Measuring Influence and Topic Dependent Interactions in Social Media Networks Based on a Counting Process Modeling Framework

by

Donggeng Xia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2015

Doctoral Committee:

Professor Moulinath Banerjee, Co-Chair
Professor George Michailidis, Co-Chair
Associate Professor Qiaozhu Mei
Assistant Professor Ambuj Tewari

To my parents

ACKNOWLEDGEMENTS

Firstly I wish to express my sincerest gratitude to my advisor Prof. George Michailidis. He introduced me to this topic of social network analysis and without his constant support, encouragements and invaluable insights this work would not have been possible. I thank him for being patient with me and teaching me the importance of hard work in every walk of life. I feel fortunate to have him as my mentor and the lessons that I learned through this journey will stay with me for the rest of my life.

I would also like to thank my committee co-chair Prof. Moulinath Banerjee, for his time and suggestions for the improvement of the theoretical proof throughout my dissertation. I also owe him additional thanks for his patient help and guidance with the course work at the beginning of my PhD. I feel lucky to have found a collaborator in Dr. Shawn Mankad, his incredible drive and work ethics is a source of constant inspiration. I also wish to thank Prof. Qiaozhu Mei and Prof. Ambuj Tewari for being members of my dissertation committee and providing many useful comments.

Moreover, I want to express my deepest appreciation to my beloved parents for their constant support. The friends I have made here in Michigan have made my PhD life so much more colorful and enjoyable.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Measuring Influence and Topic Dependent Interactions in Social Media Networks
Based on a Counting Process Modeling Framework

by

Donggeng Xia

Advisor: Professor George Michailidis

Data extracted from social media platforms, such as Twitter, are both large in scale and complex in nature, since they contain both unstructured text, as well as structured data, such as time stamps and interactions between users. Some key questions for such platforms are (i) to determine influential users, in the sense that they generate interactions between members of the platform and (ii) identifying important interactions between nodes in the corresponding user network.

Regarding the first question, common measures used both in the academic literature and by companies that provide analytics services are primarily variants of the popular web-search PageRank algorithm applied to networks that capture connections between users. In this work, we develop a modeling framework using multivariate interacting counting processes to capture the detailed actions that users undertake on such platforms, namely posting original content, reposting and/or mentioning other users' postings. Based on the proposed model, we also derive a novel influence measure. We discuss estimation of the model parameters through maximum likelihood and establish their asymptotic properties. The proposed model and the accompanying influence measure are illustrated on a data set covering a five year period of the Twitter actions of the members of the US Senate, as well as mainstream news organizations and media personalities.

We then turn our attention to the problem of identifying important interactions both globally and also based on the particular topics under discussion. We modify the previously introduced modeling framework, so that topic dependent interactions can also be identified. We extend our previous algorithm to accommodate the new framework and also establish asymptotic properties of the key model parameters. We illustrate the results on the same Twitter data set.

CHAPTER I

Introduction

1.1 Background and Literature Review

Leading business and non-profit organizations are integrating growing volumes of increasingly complex *structured* and *unstructured* data to create big data ecosystems for content distribution, as well as to gain insights for decision making. A recent, substantial area of growth has been online review and social media platforms, which have fundamentally altered the public discourse by providing easy to use forums for the distribution and exchange of news, ideas and opinions. The focus in diverse areas, including marketing, business analytics and social network analysis, is to identify trends and extract patterns in the vast amount of data produced by these platforms, so that more careful targeting of content distribution, propagation of ideas, opinions and products, as well as resource optimization is achieved (*Dave*, 2015; *Probst et al.*, 2013).

One platform that has become of central importance to both business and non-profit enterprises is Twitter. According to its second quarter 2014 financial results announcement, Twitter had more than half a billion users in July 2014, out of which more than 271 million were active ones (*Twitter*, 2014). Although Twitter lags behind in terms of active users to Facebook, it is nevertheless perceived by most businesses and non-profit organizations as an integral part of their digital presence (*Bulearca* *and Bulearca*, 2010).

The mechanics of Twitter are as follows: the basic communication unit is the account. The platform allows account users to post messages of at most 140 characters, and thus has been described as the Short Message Service (SMS) of the Internet. As of mid-2014, over half a billion messages were posted on a daily basis. Further, Twitter allows accounts to “follow” other accounts, which means the follower receives notification whenever the followed account posts a new message. Thus, the follow-follower relations serve as a primary channel for content to spread within the social networking platform. Accounts tend to interact with each other over these channels in two directed ways. First, an account can *copy* or *rebroadcast* another account’s tweet, which is referred to as a “retweeting”. Second, an account can *mention* another account within a tweet by referring to their account name with the @ symbol as a prefix. These two actions, retweeting and mentioning, are directed responses from one account to another and thus, provide the mechanisms for online conversation.

The mechanics of Twitter, together with the original messages generated by users, give rise to rich Big Data. Specifically, the content of the message, together with easily searchable key terms or topics that use the # symbol as a prefix, constitute a large corpus of unstructured text. The hashtag function enables searches to identify emerging themes and topics of discussion. In 2014, more than 2.1 billion search queries were generated (*Twitter*, 2014). Further, the following built-in capability, creates a network for *potential information flow and dissemination*, while the retweeting and mentioning actions create subnetworks of *actual interactions* between user accounts.

A key problem in all social networking platforms is that of identifying *user influence*, since such users are capable of driving action (e.g. steer discussions to particular themes and topics) or provoking interactions amongst other users and thus, are also potentially more valuable to businesses (*Trusov et al.*, 2010). In fact, as argued in *SAS* (2015), insight from social networking platforms “enhance the customer journey across all customer touch points - customer care, brand marketing, public and community relations, merchandising and more." Thus, the ranking of Twitter users based on their influence constitutes both an active research topic and a business opportunity, as manifested by services such as Klout (*Klout*, 2014) and PeerIndex (*PeerIndex*, 2014) that market and sell to businesses and other organizations influence scoring metrics. The most standard metric employed is the number of followers an account has. However, a number of studies (*Cha et al.*, 2010; *Weng et al.*, 2010) have concluded that it is not a good indicator, since most followers fail to engage with the messages that have been broadcast. For that reason, the number of retweets an account receives (*Kwak et al.*, 2010) is a better measure of influence. Since we are interested in ranking of users, more sophisticated influence measures based on the popular PageRank (*Page et al.*, 1999) and HITS (*Kleinberg*, 1999) ranking algorithms, widely used for ranking search results on the Web, have been used (*Haveliwala*, 2003; *Kwak et al.*, 2010; *Weng et al.*, 2010; *Gayo-Avello et al.*, 2011). However, these algorithms have been developed for and applied to the followers network, which clearly captures the general popularity of users, but not necessarily of their influence. For example, the twenty most followed accounts with a minimum of 25 million followers comprise of entertainers and athletes, the sole exception being President Obama.

1.2 Outline of the Thesis

In Chapter 2, we propose to measure an account/user's influence on the Twitter social media platform, by taking into consideration both their ability to produce new content by posting messages, and also to generate interactions from other accounts through retweeting and mentioning. To that end, we build a statistical model for an account's actions and interactions with other accounts. It uses a counting process framework to capture the posting, retweeting and mentioning actions. In addition, based on this model we introduce a novel *influence measure* that leverages both the follower network (that captures the potential for posted messages to generate interactions with other users) and the *intensity* over time of the basic actions involved (posting, retweeting and mentioning).

Chapter 3 considers the problem of identifying important interactions between nodes in the user network. In our proposed framework, as presented in Chapter 2, we still model actions occurring on the nodes as counting processes. However, we allow for a much more flexible parameterization than the one used in the previous chapter. Instead of having two global parameters for each node, reflecting capability to generate responses (α) and susceptibility to respond to other nodes' actions (β), we allow for independent parameters between every pair of nodes for *selected topics*. We then define an edge's importance as the expected "influence" the followed node can borrow from its follower on the other end of the edge, after a unit length of time, with a single action.

Hence,underlying the model in this thesis is the idea that conversations, and in particular the rate of directed activity, between accounts reveal their real-world position and influence. The modeling frameworks of the two chapters are illustrated on a closely knit community, namely that of the members of the United States Senate, the upper legislative house in the bicameral legislative body for the United States. Two senators are democratically elected to represent each state for six year terms. We further augment the set of Twitter accounts analyzed by including selected prominent news organizations (e.g. Financial Times, Washington Post, CNN), as well as popular bloggers (e.g. Nate Silver, Ezra Klein), the accounts of President Obama and the White House, and two influential federal agencies (the US Army and the Federal Reserve Board); for details refer to Section 3.6. Thus, we examine an ecosystem of key participants that influence the political conversation and discourse of the country.

The retweeting and mentions interactions from our data are drawn as directed edges in Figure 1.1. Given this sequence of network snapshots, we identify particular

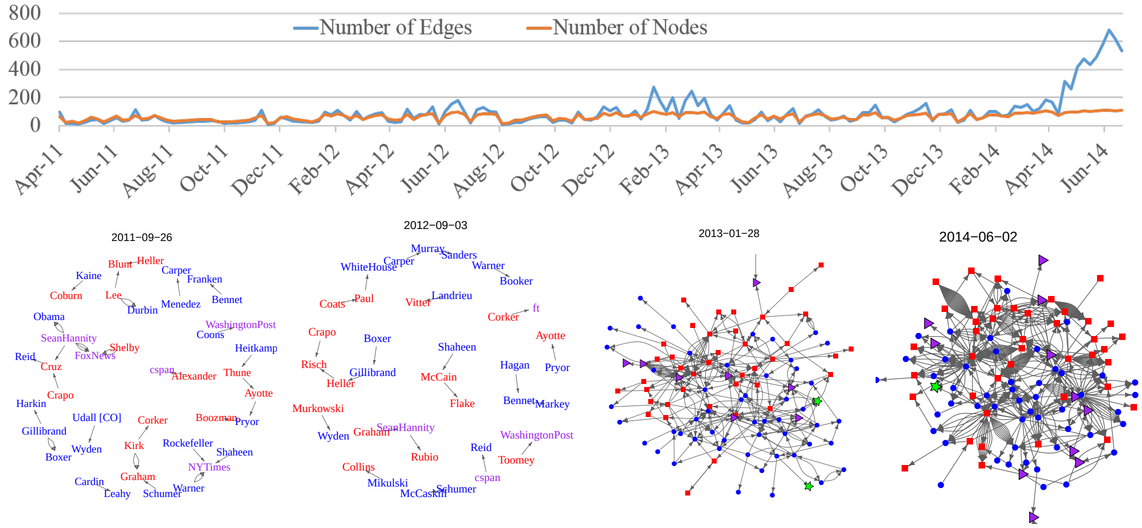

Figure 1.1: Weekly Twitter (mentions and retweet) network statistic time-series and drawings. The nodes (Twitter accounts) contain democratic senators (blue circles), republican senators (red squares), media (purple triangles), and government agencies (green stars).

senators and news agencies that tend to elicit interactions from other accounts (i.e., have many incoming edges relative to how often they tweet), thus revealing their influence on Twitter. Our results in Section 3.6 further indicate that the proposed approach produces influence measures for the U.S. Senators that correspond more closely with their legislative importance than purely network-based solutions based on the PageRank algorithm.

CHAPTER II

Measuring Influence in Twitter Ecosystems using a Counting Process Modeling Framework

2.1 Background and Literature Review

There has been a great deal of work on ranking nodes in online social networks by their influence motivated by fundamental questions in marketing, such as how to identify the best set of users to create cascades or viral campaigns. *Probst et al.* (2013), in an extensive survey article, find that the most common measures to quantify the influence of a certain node are completely based on network topology and fail to account for "further characteristics of influential users" or the actual dynamics on the social network. They identify several papers that propose variations to the core idea of measuring influence with network metrics of the followers network. An illustration of the standard methodology with data similar to ours is *Dubois and Gaffney* (2014), where Canadian political communities on Twitter are explored using degree, clustering coefficient, and other network metrics calculated from the followers network to identify "opinion leaders", i.e., accounts that steer online conversations.

To create a more nuanced influence measure that addresses the challenges highlighted by *Probst et al.* (2013) and references therein, researchers have begun to utilize the content of the communication like the underlying topic or theme of conversation, which allows for more realistic models, since some individuals are authoritative or receptive to others only along certain topical dimensions. As such, a number of recent works have extended the classical network topology measures to account for topic of conversation. *Haveliwala* (2003) and *Weng et al.* (2010) take into account topic similarity of the actual messages and the social link (followers network) structure via modified PageRank algorithms that are applied to the followers network. *Barbieri et al.* (2013) propose a similar idea for the related problem of identifying the optimal choice of initial users for inducing cascades. The model we propose relates to these previous works by also separating behavior according to the topic of conversation. Our contribution lies in measuring influence with actual conversation dynamics by combining the mentions and retweets along different topics with the followers link structure.

Our approach extends recent work in the Statistics community, which uses counting processes to combine conversation dynamics (mentions and retweets) with the followers network structure. In this stream of literature, the hazard rate represents a measure of influence and typically quantifies the effect of a message from one node on each of its followers (*Gomez-Rodriguez et al.*, 2013; *Du et al.*, 2012). Thus, as in (*Fleming and Harrington*, 2013), the interactions between nodes are modeled as *independent* counting processes. The model posited in this work exhibits certain key differences, as illustrated in Figure 2.1, because the hazard rate of a node to retweet or mention is a function of the cumulative effect of tweets from its followers. The use of *interacting* counting processes is an important modeling nuance, since it allows for more realistic account behavior. For instance, accounts that are very popular and receive many tweets on the same topic within a short period of time usually respond once both out of convenience and to avoid spamming their followers. Thus, the model we posit should result in more accurate influence measures for Twitter ecosystems like the US Senate that we investigate in Section 7.

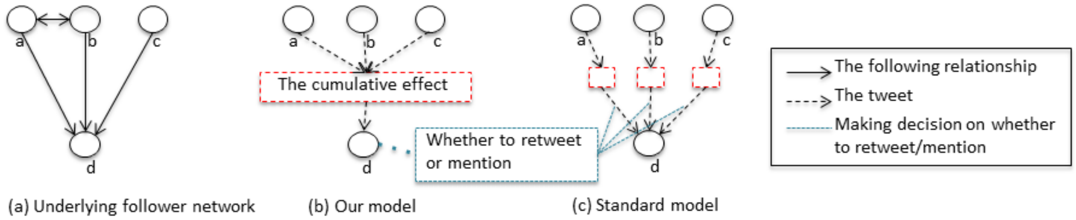

Figure 2.1: Solid lines in panel (a) represent edges in the followers network. Panel (b) illustrates the proposed model, where node d decides to retweet or mention by the cumulative effect of the three tweets from nodes a, b, and c. Panel (c) illustrates the standard counting process model on interactions between nodes as introduced in *Fleming and Harrington* (2013). Instead of considering the cumulative effect of the three tweets, node d makes a decision on whether to respond (retweet or mention) three separate times.

The remainder of this chapter is organized as follows: in Section 2.2, we review recent literature on measuring influence in online social networks. In Section 2.3, we introduce the modeling framework and the proposed influence measure. Section 2.4 presents the algorithm to obtain the model parameter estimates, as well as establish their statistical properties and those of the influence measure in Section 2.5. The performance of the model is evaluated on synthetic data sets in Section 2.6, while the US Senate application is presented in Section 2.7. Finally, some concluding remarks are drawn in Section 2.8.

2.2 The model and the influence measure

We start our presentation by defining some key quantities for future developments. Let $G = (V, L)$ denote the followers network, where V corresponds to the set of nodes of all the Twitter accounts under consideration and $L = \{L_{i,j}, 1 \leq i \neq j \leq n\}$ the edge set between them and captures whether an account follows another account.. Note that the network is bidirectional in nature and not necessarily symmetric, since account i may follow account j, but not vice versa. In principle, L can be dynamically evolving, but in this work we consider L to be static and not changing over time. As explained in the introductory section, in the Twitter platform, accounts (nodes) can undertake the following three actions: post a new message, retweet a message posted by another account that they follow and finally mention another account that they follow. Further, the vast majority of messages posted, retweeted or mentioned have key terms (with a # prefix) that identify the topic(s) that are discussed.

Next, we define the following two key counting processes. Let $N_j(t,l)$ denote the total number of retweets and mentions that account j generates on topic l by time t and let $A_j(t,l)$ denote the total number of posted messages by account j on topic l by time t. Define α_j to be a parameter that captures the long-term capability of account j to generate responses by other accounts from the content posted, and β_j a parameter that captures the long term susceptibility of account j to respond (retweet/mention) to the postings of the accounts it follows. In this thesis, we mainly focus on $N_j(t,l)$ since it reflects the interactions between accounts while $A_j(t,l)$ is frequently related to accounts' own habit of posting. We model $\{N_j(t,l)\}_{i=1}^n$ as a set of counting processes through their hazard rates, using a version of Cox (*Cox*, 1972) proportional hazard model; specifically, the hazard rate $\lambda_{j,l}(t)$ of process $N_j(t,l)$ is given by

$$\lambda_{j,l}(t) = \lambda_{0,l}(t) \exp\left(\sum_{i \neq j} L_{ij}(\alpha_i + \beta_j) \log(M_i(t,l) + 1)\right), \tag{2.1}$$

where

$$M_i(t,l) = (N_i(t,l) + A_i(t,l)) I(N_i(t,l) + A_i(t,l) \leq F) + F \cdot I(N_i(t,l) + A_i(t,l) > F).$$

$A_j(t,l) + N_j(t,l)$ is the total number of posting, retweets and mentions for account j on topic l by time t. And we consider the effect of seeing actions from account i can get saturated when the total number of actions reaches the constrain, F. We assume that the parameters $\alpha_i, \beta_i \in (-\infty, \infty)$, since accounts and their users may be positively or negatively inclined towards other accounts, as well as being more keen in joining specific conversations or passively retweeting messages from favorite accounts. The nonparametric baseline component $\lambda_{0,l}(t)$ is time varying. In general, we would expect this baseline to be small for large times t, since topics in social media platforms have a high churn rate; they become "hot" and generate a lot of action over short time scales and after awhile it stops being discussed (*Kwak et al.*, 2010). The model posits that account j interacts with other accounts at a baseline level $\lambda_{0,l}(t)$, modulated by its ability to generate responses by accounts in its followers network, as well as its own susceptibility to respond to accounts it follows postings and rebroadcasting of messages. Note that we model the retweet-mention process $N_j(t, l)$, since it reflects interactions between nodes and use the total activity process $M_j(t, l)$ as a covariate.

To complete the modeling framework, denote the set of topics in the data as $\{1, \dots, \Gamma\}$. Further, let $\mathcal{T}_j^l = \left\{T_{j,1}^l, \dots, T_{j,n_j^l}^l\right\}, t = 1, \cdots, n_j^l$, denote the set of time points that account j took action (post, retweet, mention) on topic l, until our end of observation time point t_0. Finally, for identification purposes, we require one member of the parameter vector $\Omega = (\alpha_1, \alpha_2, \dots, \alpha_n, \beta_1, \dots, \beta_n)$ to be set to a fixed value, and without loss of generality we set $\alpha_1 = 0$. Following, *Andersen and Gill* (1982), we employ a partial-likelihood function to obtain estimates of Ω. Specifically, we treat the baseline $\lambda_{0,l}(t)$ as a nuisance parameter and decomposing the full-likelihood to obtain

$$PL(t) = \prod_{1 \le l \le \Gamma} \left(\prod_{1 \le j \le n} \prod_{1 \le k \le n_j^l, T_{j,k}^l \le t} \frac{\lambda_j(T_{j,k}^l)}{\sum_{1 \le i \le n} \lambda_i(T_{j,k}^l)} \right)$$

Plugging the exact form of the hazard rate from (2.1) into the partial-likelihood function (PL), we get:

$$PL(t) = \prod_{1 \le l \le \Gamma} \left(\prod_{1 \le j \le n} \prod_{1 \le k \le n_j, T_{j,k}^l \le t} \frac{\exp\left(\sum_{i \ne j} L_{ij}(\alpha_i + \beta_j) \log(M_i(T_{j,k}^l, l) + 1)\right)}{\sum_{1 \le i \le n} \exp\left(\sum_{u \ne i} L_{ui}(\alpha_u + \beta_i) \log(M_u(T_{j,k}^l, l) + 1)\right)} \right) \tag{2.2}$$

2.2.1 The Influence Measure

Leveraging the structure of the model, we propose to measure an account's (node) influence as the total hazard rate change it will bring to its followers. Specifically, for an account j its relative hazard rate (ignoring the baseline) at time t is given by: $H_j = \exp\left(\sum_{k\neq j}\log(M_k(t,l)+1)L_{kj}(\alpha_k+\beta_j)\right)$. Further, the contribution of node i is $H_j^{(i)} = \exp\left(\log(M_i(t,l)+1)L_{ij}(\alpha_i+\beta_j)\right)$. Then, after some algebra we obtain that the total hazard rate change i brings to its followers can be written as:

$$TH^{(i)} = \sum_{j\neq i} L_{ij}\cdot\exp\left(\log(M_i(t,l)+1)(\alpha_i+\beta_j)\right). \tag{2.3}$$

Since $M_i(t,l)$ is a random value, we approximate it by its observed average value, $\bar{M}_i$, calculated from the data over all topics and time points. Hence, the influence measure becomes

$$\tilde{TH}^{(i)} = \sum_{j\neq i} L_{ij}\cdot\exp\left(\log(\bar{M}_i+1)(\alpha_i+\beta_j)\right). \tag{2.4}$$

Finally, we express it in a log-scale, so as to linearize the scale and make it compatible with the range of values of the response and susceptibility parameters α and β:

$$\Xi^{(i)} = \log\left[\sum_{j\neq i} L_{ij}\cdot\exp\left(\log(\bar{M}_i+1)(\alpha_i+\beta_j)\right)\right]. \tag{2.5}$$

In real application, we estimate $\Xi^{(i)}$ by using the estimated $\hat{\alpha}_i$ and $\hat{\beta}_j$ values.

2.3 Computation and Inference

Next, we present a Newton-type algorithm for computing the parameter estimates Ω. The logarithm of the partial likelihood function (3.3) is given by

$$
\begin{aligned}
LL(t) &= \log(PL(t)) \\
&= \sum_{1\le l\le \Gamma}\Bigg\{\sum_{1\le j\le n}\ \sum_{1\le k\le n_j^l, T_{j,k}^l\le t}\ \sum_{i\ne j} L_{ij}(\alpha_i+\beta_j)\log(M_i(T_{j,k}^l,l)+1) \\
&\quad - \sum_{1\le j\le n}\ \sum_{1\le k\le n_j^l, T_{j,k}^l\le t} \log\left[\sum_{1\le i\le n}\exp\left(\sum_{u\ne i} L_{ui}(\alpha_u+\beta_i)\log(M_u(T_{j,k}^l,l)+1)\right)\right]\Bigg\}
\end{aligned}
\tag{2.6}
$$

The objective function corresponds to $LL(t_0)$, which considers all events k in its equation (3.11). For the sake of notation simplicity, we will use LL to represent $LL(t_0)$ in the rest of the paper. Due to its smoothness, we employ Newton's algorithm that uses the gradient and the Hessian of LL. The detailed expressions for the gradient vector $G \equiv \nabla_\Omega LL$ and the Hessian $H \equiv \nabla_\Omega \nabla_\Omega (LL)$ are given in the Appendix.

Algorithm 1 Estimating the parameters by Newton's algorithm

```
Initialize the vector Ω value by α_1 = ... = α_n = β_1 ... = β_n = 0
Define s as a positive thresholding constant for the minimum step size
while t > s do
    Calculate G by using (2.11) and (2.12)
    Calculate H by using (2.13) to (2.18)
    Find the optimum positive τ value such that Ω − τ · H^{-1}G will maximize the
    log-partial-likelihood (3.11)
    Update Ω ← Ω − τ · H^{-1}G.
    In the updated Ω, set α_1 = 0.
end while
return Ω
```

To speed up calculations, we take advantage of the structure of the problem, as explained in detail in the Appendix.

The steps of the optimization are given in Algorithm 1. As stated in the algorithm,

s is a positive constant to judge the convergence of the the Newton's algorithm. The computational complexity of this algorithm is dominated by the computation of H. Denote by $m_n = \max_{1\leq j\leq n}\{n_j\}$. Based on (2.11) and (2.12), it costs $O(\Gamma n m_n)$ operations to calculate an entry of G. Further, since G is of dimension $2n$, it takes $O(\Gamma n^2 m_n)$ to obtain the entire G vector. Analogously, based on (2.13) to (2.18), it costs $O(\Gamma n m_n)$ operations to calculate an entry of H, if proper book-keeping is used on the results obtained for the gradient G. Further, since H is of dimension n^2, it takes $O(\Gamma n^3 m_n)$ to obtain the entire H matrix. Hence, the overall time complexity for each iteration of the algorithm is of the order $O(\max\{\Gamma n^3 m_n\})$. The time complexity for the whole algorithm is then $O(\max\{\Gamma n^3 m_n R\})$, where R is the number of repetitions needed for the algorithm to converge, which depends on the threshold s. Empirically, with $s = 10^{-3}$, in our simulations in Section 6 and real data analysis in Section 7, we found the algorithm generally converges in no more than 10 repetitions.

2.4 Properties of the $\hat{\Omega}$ estimates

Next, we establish that the estimator $\hat{\Omega}$ which maximizes (3.11) will converge to the true parameter Ω in probability under certain regularity conditions.

Theorem 1. Conditions:

A. (Bounded hazard rate) $C_0 \leq \lambda_{0,l}(t) \leq C_1$ for $0 \leq t \leq t_0$ $1 \leq l \leq \Gamma$,

B. (Bounded parameters) $\max_{1\leq i,j\leq n}\{|\alpha_i|, |\beta_j|\} \leq C_2$,

C. (Limited posting frequencies)

$$\begin{aligned} &P\left(A_j(t+h,l) - A_j(t,l) \geq 1\right) \leq C_3 \cdot h, \\ &P\left(N_j(t+h,l) - N_j(t,l) \geq 1\right) \leq C_3 \cdot h, \end{aligned} \tag{2.7}$$

when $t, h \geq 0, t + h \leq t_0$.

D. (Positive definite limit of Hessian) Let Ω' be any choosable parameter vector satisfying (B). For large enough Γ and some C_4, we have the holding condition to hold on the smallest eigenvalue of $-\nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega', t)$, at $\Omega' = \Omega, t = t_0$,

$$P\left(\lambda_{\min}\left(-\nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0} > C_4\right) \to 1, \text{ as } \Gamma \to \infty,$$

where

$$\begin{aligned} LT(\Omega, t) \equiv \Gamma^{-1} \Bigg\{ -\sum_{l=1}^{\Gamma} \lambda_{0,l}(u) \log \left\{ \sum_j \exp\left(\sum_{i \neq j} L_{ij}(\alpha'_i + \beta'_j) \log(M_i(t,l)+1) \right) \right\} \\ \cdot \left(\sum_{j=1}^{n} \exp\left(\sum_{1 \le i \le n, i \neq j} L_{ij}(\alpha_i + \beta_j) \log(M_i(t,l)+1) \right) \right) \Bigg\} \end{aligned} \tag{2.8}$$

In the four Conditions A, B, C and D above, C_0, C_1, C_2, C_3 and C_4 are all positive constants. Under these conditions, we will have:

$$\hat{\Omega} \to_P \Omega \text{ as } \Gamma \to \infty.$$

The detailed proof is given in Section 2.8.4.

When we have some information on the boundaries of the baseline hazard rate and parameter values, Condition A and B of Theorem 1 can be straight forwardly verified. We have the following Lemma 1 to show one example of counting processes in which Condition C naturally holds. It is however quite difficult to derive conditions under which Condition D will hold. As shown in Section 6, we propose to verify it empirically.

Lemma 1 When both $A_j(t,l)$ and $N_j(t,l)$ are both poisson processes and the hazard rate of $A_j(t,l)$ is smaller than a constant K, Condition C in Theorem 1 is satisfied.

The detailed proof is also presented in Section 2.8.3.

Based on Theorem 1, by leveraging the properties of continuous functions, we can establish the consistency of the proposed influence measure.

Proposition 1. Let $\Xi(t) = (\Xi^1(t), \cdots, \Xi^n(t))$ denote the n-dimensional vector of influence measures at time t. Further, denote by $\hat{\Xi}(t) = (\hat{\Xi}^1(t), \cdots, \hat{\Xi}^n(t))$ their empirical estimates. Under the conditions of Theorem 1, we have that

$$\left\|\hat{\Xi}(t) - \Xi(t)\right\| \to_P 0 \tag{2.9}$$

for any $t \geq 0$.

Based on Theorem 1, the proof of the proposition is straightforward, since each element of the vector $\hat{\Xi}$ is a continuous function of $\hat{\Omega}$.

2.5 Performance evaluation

In this section, we evaluate the proposed model and influence measure on synthetic data. We start by outlining the data generation mechanism.

Step 1: Building the followers network L.

The tasks employed for step 1 are presented next.

- First, for each node i, generate $K_1(i)$ from a uniform distribution on the integers $\{1, \ldots, K\}$, where $K = \lfloor *n/2 \rfloor$ and $\lfloor *\cdot \rfloor$ is the floor function that returns the maximum integer not larger than the value inside.

- Generate $F_1(i)$ for node i by randomly sampling $K_1(i)$ users from $\{1, \ldots, n\} \backslash \{i\}$. If $k \in F_1(i)$, let $L_{ik} = 1$, $1 \leq i \leq n$.

- For each node j, sample $K_2(j)$ uniformly from the set $\{1, \ldots, K\}$. Generate $F_2(i)$ for node j by randomly sampling $K_2(j)$ users from $\{1, \ldots, n\} \backslash \{j\}$. If $k \in F_2(j)$, let $L_{kj} = 1$, $1 \leq j \leq n$.

At the end of this procedure, every node in the network has at least one follower and at least an account that it follows.

Step 2. Generate the post, retweets and mentions sequences.

Since the baseline hazard rate $\lambda_{0,l}(t)$ always gets canceled out within the partial-likelihood function (3.3), we select $\lambda_{0,l}(t)$ as $\lambda_{0,l}(t) = a$, whenever $0 \leq t \leq t_0$ and $\lambda_{0,l}(t) = 0$ when $t > t_0$, where a is a small positive constant.

We then generate actions on this network with Algorithm 2 below for each topic $l \in \Gamma_1$ or Γ_2. In this algorithm, we first let each node send out a number of tweets with distribution Binomial(J, p) at $t = 0$. Then we generate the retweets and mentions in the standard survival analysis way, by using the hazard rate (2.1), as in the algorithm below.

Algorithm 2 Generate Group A actions

Initialize Indicator which is the sequence to record the nodes that have mentioned or retweeted as an empty sequence.
Initial t=0. Let each node has a tweet with probability p.
Let each node send out tweets from Binomial(J, p).
while $t < t_0$ (stopping time for all topics) **do**
    Generate survival time for each node with its hazard rate (2.1)
    Find node i with the shortest time t_s.
    if $t + t_s < t_0$ **then**
        Update t to be $t+t_s$. Record the node that has done this retweet or mention.
    end if
    if $t + t_s > t_0$ **then**
        Break
    end if
end while
return Indicator

We first illustrate the performance of the Newton estimation algorithm, on a random network of varying size. We set the parameter $a = 0.5$ for the baseline hazard rate and choose a time horizon of $t_0 = 7$, to emulate a week's worth of data. We also select the parameters Ω uniformly at random in the interval $[-0.3, 0.3]$.

Due to the bounded baseline hazard rate and simulated parameters, and since the retweets and mentions are generated as Poisson, Condition A, B, C of Theorem 1 have been satisfied. Then we empirically "check" Condition D. With a large $\Gamma = 1000$, network size $n = 10, 50$, we repeated Step 1 and 2 for 20 times to simulate the network and actions. In each repetition, the square root of the smallest eigenvalue of $\lambda_{\min}\left(-\Gamma \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0}$ is computed. The results are plotted in Figure 2.2.

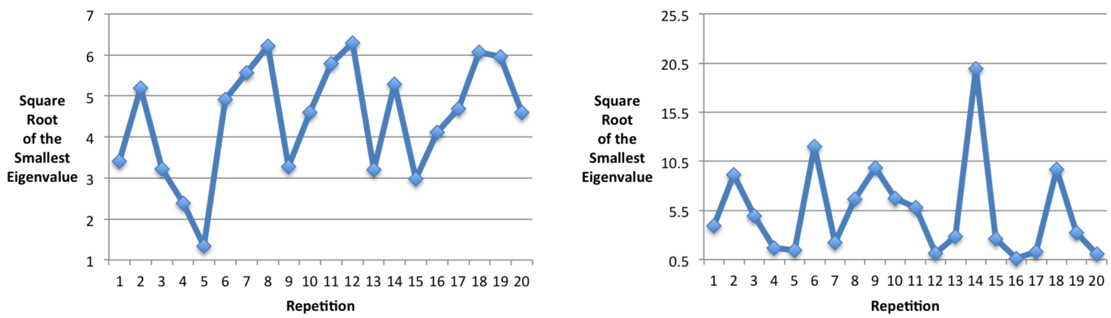

Figure 2.2: Diagnostics for Condition D with the simulated data: $\left[\lambda_{\min}\left(-\Gamma \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t)\right)\ |_{\Omega'=\Omega, t=t_0}\right]^{1/2}$ at $\Gamma = 1000$, $n = 10$ (left) and $n = 50$ (right). Due to large variations, the square root of the smallest eigenvalues is shown for better visualization.

In the plot, it can be seen that smallest eigenvalues of

$$\left[\lambda_{\min}\left(-\Gamma \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0}\right]^{1/2}$$

are generally large and greater than 0.5.

Then as we have verified all conditions are satisfied with network sizes $n = 10, 50$ and $\Gamma = 1000$, we plot in Figure 2.3 the mean squared error of the parameter and influence estimates $\frac{\|\hat{\Omega}-\Omega\|}{\sqrt{2n-1}}$ and $\frac{\|\hat{\Xi}-\Xi\|}{\sqrt{n}}$ to check the performance of our estimation algorithm, where $\|\cdot\|$ corresponds to the ℓ_2 norm of a vector. The results are based on 20 replicates of the underlying followers networks, as well as the actions (postings, retweets and mentions) data.

It can be seen that the quality of the estimates improves as a function of the number Γ of topics discussed, while it deteriorates as a function of the number of

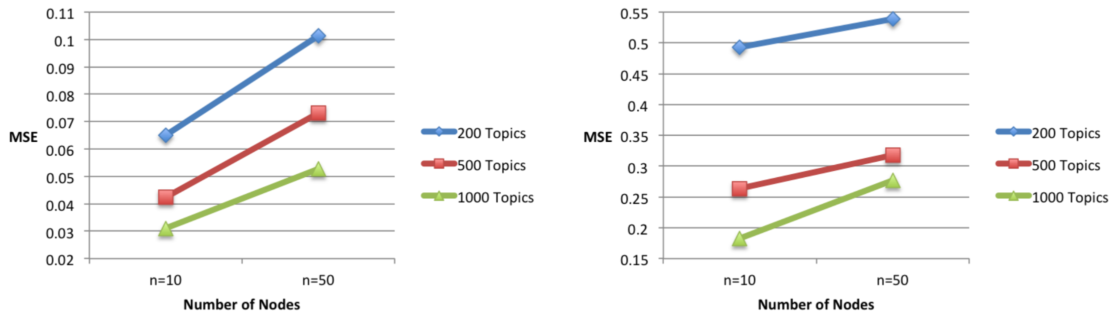

Figure 2.3: Mean squared error of the model parameter estimates Ω (left) and Ξ (right).

nodes in the followers network L. Another way to look at the quality of the estimates, is to examine the relative error of the parameter and influence estimates, given by $\frac{\|\hat{\Omega}-\Omega\|}{\|\Omega\|}$ and $\frac{\|\hat{\Xi}-\Xi\|}{\|\Xi\|}$.

It can be seen in the following Figure that especially the influence measure which is of prime interest in applications, exhibits a small (less than 10%) relative error rate.

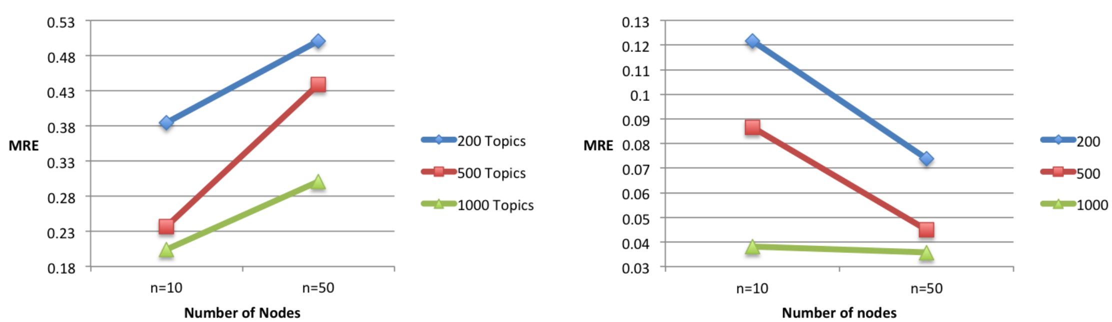

Figure 2.4: Mean relative error of the model parameter estimates Ω (left) and Ξ (right).

Next, we use a size 10 network, specially constructed to gain insight into the workings of the proposed influence measure. The settings for the data generation are as follows: $\Gamma = 500$, $\alpha_1 = 0, \alpha_2 = -2, \alpha_3 = \cdots = \alpha_{10} = 0.2$ and $\beta_1 = \cdots = \beta_{10} = 0$. Finally, the topology of the followers networks is given in Figure 2.5.

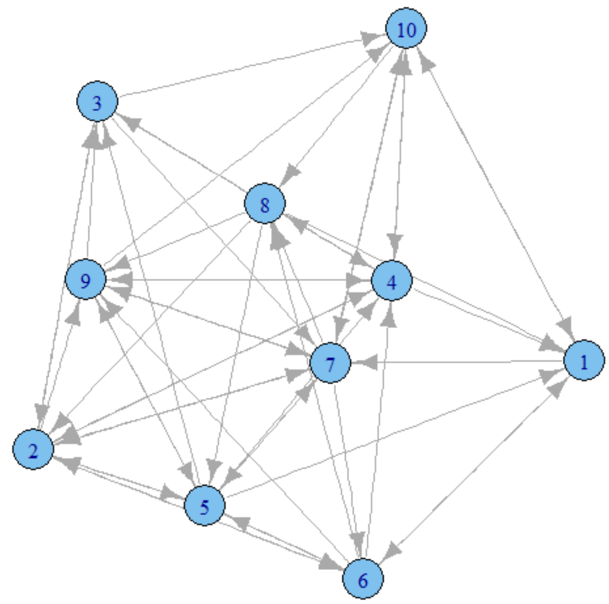

Figure 2.5: Artificial topology of a plot with "unpopular" node.

Since $\alpha_2 = -2$, node 2 is an "unpopular" one and hence can hardly generate any retweets and mentions of its postings. On the other hand, all nodes have approximately an equal number of followers, which suggests that their ranking according to the PageRank metric (or many other popular ones based on that network like *Haveliwala* (2003) and *Weng et al.* (2010)) will be approximately similar. The results based on a single realization of the user actions data generation process is shown in Figure 2.6. It can be seen that relying on the followers network structure gives a false impression, while the proposed influence measure that incorporates the actions of the accounts provides a more insightful picture.

2.6 Identifying Influential Senators

Tweets and follower lists are collected using Twitter's API and consist of approximately 200,000 tweets and 4671 follower links within the set of 120 accounts from April 2009 to July 2014. The retweeting and mentions interactions are drawn in Fig-

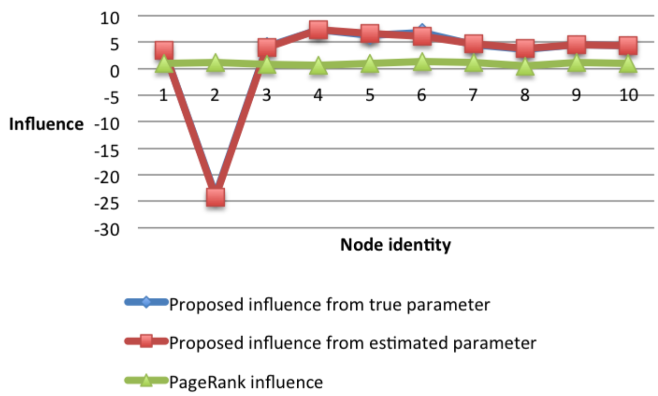

Figure 2.6: Proposed influence VS PageRank Influence, in a plot with "unpopular" node.

ure 1.1, where accounts are registered to 55 Democratic politicians (U.S. Senators and the President of the U.S.), 46 Republican Senators, 2 government organizations (U.S. Army and the Federal Reserve Board), and 16 media outlets, including newspapers (Financial Times, Washington Post, New York Times, Huffington Post), television networks (MSNBC, Fox News, CNN, CSPAN), reporters (Nate Silver (538), Ezra Klein) and television hosts (Bill O'Reilly, Sean Hannity). The figure shows some periods of increased activity, as in the months surrounding the inauguration of President Obama (January 2013), the debate on raising the debt ceiling of the US government and its temporary suspension around April 2013 and the summer of 2014 (soccer World Cup). Note that the sudden increase during the summer of 2014 may be an artifact of rate limiting data acquisition. Specifically, Twitter's API allows access to only the past 3000 tweets for any account. As a consequence, for extremely high volume users, like newspapers and television networks, our data traces their Twitter usage for months. For the least active users in our data, 3000 tweets dates back multiple years.

An inspection of actual tweets in Table 2.1 shows, consistent with *Golbeck et al.* (2010), that senators tend to retweet and mention as a means of self or legislative promotion. In fact, we see a number of references to legislative activity, such as calls for gun reform, carbon emissions, and references to actual bills on overtime pay, domestic violence protections, among others. Senators often cite news coverage by retweeting or mentioning news media accounts that support their political agenda, which would suggest that the media outlets collectively have enormous influence. This also suggests that Twitter is utilized by senators as part of a larger strategy to build and coalesce public support in order to pass bills through congress.

To test these hypotheses and also rigorously compare the proposed influence measure to PageRank applied to the followers networks (which constitutes the backbone of many ranking algorithms of Twitter accounts), we perform a regression analysis to assess how well each measure explains *legislative leadership* in Congress. Our response variable is the leadership score, published by *www.govtrack.us* (*GovTrack.us*, 2014). GovTrack creates the leadership score by applying the PageRank algorithm to the adjacency matrix of bill cosponsorship data. Thus, the leadership score for each senator is a number between 0 and 1, where higher values denote greater legislative leadership. The regression model we are interested in is

$$\text{Leadership} = \beta\text{Influence} + \Theta\text{Controls}, \tag{2.10}$$

where Influence contains the proposed measure and/or PageRank, and Controls includes party affiliation, gender, age, and number of years in the senate. Seniority endows a number of benefits including preferential assignment to committees. Thus, these control variables likely associate strongly with legislative leadership.

To estimate the proposed influence measure, the data is organized into weekly intervals after using the follow-follower relations to construct the adjacency matrix L. In Twitter it is common to use "hashtags" or the # symbol followed by a user-specified category to identify context, which, as mentioned in Section 1 can be used as an indicator of different conversations. However, we find that senators do not

Table 2.1: Actual tweets mentioning or retweeting the most influential accounts over from May 15, 2014 to July 3, 2014.

Date	Account	Tweet
05/19/2014	Menendez	".@SenBlumenthal & in #NJ the avg student loan debt is over $29K. It's unacceptable! #GameofLoans http://t.co/hUJMSeJbfd"
05/23/2014	Cornyn	"RT @nytimes: Former Defense Secretary Gates Is Elected President of the Boy Scouts http://t.co/C7STUSVIP3"
05/27/2104	Blumenthal	"RT @msnbc: @SenBlumenthal calls for reviving gun reform debate after mass shooting near Santa Barbara: http://t.co/7sqtf1IAFy"
06/02/2014	Markey	"RT @washingtonpost: A huge majority of Americans support regulating carbon from power plants http://t.co/lj6ieL5D1Y http://t.co/2CA63hTqmm"
06/17/2014	Markey	"Proud to intro new bill w @SenBlumenthal 2 protect domestic violence victims from #gunviolence http://t.co/MsgK40oLiT http://t.co/ynEHrEbh2x"
06/20/2014	Blumenthal	"Proud to stand w/ @CoryBooker & others on enhancing rules to reduce truck driver fatigue. Their safety & safety of others is paramount. -RB"
06/20/2014	Markey	"Proud to support our workers and this commonsense bill w @SenatorHarkin Keeping Track: Overtime Pay, via @nytimes http://t.co/TnAS96Hro5"
06/25/2014	Durbin	"Watch now: @OfficialCBC @HispanicCaucus @CAPAC @USProgressives @SenatorCardin on racial profiling #MoreThanAProfile http://t.co/ZX0Eu65dgi
06/25/2014	Cardin	"RT @TheTRCP: Thank you @SenatorCardin for standing with sportsmen today for #CleanWater #protectcleanwater"
06/27/2014	Markey	"Thanks @alfranken @CoryBooker @amyklobuchar @SenBlumenthal for joining me in support of community #broadband http://t.co/O8Px2MzrCg"
06/27/2014	Menendez	"Took my first #selfie at #NJs @ALJBS! Hope @CoryBooker is proud of his NJ Sen colleague. http://t.co/FrEJonUy9d
06/28/2014	Booker	"Thanks Adam RT @AIsaacs7 Props to @CoryBooker and @SenRandPaul for their bipartisanship in introducing their amendment #MedicalMarijuana"

utilize hashtags often. To overcome this challenge, we follow previous works on Twitter (*Hong and Davison*, 2010; *Ramage et al.*, 2010) by applying probabilistic topic modeling, which was first introduced in *Blei et al.* (2003). Extensive work in computer science and applied statistics has led to fast algorithms capable of analyzing extremely big text archives. Due to space constraints, for statistical and algorithmic details on the topic model, see *Blei* (2012); *Blei and Lafferty* (2007) and references therein.

Topic modeling applied to the data results in a soft clustering of tweets into 10 groups (topics), which is appropriate since a single tweet could touch on multiple issues. Thus, tweets are assigned to topics that had at least 0.25 probability. Given the fast moving landscape of social media, based on the original clustered 10 groups, new topics are assigned each week, leading to 2770 topics in total for the entire data set. After preprocessing, we apply Algorithm 3 to estimate the α and β parameters for every account using all data. The final influence measure is constructed by computing the influence measure vector $\hat{\Xi}$ over different time intervals to study how influence evolved; i.e. $\hat{\Xi}$ was computed by using the average of $M_i(\mathcal{T}_m, l)$ over all time points in $\mathcal{T}_m$ and topics, where $\mathcal{T}_m$ denotes the $m-$th time interval of interest.

The first time interval $\mathcal{T}_1$ we investigate is May 15, 2014 - July 3, 2014, which captures the most active period in our data and also represents a period when rate limiting is not a concern, i.e., the data for even high volume users extends this far. During this time several major events occurred worldwide, including the soccer World Cup, debate on immigration reform, and the Islamic State in Iraq and the Levant (also known as the ISIS or ISIL) began an offensive in northern Iraq. Table 2.2 shows the top ten most influential accounts under the proposed method and PageRank (*Page et al.*, 1999) calculated from the followers network. Both methods estimate that the Financial Times is the most influential Twitter account, and in general find that the media has an enormous influence that facilitates online conversation between

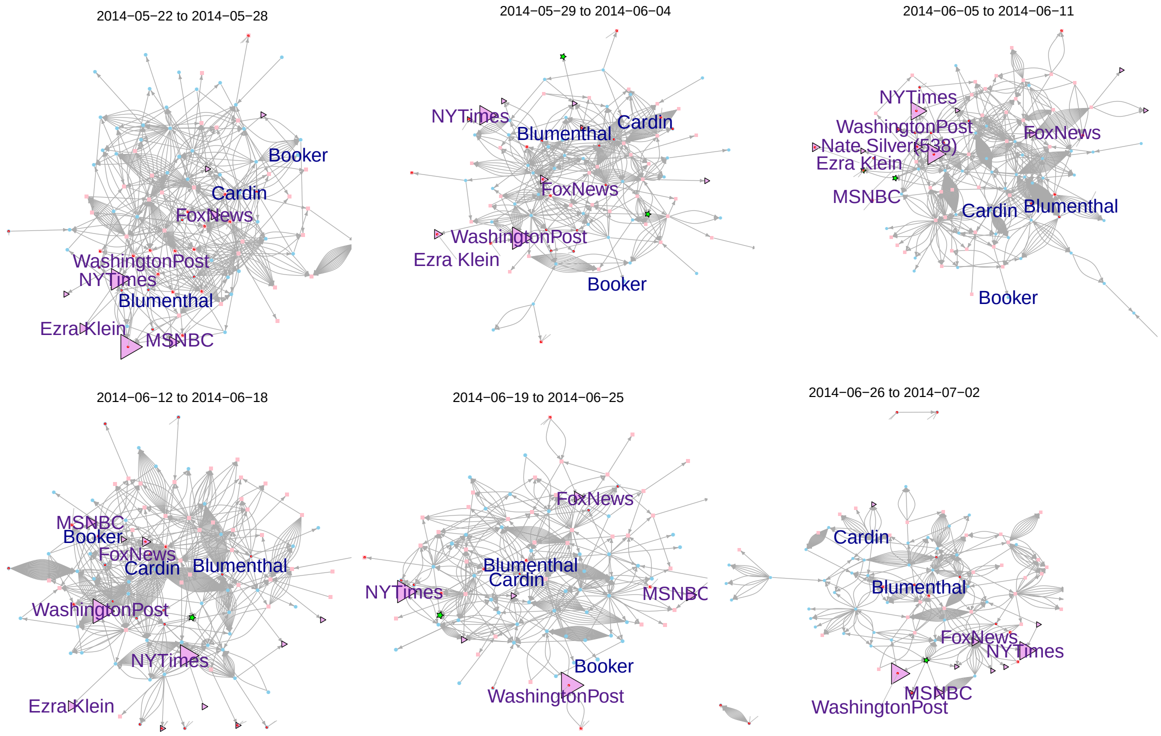

Figure 2.7: Weekly Twitter retweet and mention network drawings for the 2014 summer. Top ten most influential accounts are labeled and node sizes are proportional to the estimated influence under the proposed model. The nodes (Twitter accounts) contain democratic senators (blue circles), republican senators (red squares), media (purple triangles), and government agencies (green stars).

politicians. We see from Figure 2.7 that these top accounts were actively retweeted and mentioned throughout this period.

Next, we estimate the regression model in Equation 3.15. We note that Senators Baucus, Kerry, Cowan, Lautenberg, and Chiesa are scored by govtrack.us, but are not in our analysis. Max Baucus and John Kerry are left out, because they vacated their Senate seats to become, respectively, Ambassador to China and U.S. Secretary of State. Mo Cowan succeeded Kerry and was senator from February 1, 2013 to July 16, 2013 until a special election could be held. Cowan chose not to run in the election. Likewise, due to the death of Senator Frank Lautenberg, Jeffrey Chiesa was appointed by Governor Chris Christie to be the junior senator from New Jersey from

Table 2.2: Top ten rankings according to the proposed model and PageRank from May 15, 2014 - July 3, 2014.

Rank	Proposed Measure	PageRank
1	Financial Times	Financial Times
2	Washington Post	U.S. Army
3	NYTimes	CNN
4	MSNBC	Barack Obama
5	Ezra Klein	CSPAN
6	Fox News	New York Times
7	Cory Booker	Washington Post
8	Ben Cardin	Cory Booker
9	Nate Silver (538)	MSNBC
10	Richard Blumenthal	Wall Street Journal

Table 2.3: Estimated R-squared values for different regression models, where the proposed measure and/or PageRank is included in the set of independent variables and the influence is computed for the entire data sample. We consistently find that the proposed measure is a better indicator of legislative importance.

Response	Proposed Measure	PageRank	R^2
leadership	V		0.311
		V	0.276
	V	V	0.311
$log(\frac{\text{leadership}}{1-\text{leadership}})$	V		0.114
		V	0.098
	V	V	0.114

Table 2.4: Regression estimates, where the response variable is the raw leadership scores from GovTrack.us and influence is computed for the entire data sample. $R^2 = 0.311$; $F = 8.228$ on 5 and 92 DF (p-value: 0.000)

Variable	Estimate	Std. Error	t value	$P(>\|t\|)$
Intercept	-0.086	0.232	-0.368	0.714
Proposed Influence	0.062	0.028	2.241	0.027
Republican	-0.154	0.039	-3.945	0.000
Age	0.002	0.003	0.923	0.359
Years in Senate	0.007	0.003	2.518	0.014
Male	0.020	0.050	0.395	0.694

June 6, 2013 to October 31, 2013. He declined to run in the special election and thus, is also not included in the analysis.

Since the leadership score provided by GovTrack are between 0 and 1, we estimate two models. One model uses the raw leadership scores, and another uses $log(\frac{\text{leadership}}{1-\text{leadership}})$ for the response variable. In both cases, as shown in Table 2.3, we consistently find that the proposed influence measure explains more variation in leadership and when both the proposed and PageRank influence measures are included as independent variables, PageRank does not provide additional explanatory power. Tables 2.4 and 2.5 show a significant positive coefficient for the proposed influence measure, meaning that senators who are more influential in Twitter by successfully steering conversation of their colleagues onto particular topics, tend to be more influential in real life in passing legislation. These results are consistent across different time intervals. For instance, in the Appendix we present similar results, where influence is calculated from January 1, 2013 to March 1, 2013 corresponding to sequestration and also from November 1, 2012 to January 31, 2013 corresponding to the president's reelection and subsequent inauguration.

Table 2.5: Regression estimates, where the response variable is $log(\frac{\text{leadership}}{1-\text{leadership}})$, where leaderhip is from GovTrack.us and influence is computed for the entire data sample. $R^2 = 0.114$; $F = 2.334$ on 5 and 92 DF (p-value: 0.048)

Variable	Estimate	Std. Error	t value	$P(>\|t\|)$
Intercept	-3.590	2.604	-1.379	0.171
Proposed Influence	0.437	0.308	1.416	0.160
Republican	-1.112	0.438	-2.538	0.013
Age	0.009	0.029	0.323	0.747
Years in Senate	0.034	0.032	1.063	0.290
Male	0.470	0.563	0.834	0.407

2.7 Discussion

The goal in this paper was to characterize the influence of users in a large scale social media platform when given information about the detailed *actions* users take on it. Our comprehensive analysis of the ecosystem comprising of US Senators and influential government agency and media related accounts demonstrated that conversations, and in particular the rate of directed activity, between accounts are correlated with their real-world position and influence. We expect similar conclusions to hold broadly for other types of directed interaction data when the nodes form a clearly defined ecosystem or closely knit social group/community.

The proposed approach only utilizes network information (e.g. followers network), plus time stamps of actions (e.g. retweets and mentions), thus allowing to process a large volume of data. However, it does not consider the tone of the message (positive, negative or neutral), a topic addressed in *Taddy* (2013), where the goal is to understand how messages related to a specific topic are perceived by other users. Since in that approach the message content needs to be analyzed - a computationally demanding task - *Taddy* (2013) develops efficient sampling designs for that task. It is of interest though to combine such sampling ideas with the current approach in order to be able to address user influence issues in very large ecosystems comprising of millions of users.

The modeling and statistical inference issues, associated with large scale data obtained from these social media platforms are different from those in the related literature on network community detection (*Kolaczyk*, 2009; *Fienberg*, 2012; *Salter-Townshend et al.*, 2012), where the goal is to identify relatively dense groups of nodes (users), even though the underlying data (observed adjacency matrices) are the same. Relative to other recent work on modeling directed networks, as in *Perry and Wolfe* (2013), our study has important modeling differences motivated by the online social media platform domain. For instance, our approach incorporates the fundamental differences between actions like retweeting, mentioning, and posting. As a consequence, our final influence measure, which sums all possible influences from the social network, is able to outperform traditional topology driven approaches like PageRank (*Page et al.*, 1999). Perhaps most importantly, given the massive volumes of data generated by platforms like Twitter, we presented a fast estimation algorithm and established statistical properties for the model estimates and those of the final influence measure.

2.8 Estimation Algorithm and Proofs

2.8.1 Expressions for the gradient vector and Hessian matrix of the LL function

Some rather straightforward algebra yields the following expressions for the elements of the gradient vector $G \equiv \nabla_{\Omega} LL$:

$$
\begin{aligned}
\frac{\partial LL}{\partial \alpha_i} = \sum_{1 \le l \le \Gamma} \Bigg\{ & \sum_{1 \le j \le n, j \ne i} \sum_{1 \le k \le n_j^l} L_{ij} \log(M_i(T_{j,k}^l, l) + 1) \\
& - \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{\sum_{v \ne i} L_{iv} \log(M_i(T_{j,k}^l, l) + 1)}{\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv}(\alpha_u + \beta_v) \log(M_u(T_{j,k}^l, l) + 1)\right)} \\
& \cdot \exp\left(\sum_{u \ne v} L_{uv}(\alpha_u + \beta_v) \log(M_u(T_{j,k}^l, l) + 1)\right) \Bigg\}
\end{aligned}
\tag{2.11}
$$

for $2 \le i \le n$, and

$$
\begin{aligned}
\frac{\partial LL}{\partial \beta_j} = \sum_{1 \le l \le \Gamma} \Bigg\{ & \sum_{1 \le k \le n_j^l} \sum_{i \ne j} L_{ij} M_i(T_{j,k}^l, l) \\
& - \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{\left(\sum_{u \ne j} L_{uj} \log(M_u(T_{s,k}^l, l) + 1)\right)}{\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv}(\alpha_u + \beta_v) \log(M_u(T_{s,k}^l, l) + 1)\right)} \\
& \cdot \exp\left(\sum_{u \ne j} L_{uj}(\alpha_u + \beta_j) \log(M_u(T_{s,k}^l, l) + 1)\right) \Bigg\}
\end{aligned}
\tag{2.12}
$$

for $1 \le j \le n$.

Next, we obtain the necessary expressions for the Hessian matrix $H(LL)$. We start by computing the sub-matrix of H that includes the second partial derivatives of LL with respect to the α parameters and obtain

$$\frac{\partial^2 LL}{\partial \alpha_i^2} =$$

$$\sum_{1\le l\le \Gamma} \Bigg\{ - \sum_{1\le j\le n} \sum_{1\le k\le n_j^l} \frac{\sum_{v\neq i} L_{iv} \log(M_i^2(T_{j,k}^l, l)+1)}{\sum_{1\le v\le n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)}$$
$$\cdot \exp\left(\sum_{u\neq v} L_{uv}(T_{j,k}^l)(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)$$
$$+ \sum_{1\le j\le n} \sum_{1\le k\le n_j^l} \frac{\left[\sum_{v\neq i} L_{iv} \log(M_i(T_{j,k}^l, l)+1) \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)\right]^2}{\left[\sum_{1\le v\le n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)\right]^2} \Bigg\} \tag{2.13}$$

When $i \neq q$, we similarly have

$$\frac{\partial^2 LL}{\partial \alpha_i \partial \alpha_q} = \sum_{1\le l\le \Gamma} \Bigg\{ - \sum_{1\le j\le n} \sum_{1\le k\le n_j^l} \frac{\sum_{v\neq i,q} L_{iv} \log(M_i(T_{j,k}^l, l)+1) L_{qv} \log(M_q(T_{j,k}^l, l)+1)}{\sum_{1\le v\le n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)}$$
$$\cdot \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)$$
$$+ \sum_{1\le j\le n} \sum_{1\le k\le n_j^l} \frac{\sum_{v\neq i} L_{iv} \log(M_i(T_{j,k}^l, l)+1)}{\left[\sum_{1\le v\le n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)\right]^2}$$
$$\cdot \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)$$
$$\cdot \left[\sum_{v\neq q} L_{qv} \log(M_q(T_{j,k}^l, l)+1) \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{j,k}^l, l)+1)\right)\right] \tag{2.14}$$

Next, we obtain the sub-matrix of H that includes the second partial derivatives of

LL with respect to the β parameters and get

$$\frac{\partial^2 LL}{\partial \beta_j^2} = \sum_{1\leq l\leq \Gamma}$$

$$\left\{ \sum_{1\leq s\leq n} \sum_{1\leq k\leq n_s^l} \frac{\left[\left(\sum_{u\neq j} L_{uj} \log(M_u(T_{s,k}^l, l)+1)\right) \exp\left(\sum_{u\neq j} L_{uj}(\alpha_u+\beta_j)\log(M_u(T_{s,k}^l, l)+1)\right)\right]^2}{\left[\sum_{1\leq v\leq n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(N_u(T_{s,k}^l, l)+1)\right)\right]^2} \right.$$

$$\left. - \sum_{1\leq s\leq n} \sum_{1\leq k\leq n_s^l} \frac{\left(\sum_{u\neq j} L_{uj} \log(M_u(T_{s,k}^l, l)+1)\right)^2 \exp\left(\sum_{u\neq j} L_{uj}(\alpha_u+\beta_j)\log(M_u(T_{s,k}^l, l)+1)\right)}{\sum_{1\leq v\leq n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{s,k}^l, l)+1)\right)} \right\} \tag{2.15}$$

for $1 \leq j \leq n$. When $j \neq q$, we can similarly have

$$\frac{\partial^2 LL}{\partial \beta_j \partial \beta_q} = \sum_{1\leq l\leq \Gamma} \left\{ \sum_{1\leq s\leq n} \sum_{1\leq k\leq n_s^l} \frac{\left(\sum_{u\neq j} L_{uj} \log(M_u(T_{s,k}^l, l)+1)\right)}{\left[\sum_{1\leq v\leq n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{s,k}^l, l)+1)\right)\right]^2} \right.$$
$$\cdot \exp\left(\sum_{u\neq j} L_{uj}(\alpha_u+\beta_j)\log(M_u(T_{s,k}^l, l)+1)\right)$$
$$\left. \cdot \left(\sum_{q\neq j} L_{uq} \log(M_u(T_{s,k}^l, l)+1)\right) \exp\left(\sum_{u\neq q} L_{uj}(\alpha_u+\beta_q)\log(M_u(T_{s,k}^l, l)+1)\right) \right\} \tag{2.16}$$

Finally, we provide expressions for the cross-partials

$$\frac{\partial^2 LL}{\partial \alpha_i \partial \beta_i} = \sum_{1\leq l\leq \Gamma} \left\{ \sum_{1\leq s\leq n} \sum_{1\leq k\leq n_s^l} \frac{\left(\sum_{u\neq i} L_{ui} \log(M_u(T_{s,k}^l, l)+1)\right)}{\left[\sum_{1\leq v\leq n} \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{s,k}^l, l)+1)\right)\right]^2} \right.$$
$$\cdot \exp\left(\sum_{u\neq i} L_{ui}(\alpha_u+\beta_i)\log(M_u(T_{s,k}^l, l)+1)\right)$$
$$\left. \cdot \sum_{v\neq i} L_{iv} \log(M_i(T_{s,k}^l, l)+1) \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T_{s,k}^l, l)+1)\right) \right\} \tag{2.17}$$

When $i \neq j$,

$$
\begin{aligned}
\frac{\partial^2 LL}{\partial\alpha_i\partial\beta_j} = \sum_{1\le l\le\Gamma} \Bigg\{ & \sum_{1\le s\le n}\sum_{1\le k\le n_s^l} \frac{\left(\sum_{u\neq j} L_{uj}\log(M_u(T^l_{s,k},l)+1)\right)}{\left[\sum_{1\le v\le n}\exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T^l_{s,k},l)+1)\right)\right]^2} \\
& \cdot \exp\left(\sum_{u\neq j} L_{uj}(\alpha_u+\beta_j)\log(M_u(T^l_{s,k},l)+1)\right) \\
& \cdot \sum_{v\neq i} L_{iv}\log(M_i(T^l_{s,k},l)+1)\exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T^l_{s,k},l)+1)\right) \\
& - \sum_{1\le s\le n}\sum_{1\le k\le n_s^l} \frac{\left(\sum_{u\neq j} L_{uj}\log(M_u(T^l_{s,k},l)+1)\right) L_{ij}\log(M_i(T^l_{s,k},l)+1)}{\sum_{1\le v\le n}\exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T^l_{s,k},l)+1)\right)} \\
& \cdot \exp\left(\sum_{u\neq j} L_{uj}(\alpha_u+\beta_j)\log(M_u(T^l_{s,k},l)+1)\right)\Bigg\}
\end{aligned}
\tag{2.18}
$$

2.8.2 Implementation Issues

As outlined above, the maximum likelihood estimator is obtained by Newton's algorithm and detailed expressions for the respective gradient and Hessian are given in Section 9.1. However, the structure of the problem allows us to precompute and store several quantities for repeated use, thus saving on computational time in practice. Note that the data containing the actions are stored according to their time stamps. We start by computing four groups of quantities introduced by an action, labeled respectively by indices j, l, k and possibly some other parameters, where j indicates the node that takes the activity, l is the topic label and k represents the relative sequence number of the action, in all the actions that node j has taken under topic l.

First, we define

$$
E_{j,v,k,l} = \exp\left(\sum_{u\neq v} L_{uv}(\alpha_u+\beta_v)\log(M_u(T^l_{j,k},l)+1)\right).
$$

Then, we compute

$$SE_{j,k,l} = \sum_{1 \le v \le n} E_{j,v,k,l},$$

and

$$ME_{j,i,k,l} = \sum_{v \ne i} L_{iv} \log(M_i(T^l_{j,k}, l) + 1) E_{j,v,k}.$$

Also, we have

$$LM_{j,s,k,l} = \sum_{u \ne j} L_{uj} \log(M_u(T^l_{s,k}, l) + 1).$$

Then, based on the precomputed components values, the elements of the gradient vector $G \equiv \nabla_\Omega LL$ are obtained as follows:

$$\frac{\partial LL}{\partial \alpha_i} = \sum_{1 \le l \le \Gamma} \left\{ \sum_{1 \le j \le n, j \ne i} \sum_{1 \le k \le n^l_j} L_{ij} \log(M_i(T^l_{j,k}, l) + 1) - \sum_{1 \le j \le n} \sum_{1 \le k \le n^l_j} \frac{ME_{j,i,k,l}}{SE_{j,k,l}} \right\}$$

for $2 \le i \le n$, and

$$\frac{\partial LL}{\partial \beta_j} = \sum_{1 \le l \le \Gamma} \left\{ \sum_{1 \le k \le n^l_j} \sum_{i \ne j} L_{ij} M_i(T^l_{j,k}, l) - \sum_{1 \le s \le n} \sum_{1 \le k \le n^l_s} \frac{LM_{j,s,k,l} E_{j,s,k,l}}{SE_{s,k,l}} \right\}$$

for $1 \le j \le n$.

Regarding the Hessian, based on the four precomputed groups of quantities, we start by computing the sub-matrix of H that includes the second partial derivatives of LL with respect to the α parameters. We get

$$\frac{\partial^2 LL}{\partial \alpha_i^2} = \sum_{1 \le l \le \Gamma} \left\{ - \sum_{1 \le j \le n} \sum_{1 \le k \le n^l_j} \frac{\sum_{v \ne i} L_{iv} \log(M_i^2(T^l_{j,k}, l) + 1) E_{j,v,k,l}}{SE_{j,k,l}} + \sum_{1 \le j \le n} \sum_{1 \le k \le n^l_j} \frac{(ME_{j,i,k,l})^2}{(SE_{j,k,l})^2} \right\}$$

When $i \neq q$, we similarly have

$$\frac{\partial^2 LL}{\partial \alpha_i \partial \alpha_q} = \sum_{1 \le l \le \Gamma} \left\{ - \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{\sum_{v \neq i,q} L_{iv} \log(M_i(T_{j,k}^l, l)+1) L_{qv} \log(M_q(T_{j,k}^l, l)+1) E_{j,v,k,l}}{SE_{j,k,l}} \right.$$
$$\left. + \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{ME_{j,i,k,l} ME_{j,q,k,l}}{(SE_{j,k,l})^2} \right\}$$

Next, we obtain the sub-matrix of H that includes the second partial derivatives of LL with respect to the β parameters and get

$$\frac{\partial^2 LL}{\partial \beta_j^2} = \sum_{1 \le l \le \Gamma} \left\{ \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{(LM_{j,s,k,l} E_{j,s,k,l})^2}{(SE_{s,k,l})^2} - \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{(LM_{j,s,k,l})^2 E_{j,s,k,l}}{SE_{s,k,l}} \right\}$$

for $1 \le j \le n$. When $j \neq q$, we can similarly have

$$\frac{\partial^2 LL}{\partial \beta_j \partial \beta_q} = \sum_{1 \le l \le \Gamma} \left\{ \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{LM_{s,i,k,l} \cdot E_{j,s,k,l} \cdot LM_{q,s,k,l} \cdot E_{q,s,k,l}}{(SE_{s,k,l})^2} \right\}$$

Finally, we provide expressions for the cross-partials

$$\frac{\partial^2 LL}{\partial \alpha_i \partial \beta_i} = \sum_{1 \le l \le \Gamma} \left\{ \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{LM_{j,s,k,l} \cdot E_{s,i,k,l} \cdot ME_{j,i,k,l}}{(SE_{s,k,l})^2} \right\}$$

When $i \neq j$,

$$\frac{\partial^2 LL}{\partial \alpha_i \partial \beta_j} = \sum_{1 \le l \le \Gamma} \left\{ \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{LM_{j,s,k,l} \cdot E_{s,j,k,l} \cdot ME_{s,i,k,l}}{(SE_{s,k,l})^2} \right.$$
$$\left. - \sum_{1 \le s \le n} \sum_{1 \le k \le n_s^l} \frac{LM_{j,s,k,l} \cdot L_{ij} \log(M_i(T_{s,k}^l, l)+1) \cdot E_{s,j,k,l}}{SE_{s,k,l}} \right\}$$

2.8.3 Proof of Lemma 1

First, under the condition that $N_j(t,l)$ is a Poisson Process, we have

$$P(N_i(t+h,l) - N_i(t,l) = k) = \frac{(\mu_i(t,h,l))^k}{k!} \exp(-\mu_i(t,h,l)),$$

where $\mu_i(t,h,l) = \int_t^{t+h} \lambda_i(u,l)du$. Since

$$|\lambda_i(t,l)| = \left| \lambda_{0,l}(t) \exp\left(\sum_{k,k\neq i} L_{ki}(t)(\alpha_i' + \beta_j') \log(M_i(t,l)+1) \right) \right| \leq C_1 \exp\left(2nC_2 \log(F+1)\right),$$

we have

$$\mu_i(t,h,l) \leq C_1 n \exp\left(2nC_2 \log(F+1)\right) h.$$

Then,

$$\begin{aligned} P(N_i(t+h,l) - N_i(t,l) \geq 1) &= \sum_{k=1}^{\infty} \frac{(\mu_i(t,h,l))^k}{k!} \exp(-\mu_i(t,h,l)) \\ &= \exp(-\mu_i(t,h,l))\mu_i(t,h,l) \sum_{k=1}^{\infty} \frac{(\mu_i(t,h,l))^{k-1}}{k!} \\ &\leq \exp(-\mu_i(t,h,l))\mu_i(t,h,l) \sum_{k=1}^{\infty} \frac{(\mu_i(t,h,l))^{k-1}}{(k-1)!} = \mu_i(t,h,l) \\ &\leq C_1 \exp\left(2nC_2 \log(F+1)\right) h. \end{aligned}$$

Similarly, we can show

$$P(A_i(t+h,l) - A_i(t,l) \geq 1) \leq Kh.$$

If we let $C_3 = \max\{K, C_1 \exp\left(2nC_2 \log(F+1)\right)\}$, Condition C in Theorem 1 has been satisfied.

2.8.4 Proof of Theorem 1

Before we start the actual proof, to simplify the proof of Theorem 1, we first define some notations:

$$
\begin{aligned}
&E_l(t,\Omega') = \sum_{j=1}^{n} \lambda_{0,l}(t) \exp\left(\sum_{i,i\neq j} L_{ij}(t)(\alpha'_i + \beta'_j) \log(M_i(t,l)+1) \right) \\
&\Phi'_j = (\phi'_{1j}, \ldots, \phi'_{nj}) := (\alpha'_1 + \beta'_j, \ldots, \alpha'_n + \beta'_j)' \\
&E^{(1)}_{lj}(t,\Omega') = \left(\frac{\partial E_l(t,\Omega')}{\partial \phi'_{1j}}, \ldots, \frac{\partial E_l(t,\Omega')}{\partial \phi'_{nj}} \right) \\
&E^{(2)}_{lj}(t,\Omega') = \left(\frac{\partial^2 E_l(t,\Omega')}{\partial \phi'_{ij} \phi'_{kj}} \right)_{1\leq i,k\leq n}
\end{aligned}
\tag{2.19}
$$

To prove Theorem 1, we also need the following two lemmas.

Lemma 2.When Conditions A, B, C of Theorem 1 hold, if we define

$$
e_l(t,\Omega') = E[E_l(t,\Omega')] = \sum_j \lambda_{0,l}(t) E\left[\exp\left(\sum_{i,i\neq j} L_{ij}(t)(\alpha'_i + \beta'_j) \log(M_i(t,l)+1) \right) \right],
$$

we will have

$$
\sup_{t\in[0,t_0],|\alpha_i|\leq C_2,|\beta_j|\leq C_2} \Gamma^{-1} \left| \sum_{l=1}^{\Gamma} [E_l(t,\Omega') - e_l(t,\Omega')] \right| \to_p 0. \tag{2.20}
$$

$$
\sup_{t\in[0,t_0],|\alpha_i|\leq C_2,|\beta_j|\leq C_2} \Gamma^{-1} \sum_{j=1}^{n} \left\| \sum_{l=1}^{\Gamma} \left[E^{(k)}_{lj}(t,\Omega') - e^{(k)}_{lj}(t,\Omega') \right] \right\|_\infty \to_p 0. \tag{2.21}
$$

where $k = 1, 2$, $\|\cdot\|_\infty$ gives the largest absolute value of entries of a vector (matrix) and $e^{(1)}_j(t,\Omega')$ and $e^{(k)}_j(t,\Omega')$ are defined by

$$
\begin{aligned}
&e^{(1)}_{lj}(t,\Omega') = \left(\frac{\partial e_l(t,\Omega')}{\partial \phi'_{1j}}, \ldots, \frac{\partial e_l(t,\Omega')}{\partial \phi'_{nj}} \right) \\
&e^{(2)}_{lj}(t,\Omega') = \left(\frac{\partial^2 e_l(t,\Omega')}{\partial \phi'_{ij} \partial \phi'_{kj}} \right)_{1\leq i,k\leq n}
\end{aligned}
$$

Lemma 3. When Conditions A, B, C of Theorem 1 hold, following the definitions of $e_l(t,\Omega')$, $e_{lj}^{(1)}(t,\Omega')$ and $e_{lj}^{(2)}(t,\Omega')$ in Lemma 2, we have:

(1) $e_l(t,\Omega')$, $e_{lj}^{(1)}(t,\Omega')$ and $e_{lj}^{(2)}(t,\Omega')$ are continuous function of Ω' and t. Since Ω' and t can only be selected from compact sets, they are automatically uniform continuous.

(2) $e_l(t,\Omega')$, $e_{lj}^{(1)}(t,\Omega')$ and $e_{lj}^{(2)}(t,\Omega')$ are bounded on the selectable sets of $\alpha_i, \beta_j' \in [-C_2, C_2]$ and $t \in [0, t_0]$.

(3) $e_l(t,\Omega')$ is bounded away from zero.

Proof of Lemma 2

First, we focus on the proof of (2.20). Given $\epsilon_0 > 0, h > 0$ since the choosable set for Ω', $[-C_2, C_2]^n$ and $[0, t_0]$ are all bounded compact, we can have $[-C_2, C_2]^n \times [0, t_0]$ to be covered by a number of $n_{\epsilon_0,h}$ open sets $OS_i \equiv \{\Omega', t : \|\Omega' - \Omega_i\|_\infty < \epsilon_0, |t - t_i| < h, 1 \leq i \leq n_{\epsilon_0,h}$.

Then for any Ω', t we can always find it to fall in a certain, say OS_i. Now we can have

$$\sup_{t\in[0,t_0],|\alpha_i|\leq C_2,|\beta_j|\leq C_2} \Gamma^{-1} \left| \sum_{l=1}^{\Gamma} [E_l(t,\Omega') - e_l(t,\Omega')] \right|$$
$$= \max_{1\leq i\leq n_\epsilon} \sup_{(\Omega',t)\in OS_i} \Gamma^{-1} \left| \sum_{l=1}^{\Gamma} [E_l(t,\Omega') - e_l(t,\Omega')] \right|$$
$$\leq \max_{1\leq i\leq n_\epsilon} \left(\Gamma^{-1} \left| \sum_{l=1}^{\Gamma} [E_l(t_i,\Omega_i) - e_l(t_i,\Omega_i)] \right| + \sup_{(\Omega',t)\in OS_i} \Gamma^{-1} \left| \sum_{l=1}^{\Gamma} [E_l(t,\Omega') - E_l(t_i,\Omega_i)] \right| \right)$$
$$+ \sup_{(\Omega',t)\in OS_i} \Gamma^{-1} \left| \sum_{l=1}^{\Gamma} [e_l(t,\Omega') - e_l(t_i,\Omega_i)] \right|$$

The rest of the proof is organized as follows. In Step 1, we bound the first term in the last inequality above. In Step 2 and 3, we try to find appropriate ϵ_0 and h values to bound the second term, respectively. The results from Step 2 and 3 are combined together, and the third term is bounded in Step 4. In Step 5, we show (2.21).

Step 1 First, we show that for each selectable Ω', t, $\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}\left[E_l(t,\Omega') - e_l(t,\Omega')\right]\right| = o_P(1)$. Since $\lambda_{0,l}(t) < C_1$ and $M_i(t,l)$ values are not larger than F, by looking at each term in $E_j(t,\Omega')$, we have

$$
\begin{aligned}
&\sup_{1\le l\le\Gamma, 0\le t\le t_0} E\left[\lambda_{0,l}(t)\exp\left(\sum_{1\le i\le n, i\ne j}(\alpha_i+\beta_j)\log(M_i(t,l)+1)\right)\right]^2\\
&\le (C_1)^2 E\left[\exp\left(2\sum_{1\le i\le n, i\ne j}(\alpha_i+\beta_j)\log(M_i(t,l)+1)\right)\right]\\
&\le (C_1)^2 E\left[\exp\left(4C_2\sum_{1\le i\le n, i\ne j}\log(F+1)\right)\right]\\
&\le (C_1)^2\exp\left(4nC_2\log(F+1)\right)
\end{aligned}
$$

Let $C_5 = (C_1)^2\exp\left(4nC_2\log(F+1)\right)$. Then,

$$
\begin{aligned}
&\mathrm{Var}\left[E_l(t,\Omega')\right] = \mathrm{Var}\left[\sum_{j=1}^{n}\lambda_{0,l}(t)\exp\left(\sum_{1\le i\le n, i\ne j}(\alpha_i+\beta_j)\log(M_i(t,l)+1)\right)\right]\\
&\le n\sum_{j=1}^{n} E\left[\lambda_{0,l}(t)\exp\left(\sum_{1\le i\le n, i\ne j}(\alpha_i+\beta_j)\log(M_i(t,l)+1)\right)\right]^2\\
&< n^2 C_5,
\end{aligned}
$$

Now, due to the independency between topics, for any $\epsilon > 0$,

$$
\begin{aligned}
P\left(\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}\left[E_l(t,\Omega') - e_l(t,\Omega')\right]\right| > \epsilon\right) &\le \frac{\sum_{l=1}^{\Gamma}\mathrm{Var}(E_l(t,\Omega'))}{\Gamma^2\epsilon^2}\\
&< \frac{n^2 C_5}{\epsilon^2\Gamma}
\end{aligned}
$$

Step 2. Similarly to the derivation in Step 1,

$$
\begin{aligned}
&\left|\Gamma^{-1}\sum_{l=1}^{\Gamma}\frac{\partial E_l(t,\Omega')}{\partial \alpha_i'}\right| \\
=&\left|\Gamma^{-1}\sum_{l=1}^{\Gamma}\sum_{j=1}^{n}\frac{\partial E_l(t,\Omega')}{\partial \phi_{ij}'}\right| \\
=&\left|\Gamma^{-1}\sum_{l=1}^{\Gamma}\sum_{j=1}^{n}\lambda_{0,l}(t)\log(M_i(t,l)+1)\exp\left(\sum_{i,i\neq j}L_{ij}(t)(\alpha_i'+\beta_j')\log(M_i(t,l)+1)\right)\right| \\
\leq&\Gamma^{-1}\log(F+1)\sum_{l=1}^{\Gamma}\sum_{j=1}^{n}\left|\lambda_{0,l}(t)\exp\left(\sum_{i,i\neq j}L_{ij}(t)(\alpha_i'+\beta_j')\log(M_i(t,l)+1)\right)\right| \\
\leq&C_1\log(F+1)n\exp\left(2nC_2\log(F+1)\right),
\end{aligned}
$$

and similarly we can show,

$$
\left|\Gamma^{-1}\sum_{l=1}^{\Gamma}\frac{\partial E_l(t,\Omega')}{\partial \phi_{ij}'}\right| \leq C_1\log(F+1)n\exp\left(2nC_2\log(F+1)\right)
$$

are bounded by a constant. Let $C_6 = C_1\log(F+1)n\exp\left(2nC_2\log(F+1)\right)$, then

$$
\left\|\Gamma^{-1}\sum_{l=1}^{\Gamma}E_l(t,\Omega') - \Gamma^{-1}\sum_{l=1}^{\Gamma}E_l(t,\Omega'')\right\| \leq C_6\|\Omega'-\Omega''\|_2.
$$

Step 3. In this step, we try to find the appropriate h. For any $t \in [0,t_0]$ and Ω' satisfying Condition B in Theorem 1, let $E^{(1)}_{l,i,M}(t,\Omega') = \frac{\partial E_l(t,\Omega')}{\partial M_i(t,l)}$. For any $h > 0$ such that $t+h \in [0,t_0]$, we can then expand $E_{l,M}(t+h,\Omega')$ at t, at the first order with a continuous derivative as in equation below:

$$
E_{l,M}(t+h,\Omega') = E_{l,M}(t,\Omega') + \sum_{i=1}^{n}E^{(1)}_{l,i,M}(t+\theta h,\Omega')\cdot(M_i(t+h,\Omega')-M_i(t,\Omega')) \quad (2.22)
$$

Similar to our previous derivation, we can show that there exists a constant C_7

that $\sup_{l,i,t,\Omega'} |E^{(1)}_{l,i,M}(t+\theta h,\Omega')'| \leq C_7$. By (2.22), we then have

$$|E_{l,M}(t+h,\Omega') - E_{l,M}(t,\Omega')| \leq \sum_{i=1}^{n} C_7 \cdot (M_i(t+h,\Omega') - M_i(t,\Omega')) \leq nC_7F \quad (2.23)$$

Recall our definition of of $M(t,l)$ that

$$M_i(t,l) = (N_i(t,l) + A_i(t,l))I(N_i(t,l) + A_i(t,l) \leq F) + F \cdot I(N_i(t,l) + A_i(t,l) > F)$$

We have

$$\begin{aligned} P(M_i(t+h,l) - M_i(t,l) \geq 1) &\leq P(N_i(t+h,l) - N_i(t,l) \geq 1) + P\left(A_j(t+h,l) - A_j(t,l) \geq 1\right) \\ &\leq C_8 h, \end{aligned}$$

where $C_8 = 2C_3$. Then

$$\begin{aligned} P\left(\max_{1\leq i\leq n}[M_i(t+h,l) - M_i(t,l)] \geq 1\right) &\leq \sum_{i=1}^{n} P(M_i(t+h,l) - M_i(t,l) \geq 1) \\ &\leq nC_8 h. \end{aligned} \quad (2.24)$$

Now, we look back at $\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t+h,\Omega') - E_l(t,\Omega')]\right|$. If we want this value to be larger than $\epsilon > 0$, by (2.23), we need at least $K_\Gamma \equiv \lfloor \frac{\Gamma\epsilon}{nC_7F} \rfloor$ of the term $E_l(t+h,\Omega') - E_l(t,\Omega')$ to be non-zero, i.e. $\max_{1\leq i\leq n}[M_i(t+h,l) - M_i(t,l)]$ to be non-zero.

Then, when $0 < h = \frac{\epsilon}{2n^2C_8C_7F}$, let t_1, t_2 be (arbitrary) time points in $[t, t+h]$ and $\hat{t}_1(t,h,\Omega')$ $\hat{t}_2(t,h,\Omega')$ to be the pair of values that maximize $\left|\sum_{l=1}^{\Gamma}[E_l(t_1,\Omega') - E_l(t_2,\Omega')]\right|$. We want to mention that this pair of maximizers always exists since on any sample path, there are only a finite number of possible combinations of the M_it, l values. From the combination that maximizes the absolute difference, we can find the corresponding $\hat{t}_1(t,h,\Omega')$ and $\hat{t}_2(t,h,\Omega')$. Then, noticing that (2.23) holds for any Ω', by

(2.24) and the fact that $M_i(t,l)$ is non-decreasing in t,

$$
\begin{aligned}
&P\left(\Gamma^{-1}\sup_{t_1,t_2\in[t,t+h],\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(t_1,\Omega')-E_l(t_2,\Omega')]\right|>\epsilon\right)\\
&=P\left(\Gamma^{-1}\sup_{\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(\hat{t}_1(t,h,\Omega'),\Omega')-E_l(\hat{t}_2(t,h,\Omega'),\Omega')]\right|>\epsilon\right)\\
&\leq\sum_{K=K_\Gamma}^{\Gamma}\sum_{V_K\subset\{1,\ldots,\Gamma\}}\prod_{i\in V_K}P(\max_{1\leq i\leq n}[M_i(\hat{t}_1(t,h,\Omega')+h,l)-M_i(\hat{t}_2(t,h,\Omega'),l)]\geq 1)\\
&\leq\sum_{K=K_\Gamma}^{\Gamma}\sum_{V_K\subset\{1,\ldots,\Gamma\}}\prod_{i\in V_K}P(\max_{1\leq i\leq n}[M_i(t+h,l)-M_i(t,l)]\geq 1)\ \text{(by the nondecreasing property)}\\
&\leq\sum_{K=K_\Gamma}^{\Gamma}\begin{pmatrix}\Gamma\\ K\end{pmatrix}(nC_8h)^K(1-nC_8h)^{\Gamma-K}
\end{aligned}
$$

$\equiv P_{0,\Gamma,\epsilon}$ Let $t_i=\frac{i\cdot h}{2}, 0\leq i\leq\left\lfloor\frac{2t_0}{h}\right\rfloor$. The total time interval $[0,t_0]$ can then be covered by the series of sets, $S_i=[t_i,t_{i+1}], 1\leq i\leq\left\lfloor\frac{2t_0}{h}\right\rfloor-1, S_{\left\lfloor\frac{2t_0}{h}\right\rfloor}=\left[t_{\left\lfloor\frac{2t_0}{h}\right\rfloor},t_0\right]$. Since for any $|t_1-t_2|<h$, the two time points must be contained in the union of two subsequent S_is, we have

$$
\begin{aligned}
&P\left(\Gamma^{-1}\sup_{|t_1-t_2|<h,\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(t_1,\Omega')-E_l(t_2,\Omega')]\right|>\epsilon\right)\\
&=P\left(\Gamma^{-1}\max_{0\leq i\leq\left\lfloor\frac{2t_0}{h}\right\rfloor-1}\sup_{t_1,t_2\in S_i\bigcup S_{i+1},\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(t_1,\Omega')-E_l(t_2,\Omega')]\right|>\epsilon\right)\\
&\leq\sum_{i=0}^{\left\lfloor\frac{2t_0}{h}\right\rfloor-1}P\left(\Gamma^{-1}\sup_{t_1,t_2\in S_i\bigcup S_{i+1},\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(t_1,\Omega')-E_l(t_2,\Omega')]\right|>\epsilon\right)\\
&\leq\left\lfloor\frac{2t_0}{h}\right\rfloor\sum_{K=K_\Gamma}^{\Gamma}\begin{pmatrix}\Gamma\\ K\end{pmatrix}(nC_8h)^K(1-nC_8h)^{\Gamma-K}\equiv P_{\Gamma,\epsilon}
\end{aligned}
\tag{2.25}
$$

$P_{\Gamma,\epsilon}$ can be viewed as

$$\left\lfloor\frac{2t_0}{h}\right\rfloor P\left(\sum_{i=1}^{\Gamma}Y_i\geq K_\Gamma\right),$$

where $Y_i, 1 \le i \le \Gamma$ are i.i.d random variables with Binomial$(1, nC_8h)$. Since $\frac{K_\Gamma}{\Gamma} = 2nC_8h$, and h does no depends on Γ, by the law of large numbers, we can show that $P_{\Gamma,\epsilon}$ converges to zero as $\Gamma \to \infty$.

Step 4. Actually what we have shown in Step 2 and 3 is stronger than what we need. In this step, we show that (2.20) holds.

For any $\epsilon > 0$, for given Ω''' and $t, \in [0, t_0]$, from Step 2 and 3, we can see that for any Ω', Ω'' and t, t' that satisfy $\|\Omega'' - \Omega'\|_\infty < \frac{\epsilon}{C_6}$, $|t'' - t| < h = \frac{\epsilon}{2n^2C_8C_7F}$, we have

$$
\begin{aligned}
&P\left(\Gamma^{-1}\sup_{\Omega',\Omega'',t,t'}\left|\sum_{l=1}^{\Gamma}[E_l(t',\Omega'') - E_l(t,\Omega')]\right| > 2\epsilon\right)\\
\le &P\left(\Gamma^{-1}\sup_{t,t',\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(t',\Omega') - E_l(t,\Omega')]\right| + \Gamma^{-1}\sup_{t,t',\Omega',\Omega''}\left|\sum_{l=1}^{\Gamma}[E_l(t',\Omega'') - E_l(t',\Omega')]\right| > 2\epsilon\right)\\
\le &P\left(\Gamma^{-1}\sup_{t,t',\Omega'}\left|\sum_{l=1}^{\Gamma}[E_l(t',\Omega') - E_l(t,\Omega')]\right| > \epsilon\right)\\
+ &P\left(\Gamma^{-1}\sup_{t,t',\Omega',\Omega''}\left|\sum_{l=1}^{\Gamma}[E_l(t',\Omega'') - E_l(t',\Omega')]\right| > \epsilon\right)\\
\le &P_{\Gamma,\epsilon} + 0 = P_{\Gamma,\epsilon} \to 0, \text{ as } \Gamma \to \infty.
\end{aligned}
$$

Then, due to the pointwise convergence in probability as proved in Step 1, for any Ω'' and t' that satisfy $\|\Omega'' - \Omega'\|_\infty < \frac{\epsilon}{C_6}$, $|t'' - t| < h = \frac{\epsilon}{2n^2C_8C_7F}$, we also have

$$
\Gamma^{-1}\sup_{\Omega'',t'}\left|\sum_{l=1}^{\Gamma}[e_l(t',\Omega'') - e_l(t,\Omega')]\right| < 2\epsilon
$$

Since the choosable set for Ω', $[-C_2, C_2]^n$ and $[0, t_0]$ are all bounded compact, we can have $[-C_2, C_2]^n \times [0, t_0]$ to be covered by a number of n_ϵ open sets $OS_i \equiv \{\Omega', t : \|\Omega' - \Omega_i\|_\infty < \frac{\epsilon}{C_6}, |t - t_i| < h = \frac{\epsilon}{2n^2C_8C_7F}\}, 1 \le i \le n_\epsilon$.

Then for any Ω', t we can always find it to fall in a certain, say OS_i. Now we can have

$$
\begin{aligned}
&\sup_{t\in[0,t_0],|\alpha_i|\le C_2,|\beta_j|\le C_2} \Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-e_l(t,\Omega')]\right| \\
&= \max_{1\le i\le n_\epsilon}\sup_{(\Omega',t)\in OS_i}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-e_l(t,\Omega')]\right| \\
&\le \max_{1\le i\le n_\epsilon}\left(\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t_i,\Omega_i)-e_l(t_i,\Omega_i)]\right| + \sup_{(\Omega',t)\in OS_i}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-E_l(t_i,\Omega_i)]\right|\right) \\
&+ \sup_{(\Omega',t)\in OS_i}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[e_l(t,\Omega')-e_l(t_i,\Omega_i)]\right| \\
&\le \max_{1\le i\le n_\epsilon}\left(\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t_i,\Omega_i)-e_l(t_i,\Omega_i)]\right| + \sup_{(\Omega',t)\in OS_i}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-E_l(t_i,\Omega_i)]\right| + 2\epsilon\right)
\end{aligned}
$$

Then, combing what we have proved in Step 1,

$$
\begin{aligned}
&P\left(\sup_{t\in[0,t_0],|\alpha_i|\le C_2,|\beta_j|\le C_2}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-e_l(t,\Omega')]\right| > 6\epsilon\right) \\
&\le P\left(\max_{1\le i\le n_\epsilon}\left(\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t_i,\Omega_i)-e_l(t_i,\Omega_i)]\right|\right) > 2\epsilon\right) \\
&+ P\left(\max_{1\le i\le n_\epsilon}\left(\sup_{(\Omega',t)\in OS_i}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-E_l(t_i,\Omega_i)]\right|\right) > 2\epsilon\right) \\
&\le \sum_{i=1}^{n_\epsilon} P\left(\left(\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t_i,\Omega_i)-e_l(t_i,\Omega_i)]\right|\right) > 2\epsilon\right) \\
&+ \sum_{i=1}^{n_\epsilon} P\left(\left(\sup_{(\Omega',t)\in OS_i}\Gamma^{-1}\left|\sum_{l=1}^{\Gamma}[E_l(t,\Omega')-E_l(t_i,\Omega_i)]\right|\right) > 2\epsilon\right) \\
&\le \frac{n_\epsilon n^2 C_5}{4\epsilon^2\Gamma} + n_\epsilon P_{\Gamma,\epsilon} \to 0, \text{ as } \Gamma\to\infty.
\end{aligned}
$$

We have proved (2.20).

Step 5. In this step, we show that (2.21) holds.

Since $E_l(t,\Omega'), 1\le l\le\Gamma$ are independent, $\Gamma^{-1}\sum_{l=1}^{\Gamma}|E_l(t,\Omega')-e_l(t,\Omega')| \to_P 0$ is equivalent to $\Gamma^{-1}\sum_{l=1}^{\Gamma}|E_l(t,\Omega')-e_l(t,\Omega')|\to 0$, a.e. Since $\Gamma^{-1}\sum_{l=1}^{\Gamma}E_l(t,\Omega')$ has bounded continuous second order derivatives, we have $\sum_{j=1}^{n}\|\Gamma^{-1}\sum_{l=1}^{\Gamma}[E_{lj}^{(1)}(t,\Omega')-$

$e_{lj}^{(1)}(t,\Omega')]\|_\infty \to 0$, a.e. Similarly, since $E_l(t,\Omega')$ has bounded continuous third order derivatives, we have $\sum_{j=1}^n \|\Gamma^{-1}\sum_{l=1}^{\Gamma}[E_{lj}^{(2)}(t,\Omega') - e_{lj}^{(2)}(t,\Omega')]\|_\infty \to 0$, a.e. Then, similar to the derivations in Step 2, 3 and 4, we can show all the entries of $\Gamma^{-1}\sum_{l=1}^{\Gamma} E_{lj}^{(1)}(t,\Omega')$ and $\Gamma^{-1}\sum_{l=1}^{\Gamma} E_{lj}^{(2)}(t,\Omega')$ have similar properties in Ω' and t. Then similar to Step 4, we can show (2.21).

Proof of Lemma 3

Following Step 5 in the Proof of Lemma 1, considering the existence of all the third order derivatives, it becomes obvious that $e_l(t,\Omega')$, $e_{lj}^{(1)}(t,\Omega')$ and $e_{lj}^{(2)}(t,\Omega')$ are continuous in Ω' and t. Since the selectable sets of $\alpha_i, \beta'_j \in [-C_2, C_2]$ and $t \in [0, t_0]$ are all compact, $e_j(t,\Omega')$, $e_j^{(1)}(t,\Omega')$ and $e_j^{(2)}(t,\Omega')$ are also bounded. Actually, an actual bound can be got following our argument in Step 1 of Lemma proof. At last, since

$$\begin{aligned}
E_l(t,\Omega') &= \sum_{j=1}^n \lambda_{j,l}(t) \\
&= \sum_{j=1}^n \lambda_{0,l}(t)\exp\left(\sum_{i,i\neq j} L_{ij}(t)(\alpha'_i + \beta'_j)\log(M_i(t,l)+1)\right) \\
&\geq \sum_{j=1}^n C_0 \exp\left(-2nC_2\log(F+1)\right) \\
&= nC_0\exp\left(-2nC_2\log(F+1)\right)
\end{aligned}$$

Then, by the properties of almost sure convergence, we can also get

$$e_j(t,\Omega') \geq nC_0\exp\left(-2nC_2\log(F+1)\right).$$

Lemma 3 has been proved.

To prove Theorem 1, we also need the following w lemmas, which are originally the Theorem II.1 and Corollary II.2 of (*Andersen and Gill*, 1982).

Lemma 4. Let E be an open convex subset of $\mathcal{R}^p$ and let $F_1, F_2, \ldots,$ be a sequence of random concave functions on E such that for any $x \in E$, $F_\Gamma(x) \to_P f(x)$ as $n \to \infty$

where f is some non-random function on E. If f is also concave, then for all compact $A \subset E$,

$$\sup_{x \in A} |F_\Gamma(x) - f(x)| \to_P 0, \text{ as } \Gamma \to \infty$$

Lemma 5. Suppose f has a unique maximum at $\hat{x} \in E$. Let $\hat{x}_\Gamma$ maximize F_Γ. Then under the condition of Lemma 4, $\hat{x}_\Gamma \to \hat{x}$ as $n \to \infty$.

Proof of Theorem 1

We prove the Theorem by combining the results in Lemma 1 and Lemma 2, in the following 3 steps.

Step 1

In this step, we first analyze some properties of the counting processes and represent the log-likelihood function by using integrals of counting processes as a preparation. We notice that by stating that $\lambda_j, l(t)$ is the hazard rate of $N_j(t,l)$, we actually have that the processes $K_j(t,l)$ defined by

$$\begin{aligned} K_j(t,l) &= N_j(t,l) - \int_0^t \lambda_{j,l}(t) du \\ &= N_j(t,l) - \int_0^t \lambda_{0,l}(t) \exp\left(\sum_{i, i \neq j} L_{ij}(u)(\alpha'_i + \beta'_j) \log(M_i(u,l)+1) \right) du \end{aligned} \tag{2.26}$$

$j = 1, \ldots, n, t \in [0, t_0]$, are local martingales on the time interval $[0, t_0]$. As a consequence, they are in fact local square integrable martingales, with

$$\langle K_j(\cdot,l), K_j(\cdot,l)\rangle(t) = \int_0^t \lambda_j(u,l) du, \langle K_i(\cdot,l_1), K_j(\cdot,l_2)\rangle = 0, i \neq j \text{ or } l_1 \neq l_2, \tag{2.27}$$

i.e. $K_i(t,l_1)$ and $K_j(t,l_2)$ are orthogonal when $i \neq j$ or $l_1 \neq l_2$.

Let $TN(t,l) = \sum_i N_i(t,l)$, also based on the definition of $N_j(t,l)$, we have

$$LL(\Omega',t) = \sum_{l=1}^{\Gamma}\left[\sum_{j=1}^{n}\int_0^t \sum_{1\leq i\leq n, i\neq j} L_{ij}(\alpha_i' + \beta_j')\log(M_i(u,l)+1)dN_i(u,l) - \right.$$
$$\left.\int_0^t \log\left(\sum_{j=1}^{n}\exp\left(\sum_{i,i\neq j} L_{ij}(u)(\alpha_i' + \beta_j')\log(M_i(u,l)+1)\right)\right) dTN(t,l)\right] \tag{2.28}$$

Now based on (2.28), we consider the process

$$\begin{aligned} X(\Omega',t) &= \Gamma^{-1}(LL(\Omega',t) - LL(\Omega,t)) \\ &= \Gamma^{-1}\left\{\sum_{l=1}^{\Gamma}\int_0^t \sum_{j=1}^{n}\sum_{1\leq i\leq n, i\neq j} L_{ij}(\alpha_i' + \beta_j' - \alpha_i - \beta_j)\log(M_i(u,l)+1)dN_j(u,l)\right. \\ &\left. - \int_0^t \log\left\{\frac{\sum_{j=1}^{n}\exp\left(\sum_{1\leq i\leq n, i\neq j} L_{ij}(\alpha_i' + \beta_j')\log(M_i(u,l)+1)\right)}{\sum_{j=1}^{n}\exp\left(\sum_{1\leq i\leq n, i\neq j} L_{ij}(\alpha_i + \beta_j)\log(M_i(u,l)+1))\right)}\right\} dTN(u,l)\right\} \end{aligned} \tag{2.29}$$

where recall that Γ denotes the number of topics under consideration.

Step 2

By definition, $\hat{\Omega}$ maximizes $X(\Omega',t)$ defined in (2.29). In this step, we find another easier to analyze function to approximate $X(\Omega',t)$. Notice that if we replace $dN_j(t,l)$

with the hazard rates of $N_j(t,l)$, $\lambda_{j,l}(t)$ in (2.29), we can get

$$\begin{aligned}
R(\Omega',t) = \Gamma^{-1} \Bigg\{ & \sum_{l=1}^{\Gamma} \int_0^t \lambda_{0,l}(u) \sum_{j=1}^{n} \sum_{1\le i\le n, i\neq j} L_{ij}(\alpha_i' + \beta_j' - \alpha_i - \beta_j)\log(M_i(u,l)+1) \\
& \cdot \exp\left(\sum_{i\neq j} L_{ij}(t)(\alpha_i+\beta_j)\log(M_i(u,l)+1)\right) du \\
& - \int_0^t \lambda_{0,l}(u) \log\left\{ \frac{\sum_j \exp\left(\sum_{i\neq j} L_{ij}(\alpha_i'+\beta_j')\log(M_i(u,l)+1)\right)}{\sum_j \exp\left(\sum_{i\neq j} L_{ij}(\alpha_i+\beta_j)\log(M_i(u,l)+1)\right)} \right\} \\
& \cdot \left(\sum_{j=1}^{n} \exp\left(\sum_{1\le i\le n, i\neq j} L_{ij}(\alpha_i+\beta_j)\log(M_i(u,l)+1)\right)\right) du \Bigg\} \\
= & \int_0^t LT(\Omega',u)du.
\end{aligned} \tag{2.30}$$

For each Ω', $X(\Omega',t) - R(\Omega',\cdot)$ can be written as sums of $K_j(t,l)$ defined in (2.26). By Theorem 2.4.3 in (*Fleming and Harrington*, 2013), we have

$$< X(\Omega',t) - R(\Omega',t), X(\Omega',t) - R(\Omega',t) >= B(\Omega',t),$$

where

$$B(\Omega',t) = \Gamma^{-2} \sum_{l=1}^{\Gamma} \int_0^t S(u,l)\lambda_{0,l}(u)du,$$

where $S(u,l)$ is given by

$$\begin{aligned}
S(u,l) = \sum_{j=1}^{n} \Bigg(& \sum_{1\le i\le n, i\neq j} L_{ij}(\alpha_i' + \beta_j' - \alpha_i - \beta_j)\log(M_i(u,l)+1) \\
& - \log\left\{ \frac{\sum_{j=1}^{n} \exp\left(\sum_{i,i\neq j} L_{ij}(\alpha_i'+\beta_j')\log(M_i(u,l)+1)\right)}{\sum_{j=1}^{n} \exp\left(\sum_{i,i\neq j} L_{ij}(\alpha_i+\beta_j)\log(M_i(u,l)+1)\right)} \right\} \Bigg)^2
\end{aligned}$$

Let

$$Q_l(\Omega', t) = \int_0^t S(u,l)\lambda_{0,l}(u)du.$$

Then $Q_l, l = 1 \ldots, \Gamma$ are independent. We can write $B(\Omega', t)$ as

$$B(\Omega', t) = \Gamma^{-2} \sum_{l=1}^{\Gamma} Q_l(\Omega', t)$$

Similar to the proof of Lemma 2, since $\alpha_i', \beta_j', \alpha_i, \beta_j, M_i(u,l), 1 \le i, j \le n, u \in [0, t_0]$, we can find a constant C_8, such that $E[(Q_l(\Omega', t))^2] < C_8$. Then we will have

$$\left| \Gamma B(\Omega', t) - \Gamma^{-1} \sum_{l=1}^{\Gamma} E[Q_l(\Omega', t)] \right| \to_P 0.$$

Therefore by the inequality of Lenglart (I.2), we see that $X(\Omega', t)$ should converge in probability to the same limit as $R(\Omega', t)$ for each Ω', when at least one of they converges.

Step 3

In this step, we show $R(\Omega', t)$ converges and analyze the limit function. Note that by using the notations in (3.24), $R(\Omega', t)$ can be simplified to

$$R(\Omega', t) = \int_0^t \Gamma^{-1} \sum_{l=1}^{\Gamma} \left[\sum_{j=1}^{n} (\Phi_j' - \Phi_j)' E_{lj}^{(1)}(u, \Omega) - \log \left\{ \frac{E_l(u, \Omega')}{E_l(u, \Omega)} \right\} E_l(u, \Omega) \right] du$$

It follows that by our assumption, $\lambda_{0,l} < C_1$, and Lemma 1 and Lemma 2, for each Ω', as $\Gamma \to \infty$,

$$|R(\Omega', t_0) - P(\Omega', t_0)| \to_P 0,$$

where

$$P(\Omega', t_0) = \int_0^{t_0} \Gamma^{-1} \sum_{l=1}^{\Gamma} \left[\sum_{j=1}^{n} (\Phi_j' - \Phi_j)' e_{lj}^{(1)}(u, \Omega) - \log \left\{ \frac{e_l(u, \Omega')}{e_l(u, \Omega)} \right\} e_l(u, \Omega) \right] du \tag{2.31}$$

Following our argument in Step 3, equivalently we should have

$$|X(\Omega', t_0) - P(\Omega', t_0)| \to_P 0, \tag{2.32}$$

By Lemma 4, since $X(\Omega', t_0)$ are concave (as will be shown in Lemma 5 below) for all Γ, we have convergence in (2.32) is equivalent to

$$\sup_{\Omega'} |X(\Omega', t_0) - P(\Omega', t_0)| \to_P 0$$

Also as shown in Lemma 4 below, $P(\Omega', t_0)$ is concave and uniquely maximized at $\Omega' = \Omega$. Since by definition, $\hat{\Omega}$ maximizes $X(\Omega', t)$, then by Lemma 5, $\hat{\Omega} \to \Omega$.

Lemma 6. Under the Conditions A, B, C,D of Theorem 1, the $P(\Omega', t_0)$ defined in (2.31) is concave and uniquely maximized at $\Omega' = \Omega$.

Lemma 7. $LT(\Omega', t)$ and $X(\Omega', t)$ are both concave in Ω'.

Proof of Lemma 6:

We establish the concavity of $P(\Omega', t_0)$ and its unique maximizer based the evaluation of its first and second derivative of $P_1(\Omega', t_0)$ to show its convexity. By Lemma 2, we may evaluate the first and second derivatives of $P(\Omega', t_0)$ inside the integral(cf. Bartle, 1966, Corollary5.9). We compute the first derivatives as

$$\frac{\partial P(\Omega', t_0)}{\partial \beta_j'} = \int_0^{t_0} \Gamma^{-1} \sum_{l=1}^{\Gamma} \left[I_n' e_{lj}^{(1)}(u, \Omega) - I_n' e_{lj}^{(1)}(u, \Omega') \frac{e_l(u, \Omega)}{e_l(u, \Omega')} \right] du$$

and

$$\frac{\partial P(\Omega',t_0)}{\partial \alpha_i'} = \int_0^{t_0} \Gamma^{-1} \sum_{l=1}^{\Gamma} \left[\sum_{j=1}^{n} G_i' e_{lj}^{(1)}(u,\Omega) - (\alpha_1',\ldots,\alpha_n')' \sum_j G_i' e_{lj}^{(1)}(u,\Omega') \frac{e_l(u,\Omega)}{e_l(u,\Omega')} \right] du \tag{2.33}$$

where I_n is the $n \times n$-dimensional diagonal matrix and G_i is a n-dimensional vector with all zeros expect one on the i-th entry.

Note that the above parital derivatives are all zero at $\Omega' = \Omega$. Further, the Hessian matrix of $P(\Omega', t_0)$ can be written as

$$\nabla_{\Omega'} \nabla_{\Omega'} P(\Omega', t_0) = -T'DT, \tag{2.34}$$

where $T = \left(\frac{\partial(\Phi_1')'}{\partial \alpha_2'}, \ldots, \frac{\partial(\Phi_1')'}{\partial \alpha_n'}, \ldots, \frac{\partial(\Phi_n')'}{\partial \alpha_2'}, \ldots, \frac{\partial(\Phi_n')'}{\partial \alpha_n'}, \frac{\partial(\Phi_1')'}{\partial \beta_1'}, \ldots, \frac{\partial(\Phi_1')'}{\partial \beta_n'}, \ldots, \frac{\partial(\Phi_n')'}{\partial \beta_1'}, \ldots, \frac{\partial(\Phi_n')'}{\partial \beta_n'} \right)'$ is a matrix of zeros and ones, which describes the linear combination relationship between ϕ_{ij}' and $\alpha_i' + \beta_j'$. Matrix D is a block diagonal matrix of dimension $n^2 \times n^2$, with a number of n block matrix of size $n \times n$, $D_j, 1 \le j \le n$, on the diagonal, where

$$D_j = \int_0^{t_0} \Gamma^{-1} \sum_{l=1}^{\Gamma} T' \left[e_{lj}^{(2)}(u,\Omega) + e_{lj}^{(1)}(u,\Omega)^{\otimes 2} \frac{e_l(u,\Omega)}{e_l(u,\Omega')} \right] du,$$

where "$\otimes 2$" denotes the outer product of a vector. Then the entries of $\nabla_{\Omega'} \nabla_{\Omega'} P(\Omega', t_0)$ will be linear combinations of e_l and $e_{lj}^{(k)}$, $k = 1, 2, 1 \le l \le \Gamma, 1 \le j \le n$.

The exact form of $\nabla_{\Omega'} \nabla_{\Omega'} P(\Omega', t_0)$ in (2.34) may look intimidating. But from the definition of $LT(\Omega', t)$ in (2.8) we can see $LT(\Omega', t)$ and $P(\Omega', t)$ in (2.30) have exactly the same Hessian matrix. Then, again Lemma 1 and 2 implies as $\Gamma \to \infty$,

$$\left\| \int_0^{t_0} \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t) dt - \nabla_{\Omega'} \nabla_{\Omega'} P(\Omega', t_0) \right\|_\infty \to_P 0. \tag{2.35}$$

By Condition D of Theorem 1, we have

$$\lim_{\Gamma\to\infty} P\left(\lambda_{\min}\left(-\nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t_0)|_{\Omega'=\Omega}\right) > C_4\right) = 1$$

By (2.25) we have proved and the continuity of the function $\lambda_{\min}(\cdot)$, we can find a constant δ (not depending on Γ and Ω), such that

$$\lim_{\Gamma\to\infty} P\left(\min_{t\in[t_0-\delta,t_0]} \lambda_{\min}\left(-\nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}\right) > \frac{C_4}{2}\right) = 1 \tag{2.36}$$

We can write

$$\int_0^{t_0} \nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}dt = \int_{t_0-\delta}^{t_0} \nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}dt + \int_0^{t_0-\delta} \nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}dt \tag{2.37}$$

By the definition of integration and the fact that the first integration on the right hand side of the equation above exists, from (2.36), we should also have

$$\lim_{\Gamma\to\infty} P\left(\lambda_{\min}\left(-\int_{t_0-\delta}^{t_0} \nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}dt\right) > \frac{C_4\delta}{2}\right) = 1 \tag{2.38}$$

Also, since $LT(\Omega',t)$ is concave (as shown in Lemma 5), also by the definition of integration and the fact that the second integration on the right hand side of (2.37) exists, we have the following Hessian matrix

$$\int_0^{t_0-\delta} \nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}dt$$

to be at least semi-positive definite. Based on (2.38) Actually, we have already shown that

$$\lim_{\Gamma\to\infty} P\left(\lambda_{\min}\left(-\int_0^{t_0} \nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega',t)|_{\Omega'=\Omega}dt\right) > \frac{C_4\delta}{2}\right) = 1$$

Therefore, combining the result in (2.35), we have

$$\lambda_{\min}\left(-\int_0^{t_0}\nabla_{\Omega'}\nabla_{\Omega'}P(\Omega',t_0)|_{\Omega'=\Omega}\right)\geq\frac{C_4\delta}{2}\tag{2.39}$$

Since $P(\Omega',t_0)$ has zero derivative at $\Omega'=\Omega$, and by (2.39), $P(\Omega',t_0)$ is uniquely maximized at Ω. Lemma 4 has been proved.

Proof of Lemma 7:

Considering the definition of $X(\Omega',t)$ and $LT(\Omega',t)$ as in (2.29) and (2.8), ignoring the terms linear in Ω', we notice that $X(\Omega',t)$ and $LT(\Omega',t)$ are both positively weighted sums (with weights independent of Ω') of

$$LE_l(\Omega',t)\equiv-\log\left[\sum_{j=1}^{n}\exp\left(\sum_{1\leq i\leq n,i\neq j}L_{ij}(\alpha_i'+\beta_j')\log(M_i(u,l)+1)\right)\right].$$

Then, to finish the proof of the lemma, it is equivalent to show the concavity of $LE_l(\Omega',t)$. Let

$$SE_l(\Omega',t)=\sum_{j=1}^{n}\exp\left(\sum_{1\leq i\leq n,i\neq j}L_{ij}(\alpha_i'+\beta_j')\log(M_i(u,l)+1)\right)$$

For any $a,b>0$ and $a+b=1$ and $\Omega_k'=(\alpha_{k,2}',\ldots,\alpha_{k,n}',\beta_{k,1}'\ldots,\beta_{k,n}'),k=1,2,$

satisfying Condition B of Theorem 1, by Jensen's inequality, we have

$$
\begin{aligned}
& SE_l(a\Omega_1' + b\Omega_2', t) \\
& = \sum_{j=1}^{n} \exp\left(\sum_{1\le i\le n, i\ne j} L_{ij}(a\alpha_{1,i}' + a\beta_{1,j}' + b\alpha_{2,i}' + b\beta_{2,j}') \log(M_i(u,l)+1) \right) \\
& \le \left(\sum_{j=1}^{n} \exp\left(\sum_{1\le i\le n, i\ne j} L_{ij}(\alpha_{1,i}' + \beta_{1,j}') \log(M_i(u,l)+1) \right) \right)^{a} \\
& \cdot \left(\sum_{j=1}^{n} \exp\left(\sum_{1\le i\le n, i\ne j} L_{ij}(\alpha_{2,i}' + \beta_{2,j}') \log(M_i(u,l)+1) \right) \right)^{b} \\
& = \left[SE_l(\Omega_1', t)\right]^{a} \left[SE_l(\Omega_2', t)\right]^{b}
\end{aligned}
$$

And the above inequality is just equivalent to

$$
-\log(SE_l(a\Omega_1' + b\Omega_2', t)) \ge -a\log(SE_l(\Omega_1', t)) - b\log(SE_l(\Omega_2', t)).
$$

Lemma 5 has been proved.

2.9 Additional Senator Results

Table 2.6 shows the top ten most influential accounts under the proposed method for different time periods. We see consistent results with the findings from summer 2014. Important newspapers like the Financial Times and Washington Post still appear in the top ten when utilizing the full data. Other prominent accounts include senators that have leadership positions, like Harry Reid (Senate Majority Leader) and several others with high profile committee chairmanships or ranking appointments.

Tables 2.7 and 2.8 show regression results for the sequestration period, and Tables 2.9 and 2.10 show regression results for the inauguration period. The results are consistent with the results presented in the main text. Regressing directly on the leadership scores shows a strongly significant and positive coefficient for the proposed

Table 2.6: Top ten rankings under the proposed model for different time intervals.

Rank	Sequestration	2014 Inauguration	Entire Data
1	Leahy	Leahy	Financial Times
2	Grassley	Grassley	Grassley
3	Mikulski	Begich	Leahy
4	Begich	Mikulski	Cruz
5	Shaheen	Johanns	Washington Post
6	McCaskill	Reid	Reid
7	Reid	McCaskill	Begich
8	Blunt	Graham	Mikulski
9	Graham	Shaheen	Ezra Klein
10	Collins	Hagan	Schatz

Table 2.7: Regression estimates, where the response variable is the raw leadership scores from GovTrack.us and influence is computed from January 1, 2013 to March 1, 2013. $R^2 = 0.327$; $F = 8.839$ on 5 and 92 DF (p-value: 0.000)

Variable	Estimate	Std. Error	t value	$P(>\|t\|)$
Intercept	-0.153	0.228	-0.669	0.505
Proposed Influence	0.074	0.028	2.689	0.009
Republican	-0.153	0.039	-3.960	0.000
Age	0.002	0.003	0.833	0.407
Years in Senate	0.007	0.003	2.532	0.013
Male	0.020	0.050	0.397	0.692

influence measure. The regressions with transformed leadership scores show effects are moderately significant.

Table 2.8: Regression estimates, where the response variable is $log(\frac{\text{leadership}}{1-\text{leadership}})$, where leaderhip is from GovTrack.us and influence is computed from January 1, 2013 to March 1, 2013. $R^2 = 0.119$; $F = 2.466$ on 5 and 92 DF (p-value: 0.038)

Variable	Estimate	Std. Error	t value	$P(>\|t\|)$
Intercept	-3.925	2.574	-1.525	0.131
Proposed Influence	0.504	0.312	1.614	0.110
Republican	-1.103	0.437	-2.526	0.013
Age	0.008	0.029	0.267	0.790
Years in Senate	0.033	0.031	1.062	0.291
Male	0.465	0.561	0.830	0.409

Table 2.9: Regression estimates, where the response variable is the raw leadership scores from GovTrack.us and influence is computed from November 1, 2012 to January 31, 2013. $R^2 = 0.328$; $F = 8.839$ on 5 and 92 DF (p-value: 0.000)

Variable	Estimate	Std. Error	t value	$P(>\|t\|)$
Intercept	-0.132	0.220	-0.597	0.552
Proposed Influence	0.072	0.026	2.726	0.008
Republican	-0.154	0.039	-3.978	0.000
Age	0.002	0.003	0.797	0.427
Years in Senate	0.007	0.003	2.616	0.010
Male	0.020	0.050	0.395	0.693

Table 2.10: Regression estimates, where the response variable is $log(\frac{\text{leadership}}{1-\text{leadership}})$, where leaderhip is from GovTrack.us and influence is computed from November 1, 2012 to January 31, 2013. $R^2 = 0.117$; $F = 2.402$ on 5 and 92 DF (p-value: 0.043)

Variable	Estimate	Std. Error	t value	$P(>\|t\|)$
Intercept	-3.578	2.495	-1.434	0.155
Proposed Influence	0.452	0.297	1.521	0.132
Republican	-1.105	0.437	-2.527	0.013
Age	0.007	0.029	0.255	0.800
Years in Senate	0.035	0.031	1.113	0.269
Male	0.460	0.561	0.819	0.415

CHAPTER III

Measuring Topic Dependent Edge Importance in Twitter Ecosystems Using a Counting Process Modeling Framework

3.1 Introduction

Over the past two decades, the functional properties of edges in complex networks have gained much attention in literature, (*Miritello et al.*, 2011) (*Dorogovtsev and Mendes*, 2002) (*Newman*, 2003). The importance (weights) of edges, a key factor in determining structural properties of a network has also been fairly extensively studied. For example, in (*Tong et al.*, 2012), an edge's importance is measured by its topology strength in the connection network, while in (*Toivonen et al.*, 2007), it is modeled by the short time probability that a node can send messages to other nodes in the network. In our proposed framework, as presented in the previous chapter, we take into consideration the actions occurring from the nodes (such as posting, retweeting and mentioning) on *multiple edges simultaneously.* To capture the temporal evolution of actions, we model them as a counting process.

However, we allow for a much more flexible parameterization than the one used in the previous chapter. Instead of having two global parameters for each node, reflecting capability to generate responses (α) and susceptibility to respond to other nodes' actions (β), we allow for independent parameters between every pair of nodes for *selected topics*. With this extension we are trying to incorporate the heterogeneous nature of topics discussed and the fact that for specific topics, selected nodes have either greater susceptibility due to their particular interest, or greater capability for generating responses, due to their perceived expertise. For example, if we considered a general Twitter network, it is reasonable to assume that sport fans would engage in a different manner when the topic under discussion involves sports (and even more so, if it involves their favorite sport or favorite sport team) or the node posting is a sports-writer or an athlete and hence it is generally perceived that (s)he has expertise or additional information.

The remainder of the chapter is organized as follows: in Section 3.2, we introduce the counting process modeling framework and the proposed edge importance measure. Section 3.3 presents the computational algorithm we use to obtain the parameter estimates, as well as establish their statistical properties and those of the influence measure in Section 3.4. In Section 3.5, we use simulation studies to evaluate the performance of the model, while the US Senate application is presented in Section 3.6. Finally, a short summary is given in Section 3.7.

3.2 The model and the influence measure

Leveraging the model developed in the previous chapter, we have similar definitions of some key quantities for the network under consideration. We continue to represent the followers network as $G = (V, L)$, where V corresponds to the set of nodes of all the Twitter accounts under consideration and $L = \{L_{i,j}, 1 \leq i \neq j \leq n\}$ the edge set between them. This network establishes *potential channels of communication* between accounts, since if an account follows another, then they can actively interact. In principle, L can be dynamically evolving, but in this work we continue to consider L to be static and not changing over time, which is a reasonable assumption for periods of time extending to months. We continue to consider the following three actions: posting a new message, retweeting a message posted by another account that they follow and finally mentioning another account that they follow in a new posted message.

Next, we define the following two key counting processes. Still we use $N_j(t)$ to denote the total number of retweets and mentions that account j generates on topic l by time t and let $A_j(t,l)$ denote the total number of posted messages by account j on topic l by time t. Denote the set of topics discussed by Γ. In the model presented in the previous chapter, all topics were treated identically; thus, the cardinality of the set Γ corresponds to the total "sample size". However, in practice different topics elicit different responses and/or interactions for different sets of users, as briefly discussed in the introductory section. In principle, the set Γ can be partitioned into several groups. However, for the sake of simplicity we consider two groups; namely, $\Gamma = \Gamma 1 \bigcup \Gamma_2$. We assume that an account has different long-term interaction parameters with its followers on all topic sets. On topic set Γ_1, we use α_{ij} to denote account i's interactions with its followers $j \in L$, while on set Γ_2, we assume the interaction to be $\alpha_{ij} + \delta_i$. The δ_i parameter captures the *differential* capability of account i to elicit responses from all the other accounts on the followers' network for the subset of topics in Γ_2.

We continue to adopt the previously described modeling strategy and model $\{N_j(t,l)\}_{i=1}^{n}$ as a set of counting processes through their hazard rates, using a version of Cox (*Andersen and Gill*, 1982) proportional hazard model; specifically, the hazard rate $\lambda_{j,l}(t)$ of process $N_j(t,l)$ is given by

$$\lambda_{j,l}(t) = \lambda_{0,l}(t) \exp\left(\sum_{i \neq j} L_{ij}(\alpha_{ij}) \log(M_i(t,l)+1)\right), \tag{3.1}$$

when $l \in \Gamma_1$ and

$$\lambda_{j,l}(t) = \lambda_{0,l}(t) \exp\left(\sum_{i \neq j} L_{ij}(\alpha_{ij} + \delta_i) \log\left(M_i(t,l) + 1\right)\right), \tag{3.2}$$

when $l \in \Gamma_2$, where

$$M_i(t,l) = (N_i(t,l) + A_i(t,l)) I(N_i(t,l) + A_i(t,l) \leq F) + F \cdot I(N_i(t,l) + A_i(t,l) > F).$$

$A_j(t,l) + N_j(t,l)$ is the total number of postings, retweets and mentions for account j on topic l by time t. We also consider the effect of seeing actions from account i can get saturated when the total number of actions reaches the constraint, F. We assume that the parameters $\alpha_{ij}, \delta_i \in (-\infty, \infty)$, since accounts and their users may be positively or negatively inclined towards other accounts, and the accounts' capability can vary in a differential manner between the two topic sets. The nonparametric baseline component $\lambda_{0,l}(t)$ is time varying. In general, due to limited time span, we assume all the observations are made within $[0, t_0]$, since as observed in real data sets, interest in a particular topic wanes fairly quickly. The model posits that account j interacts with other accounts at a baseline level $\lambda_{0,l}(t)$, modulated by its (different) ability to generate responses from its followers (in the two topic sets). Note that we continue to model the retweet-mention process $N_j(t,l)$, since it reflects interactions between nodes and use the total effective activity process $M_j(t,l)$ as a covariate.

To complete the modeling framework, let $\mathcal{T}_j^l = \left\{T_{j,1}^l, \ldots, T_{j,n_j^l}^l\right\}, t = 1, \cdots, n_j^l$, denote the set of time points that account j took action (post, retweet, mention) on topic l, until the last observation time point t_0. Finally, for identification purposes, we require $\alpha_{ij} = 0$ when $L_{ij} = 0$, i.e., when account j does not follow i. Use $\Psi = (\alpha_{ij})_{1 \leq i,j \leq n}$ to denote the α matrix and Δ to represent $(\delta_1, \ldots, \delta_n)$. Then the parameters of interest are $\Omega = \{\Psi \cdot L, \Delta\}$, where "$\cdot$" denotes the point-wise multiplication of two matrices. As before and following, *Andersen and Gill* (1982), we employ a partial-likelihood function to obtain estimates of Ω. Specifically, we treat the baseline $\lambda_{0,l}(t)$ as a nuisance parameter and decomposing the full-likelihood to obtain

$$PL_1(t) = \prod_{1\le l\le \Gamma} \left(\prod_{1\le j\le n} \prod_{1\le k\le n_j^l, T_{j,k}^l \le t} \frac{\lambda_j(T_{j,k}^l)}{\sum_{1\le i\le n} \lambda_i(T_{j,k}^l)} \right)$$

Plugging the exact form of the hazard rate from (3.1) and (3.2) into the partial-likelihood function (PL), we get:

$$\begin{aligned} PL_1(t) = & \prod_{l\in\Gamma_1} \left(\prod_{1\le j\le n} \prod_{1\le k\le n_j, T_{j,k}^l \le t} \frac{\exp\left(\sum_{i\ne j} L_{ij}\alpha_{ij} \log\left(M_i(T_{j,k}^l, l) + 1\right)\right)}{\sum_{1\le i\le n} \exp\left(\sum_{u\ne i} L_{ui}\alpha_{ui} \log\left(M_u(T_{j,k}^l, l) + 1\right)\right)} \right) \\ + & \prod_{l\in\Gamma_2} \left(\prod_{1\le j\le n} \prod_{1\le k\le n_j, T_{j,k}^l \le t} \frac{\exp\left(\sum_{i\ne j} L_{ij}(\alpha_{ij} + \delta_i) \log\left(M_i(T_{j,k}^l, l) + 1\right)\right)}{\sum_{1\le i\le n} \exp\left(\sum_{u\ne i} L_{ui}(\alpha_{ui} + \delta_u) \log\left(M_u(T_{j,k}^l, l) + 1\right)\right)} \right) \end{aligned} \tag{3.3}$$

3.2.1 The Edge Importance Measure

Next, we define the edge importance measure, leveraging the structure of the new model. Specifically, for an edge (i, j) with $L_{ij} = 1$, when $\lambda_{0,l}(t) = 1$, and $M_i(t, l) = 1$, other $M_k(t, l) = 0, t \in [0, t_0], k \ne i, j$, account j's hazard rate within unit time $[0, 1]$ on topic set Γ_1 is given by

$$H_j^{(1)} = \exp\left(\log(2)\alpha_{ij}\right),$$

and its hazard rate on topic set Γ_2 is

$$H_j^{(2)} = \exp\left(\log(2)(\alpha_{ij} + \delta_i)\right).$$

From the properties of the exponential distribution, the probability that account j will retweet i, within $[0,1]$ on topic set Γ_1 is

$$P_{ij}^{(1)} = 1 - \exp\left[-\exp\left(\log(2)\alpha_{ij}\right)\right],$$

and the probability that account j will retweet i, within $[0,1]$ on topic set Γ_2 is

$$P_{ij}^{(2)} = 1 - \exp\left[-\exp\left(\log(2)(\alpha_{ij} + \delta_i)\right)\right].$$

Following our definition of accounts' influences in the previous chapter, the influence of j on topic set Γ_1 is given by

$$\Xi_j^{(1)} = \log\left[\sum_{1 \leq k \leq n, k \neq j} L_{jk} \cdot \exp\left(\log(\bar{M}_j^{(1)} + 1)\alpha_{jk}\right)\right], \tag{3.4}$$

and its influence on topic set Γ_2 can be written as

$$\Xi_j^{(2)} = \log\left[\sum_{1 \leq k \leq n, k \neq j} L_{jk} \cdot \exp\left(\log(\bar{M}_j^{(2)} + 1)(\alpha_{jk} + \delta_j)\right)\right], \tag{3.5}$$

where $\bar{M}_j^{(1)}$ and $\bar{M}_j^{(2)}$ are node j's average number of actions on topic set Γ_1 and Γ_2 respectively. Therefore, on topic set Γ_1, on edge (i,j), the influence i can borrow from j with a single action, can be expressed as

$$\begin{aligned} \varsigma_{ij}^{(1)} &= P_{ij}^{(1)} \cdot \Xi_j^{(1)} \\ &= \left(1 - \exp\left[-t_0 \exp\left(\log(2)\alpha_{ij}\right)\right]\right) \log\left[\sum_{1 \leq k \leq n, k \neq j} L_{jk} \cdot \exp\left(\log(\bar{M}_j^{(1)} + 1)\alpha_{jk}\right)\right], \end{aligned} \tag{3.6}$$

and on topic set Γ_2, the influence i that can be borrowed can be given by

$$\begin{aligned}\varsigma_{ij}^{(2)} &= P_{ij}^{(2)} \cdot \Xi_j^{(2)} \\ &= (1-\exp\left[-t_0 \exp\left(\log(2)(\alpha_{ij}+\delta_i)\right)\right]) \log\left[\sum_{1\le k\le n, k\ne j} L_{jk}\cdot \exp\left(\log(\bar{M}_j^{(2)}+1)(\alpha_{jk}+\delta_j)\right)\right].\end{aligned} \tag{3.7}$$

We define $\varsigma_{ij}^{(1)}$ and $\varsigma_{ij}^{(2)}$ to be edge (i,j)'s importance on topic set Γ_1 and Γ_2, respectively. When $L_{ij}=0$, let $\varsigma_{ij}^{(1)} = \varsigma_{ij}^{(2)} = 0$.

3.2.2 The set of influential edges

In this subsection, we use our definition of edge importance to capture the edges that are essential for propagating or impairing the information flow process. Let $S_L^{(c)} = \{\varsigma_{ij}^{(c)} : L_{ij} = 1\}, c = 1,2,$ denote all the edge importance values on existing edges, on the two topic sets. For any given probability p, we can build the following two edge collections, which correspond to the p proportion of the most spawning and jamming edges as follows.

$$S_1(p)^{(c)} = \{(i,j) : \varsigma_{ij}^{()} \ge q_{1-p}^{(c)}\} \tag{3.8}$$

and

$$S_2(p)^{(c)} = \{(i,j) : \varsigma_{ij} \le q_p^c\} \tag{3.9}$$

where q_p^c are the quantile of set $S_L^{(c)}$ of probability p, $c = 1,2$.

3.3 Computation and Inference

Next, we present a Newton-type algorithm for computing the parameter estimates Ω. The logarithm of the partial likelihood function (3.3) is given by

$$
\begin{aligned}
&LL_0(t) = \log(PL(t)) \\
&= \sum_{l\in\Gamma_1}\Bigg\{\sum_{1\le j\le n}\ \sum_{1\le k\le n_j^l, T_{j,k}^l\le t}\ \sum_{i\ne j} L_{ij}\alpha_{ij}\log(M_i(T_{j,k}^l,l)+1) \\
&\quad - \sum_{1\le j\le n}\ \sum_{1\le k\le n_j^l, T_{j,k}^l\le t}\log\left[\sum_{1\le i\le n}\exp\left(\sum_{u\ne i} L_{ui}\alpha_{ui}\log(M_u(T_{j,k}^l,l)+1)\right)\right]\Bigg\} \\
&+ \sum_{l\in\Gamma_2}\Bigg\{\sum_{1\le j\le n}\ \sum_{1\le k\le n_j^l, T_{j,k}^l\le t}\ \sum_{i\ne j} L_{ij}(\alpha_{ij}+\delta_i)\log(M_i(T_{j,k}^l,l)+1) \\
&\quad - \sum_{1\le j\le n}\ \sum_{1\le k\le n_j^l, T_{j,k}^l\le t}\log\left[\sum_{1\le i\le n}\exp\left(\sum_{u\ne i} L_{ui}(\alpha_{ui}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)\right]\Bigg\}
\end{aligned}
\tag{3.10}
$$

To introduce sparsity on the differences of influence levels among topics, we add an L_1 penalty on Δ to obtain the following penalized partial likelihood function.

$$
\begin{aligned}
&LL(t) = \log(PL(t)) - \sum_{i=1}^{n} \gamma|\delta_i| \\
&= \sum_{l\in\Gamma_1} \Bigg\{ \sum_{1\le j\le n} \sum_{1\le k\le n_j^l, T_{j,k}^l \le t} \sum_{i\ne j} L_{ij}\alpha_{ij} \log(M_i(T_{j,k}^l, l)+1) \\
&\qquad - \sum_{1\le j\le n} \sum_{1\le k\le n_j^l, T_{j,k}^l \le t} \log\left[\sum_{1\le i\le n} \exp\left(\sum_{u\ne i} L_{ui}\alpha_{ui} \log(M_u(T_{j,k}^l, l)+1)\right)\right] \\
&+ \sum_{l\in\Gamma_2} \Bigg\{ \sum_{1\le j\le n} \sum_{1\le k\le n_j^l, T_{j,k}^l \le t} \sum_{i\ne j} L_{ij}(\alpha_{ij}+\delta_i) \log(M_i(T_{j,k}^l, l)+1) \\
&\qquad - \sum_{1\le j\le n} \sum_{1\le k\le n_j^l, T_{j,k}^l \le t} \log\left[\sum_{1\le i\le n} \exp\left(\sum_{u\ne i} L_{ui}(\alpha_{ui}+\delta_u) \log(M_u(T_{j,k}^l, l)+1)\right)\right] \\
&\qquad - \sum_{i=1}^{n} \gamma|\delta_i|
\end{aligned}
\tag{3.11}
$$

The objective function corresponds to $LL(t_0)$, which considers all events k in its equation (3.11). For the sake of notational simplicity, we will use LL to represent $LL(t_0)$ in the rest of the paper. To maximize $LL(t_0)$, due to the smoothness of $LL_0(t_0)$, we first compute the gradient $G_0 \equiv \nabla_\Omega LL_0(t_0)$ and the Hessian $H \equiv \nabla_\Omega \nabla_\Omega (LL_0(t_0))$ of $LL_0(t_0)$. Then on the last n entries of G_0, which is represented as G_n, update it to $G_n - \gamma \text{sign}(\Delta)$. Denote the updated G_0 as G. G and H are used as the approximate gradient and Hessian of $LL(t_0)$ in Newton's algorithm. The detailed expressions for G_0 and the Hessian H are given in the Appendix.

The steps of the optimization, with a given penalty γ in computing G, are given in Algorithm 1. To speed up calculations, we take advantage of the structure of the problem in actual applications, as explained in detail in the Appendix. Then γ is selected with the one that minimize the estimation of the parameter. As stated in the

Algorithm 3 Estimating the parameters by Newton's algorithm

Initialize Ω value by $\alpha_{ij} = 0, \delta_i = 0, 1 \leq i, j \leq n$
Define s as a positive thresholding constant for the minimum step size
while $\tau > s$ **do**
  Calculate G by using (3.16), (3.20) and (3.22)
  Calculate H by using (3.18) to (3.21)
  Find the optimum positive τ value such that $\Omega - \tau \cdot H^{-1}G$ will maximize the log-partial-likelihood (3.11), when H is non-singular and find the optimum positive τ value such that $\Omega - \tau \cdot (H - \theta I)^{-1}G$ will maximize the log-partial-likelihood (3.11), otherwise.
  Update $\Omega \leftarrow \Omega - \tau \cdot H^{-1}G$ or $\Omega - \tau \cdot (H - \theta I)^{-1}G$, depending on H's singularity
end while
return Ω

algorithm, s, θ are positive constants to judge the convergence of the the Newton's algorithm and to solve singularity problems when computing the inverse matrices. The computational complexity of this algorithm is dominated by the computation of H. Denote by $m_n = \max_{1 \leq j \leq n}\{n_j\}$. Based on (3.16), (3.20) and (3.22), it costs $O(\Gamma n^2 m_n)$ operations to calculate an entry of G. Further, since G is of dimension $2n$, it takes $O(\Gamma n^3 m_n)$ to obtain the entire G vector. Analogously, based on (3.18) to (3.21), it costs $O(\Gamma n m_n)$ operations to calculate an entry of H, if proper book-keeping is used on the results obtained for the gradient G. Further, since H is of dimension n^4, it takes $O(\Gamma n^5 m_n)$ to obtain the entire H matrix. Hence, the overall time complexity for each iteration of the algorithm is of the order $O(\max\{\Gamma n^5 m_n\})$.

3.4 Properties of the $\hat{\Omega}$ estimates

Next, we establish that the estimator $\hat{\Omega}$ which maximizes (3.11) will converge to the true parameter $\Omega = \{\Psi \cdot L, \Delta\}$ in probability under certain regularity conditions.

Theorem 2. Conditions:

A. (Bounded hazard rate) $C_0 \leq \lambda_{0,l}(t) \leq C_1$ for $0 \leq t \leq t_0$ $1 \leq l \leq \Gamma$,

B. (Bounded parameters) $\max_{1 \leq i,j \leq n}\{|\alpha_{ij}|, |\delta_i|\} \leq C_2$,

C. (Limited posting frequencies)

$$P\left(A_j(t+h,l) - A_j(t,l) \geq 1\right) \leq C_3 \cdot h, \text{ when } h > 0, t+h \leq t_0, \tag{3.12}$$

D. (Balanced sets sizes) $\max\left\{\frac{\|\Gamma_1\|_0}{\|\Gamma_2\|_0}, \frac{\|\Gamma_2\|_0}{\|\Gamma_1\|_0}\right\} < C_4$

E. (Positive definite limit of Hessian) Let Ω' be any choosable parameter vector satisfying (B). For large enough Γ and some C_5, we have the holding condition to hold on the smallest eigenvalue of $-\nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega', t)$, at $\Omega' = \Omega, t = t_0$,

$$P\left(\lambda_{\min}\left(-\nabla_{\Omega'}\nabla_{\Omega'}LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0} > C_5\right) \to 1, \text{ as } \|\Gamma_1\|_0 \to \infty,$$

where

$$\begin{aligned} LT(\Omega', t) \equiv (\|\Gamma_1\|_0)^{-1} &\left\{\sum_{l\in\Gamma_1} -\lambda_{0,l}(u) \log\left\{\sum_j \exp\left(\sum_{i\neq j} L_{ij}\alpha'_{ij} \log(M_i(t,l)+1)\right)\right\}\right. \\ &\left.\cdot \left(\sum_{j=1}^n \exp\left(\sum_{1\leq i\leq n, i\neq j} L_{ij}\alpha_{ij} \log(M_i(t,l)+1)\right)\right)\right\} \\ +(\|\Gamma_1\|_0)^{-1} &\left\{\sum_{l\in\Gamma_1} -\lambda_{0,l}(u) \log\left\{\sum_j \exp\left(\sum_{i\neq j} L_{ij}(\alpha'_{ij}+\delta'_i) \log(M_i(t,l)+1)\right)\right\}\right. \\ &\left.\cdot \left(\sum_{j=1}^n \exp\left(\sum_{1\leq i\leq n, i\neq j} L_{ij}(\alpha_{ij}+\delta_i) \log(M_i(t,l)+1)\right)\right)\right\} \end{aligned} \tag{3.13}$$

where $\|\cdot\|_0$ computes the L_0 (size) of a set.

In the five Conditions A through E above, C_0, C_1, C_2, C_3, C_4 and C_5 are all positive constants. Under these conditions, we will have:

$$\hat{\Omega} \to_P \Omega \text{ as } \|\Gamma_1\|_0, \|\Gamma_2\|_0 \to \infty.$$

The detailed proof is given in the Appendix.

Since Condition C in Theorem 2 is the same as the one in Theorem 1, Lemma 1 in Section 2.4 already gives one example in which this condition will hold. Based on Theorem 2, by leveraging the properties of continuous functions, we can establish the consistency of the proposed edge importance measure.

Proposition 2. Let $\varsigma^{(1)} = \left(\varsigma_{ij}^{(1)}\right)_{1\le i,j\le n}$ and $\varsigma^{(2)} = \left(\varsigma_{ij}^{(2)}\right)_{1\le i,j\le n}$ denote the $n \times n$ dimensional matrix of edge importance. Further, denote by $\hat{\varsigma}^{(1)} = \left(\hat{\varsigma}_{ij}^{(1)}\right)_{1\le i,j\le n}$ and $\hat{\varsigma}^{(2)} = \left(\hat{\varsigma}_{ij}^{(2)}\right)_{1\le i,j\le n}$ their empirical estimates. Under the conditions of Theorem 1, we have that when $\|\Gamma_1\|_0, \|\Gamma_2\|_0 \to \infty$

$$\left\|\hat{\varsigma}^{(1)} - \varsigma^{(1)}\right\|_2 + \left\|\hat{\varsigma}^{(2)} - \varsigma^{(2)}\right\|_2 \to_P 0, \tag{3.14}$$

where $\|\cdot\|_2$ computes the L_2 norm. As a result, with any probability $p \in (0,1)$, when $\left(\varsigma_{ij}^{(1)}\right)_{1\le i,j\le n}$ and $\left(\varsigma_{ij}^{(2)}\right)_{1\le i,j\le n}$ values are all distinct, we have

$$\|\hat{S}_1(p)^{(c)} \backslash S_1(p)^{(c)}\|_0 \to_P 0,$$

where $\|\cdot\|_0$ is the L_0 norm of a set.

From Theorem 2, the proof of the proposition is straightforward, since each element of the matrix $\hat{\varsigma}^{(1)}$ and $\hat{\varsigma}^{(1)}$ is a continuous function of $\hat{\Omega}$.

3.5 Performance evaluation

The key steps for obtaining the synthetic data are identical to those in the previous chapter; namely, generating the followers' network and generating actions for the two sets of topics Γ_1 and Γ_2, respectively.

As before, we first illustrate the performance of the Newton estimation algorithm, on a random network of varying size. We set the parameter $a = 0.5$ for the baseline hazard rate and choose a time horizon of $t_0 = 10$, to emulate ten days worth of data.

We also select the parameters Ω uniformly at random in the interval $[-0.4, 0.4]$. Due to the bounded baseline hazard rate and simulated parameters, and since the retweets and mentions are generated as Poisson, Condition A, B, C of Theorem 1 have been satisfied. Then we empirically "check" Condition D. With a large $\|\Gamma_1\|_0 = \|\Gamma_2\|_0 = 500$, network size $n = 10, 30$, we repeated Step 1 and 2 for 20 time s to simulate the network and actions. In each repetition, the square root of the smallest eigenvalue of $\lambda_{\min}\left(-\Gamma \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0}$ is computed. The results are plotted in Figure 3.1. In the plot, it can be seen that smallest eigenvalues of

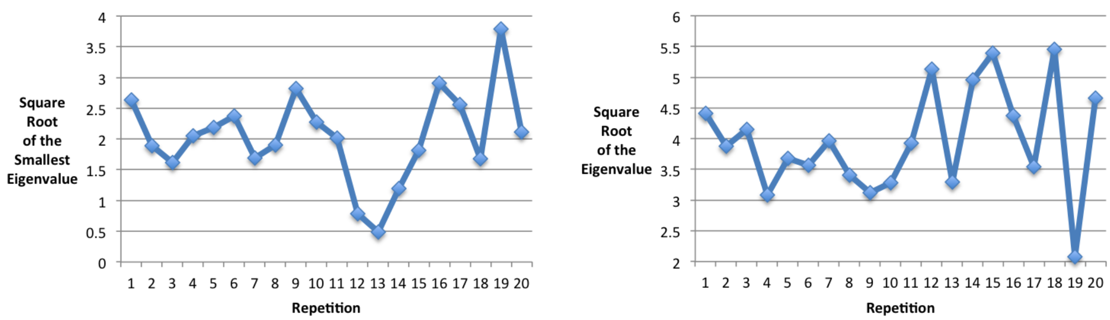

Figure 3.1: $[\lambda_{\min}\left(-\Gamma \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0}]^{1/2}$ at $\|\Gamma_1\|_0 = \|\Gamma_2\|_0 = 500$, $n = 10$ (left) and $n = 30$ (right).

$[\lambda_{\min}\left(-\Gamma \nabla_{\Omega'} \nabla_{\Omega'} LT(\Omega', t)\right)|_{\Omega'=\Omega, t=t_0}]^{1/2}$ are generally larger or not smaller than 0.5. Due to their large variations, we took square root to make the values of the smallest eigenvalues easier to be reflected in plots. Due to their large variations, we took square root to make the values of the smallest eigenvalues easier to be reflected in plots.

Then as we have "checked", with network sizes $n = 10, 30$ and number of topics generated $\|\Gamma_1\|_0 = \|\Gamma_2\|_0 = 500$ and another set of smaller topic sizes, and penalty γ, we obtain three sets of values to estimate the relative error of the parameter and importance estimates, $\frac{\|\hat{\Omega}-\Omega\|_2}{\|\Omega\|_2}$, $\frac{\|\hat{\varsigma}^{(1)}-\varsigma^{(1)}\|_2}{\|\varsigma^{(1)}\|}$ and $\frac{\|\hat{\varsigma}^{(2)}-\varsigma^{(2)}\|_2}{\|\varsigma^{(2)}\|}$. At $n = 10$, with $\gamma \in \{1, 2 \ldots, 50\}$, and $\|\Gamma_1\|_0 = \|\Gamma_2\|_0 = 500$, we first observed that $\frac{\|\hat{\Omega}-\Omega\|_2}{\|\Omega\|_2}$ was minimized around $\gamma = 10$ and actually the errors only had small differences, based on the average of five replicates for each chosen γ. Due to the high time complexity of the computation algorithm, we then only applied it to the network with $n = 30$ under $\gamma = 0, 10$ and 50. We obtain the following Figure 3.2 to show the relative error of the parameter and importance estimates

The results are based on 20 replicates of the underlying followers networks, as well as the actions (postings, retweets and mentions) data. It can be seen in Figure 3.2 that with large enough topic sets, $\|\Gamma_1\| = \|\Gamma_2\| = 500$, the parameters and importance measures, exhibit a small (less than 10%) relative error rate, where $\|\cdot\|_0$ corresponds to ℓ_0 norm.

Based on the estimated parameters $\hat{\alpha}_{ij}$ of the simulations, we examine more closely the estimation results in the setting with $\|\Gamma_1\|_0 = \|\Gamma_2\|_0 = 500$. We estimate the important edge collections $\hat{S}_k(p)^{(c)}$ as defined in (3.8) and (3.9), $c, k = 1, 2$. Then, we check the performance of $\hat{S}_k(p)^{(c)}$ by looking at the proportions of edges in the estimated sets, that are also coherent with the original ones:

$$\frac{\|\hat{S}_k(p)^{(c)} \bigcap S_k(p)^{(c)}\|_0}{\|\hat{S}_k(p)^{(c)}\|_0},$$

with $c, k = 1, 2$. The estimation results are shown in Figure 3.3. Sinceour interest mainly focus on capturing the most spawning and jamming edges, we may only plot with small p values.

From Figure 3.3, we can see that, generally $\hat{S}_k(p)^{(c)}$ estimates $S_k^{(c)}(p)$ well with its edges coinciding with those of $S_k(p)^{(c)}$ with an accuracy of more than 80%, $c, k = 1, 2$. Since the two sets (at the same p) contain almost the same number of edges, the rate in Figure 3.3 can also be looked at as the proportion of edges in $S_k^{(c)}(p)$ that are captured by $\hat{S}_k(p)^{(c)}$. To include most of the extreme edges $S_k^{(c)}(p)$, we propose to use $\hat{S}_k(p)^{(c)}$. And at the same time, we hope the other edges (not in $S_k^{(c)}(p)$) are still worth considering in the sense that they are included in $S_k(4p)^{(c)}$, $c, k = 1, 2$. The

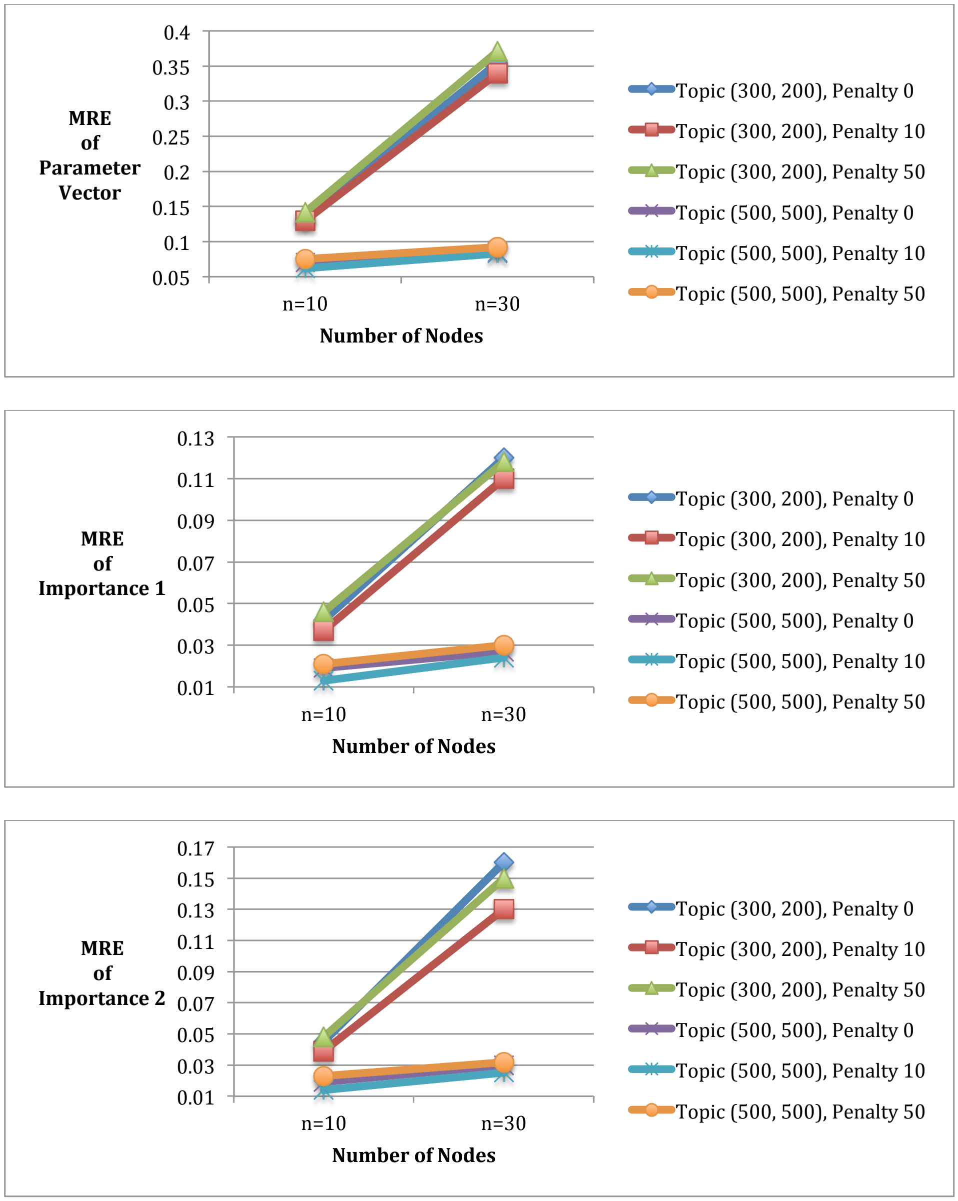

Figure 3.2: Mean relative error of the model parameter estimates Ω (up), $\varsigma^{(1)}$(middle) and $\varsigma^{(2)}$ (down)

Figure 3.3: Proportion of correct edges, from up to down: n=10, S_1 (first); n=10,S_2(Second); n=30, S_1 (Third); n=30, S_2(fourth).

results of this idea are depicted in Figure 3.4 below. In the figure, we computed

$$\frac{\|\hat{S}_k(2p)^{(c)} \bigcap S_k(p)^{(c)}\|_0}{\|S_k(p)^{(c)}\|_0},$$

the coverage for real p proportional most important edges and

$$\frac{\|\hat{S}_k(2p)^{(c)} \bigcap S_k(4p)^{(c)}\|_0}{\|\hat{S}_k(2p)^{(c)}\|_0},$$

the accuracy rate, with $c, k = 1, 2$ for the network $n = 30$.

As can be seen from Figure 3.4, at small probability p values, generally more than 90% of edges in $\hat{S}_k(2p)^{(c)}$ can be captured in $S_k(4p)^{(c)}$. With the exception of $p = 0.01$, around (or more than) 90% of edges in $S_k^{(c)}(p)$ have been included in $\hat{S}_k(2p)^{(c)}$.

3.6 Identifying Important Connections between Senators

Next, we re-examine our Tweeter data. Recall that there are about 200,000 tweets and 4671 follower links within the set of 120 accounts from April 2009 to July 2014. The recorded accounts are registered to 55 Democratic politicians (U.S. Senators and the President of the U.S.), 46 Republican Senators, 2 government organizations (U.S. Army and the Federal Reserve Board), and 16 media outlets, including newspapers (Financial Times, Washington Post, New York Times, Huffington Post), television networks (MSNBC, Fox News, CNN, CSPAN), reporters (Nate Silver (538), Ezra Klein) and television hosts (Bill O'Reilly, Sean Hannity).

In the previous chapter, the focus was on identifying the most influential senators. This time, we are interested in exploring the importance of edges (connections) in the followers network. Also, as both in Table 1 in the previous chapter and in *Golbeck et al.* (2010), it is understood that senators tend to retweet and mention as a means of self or legislative promotion. We have increased frequencies of data during

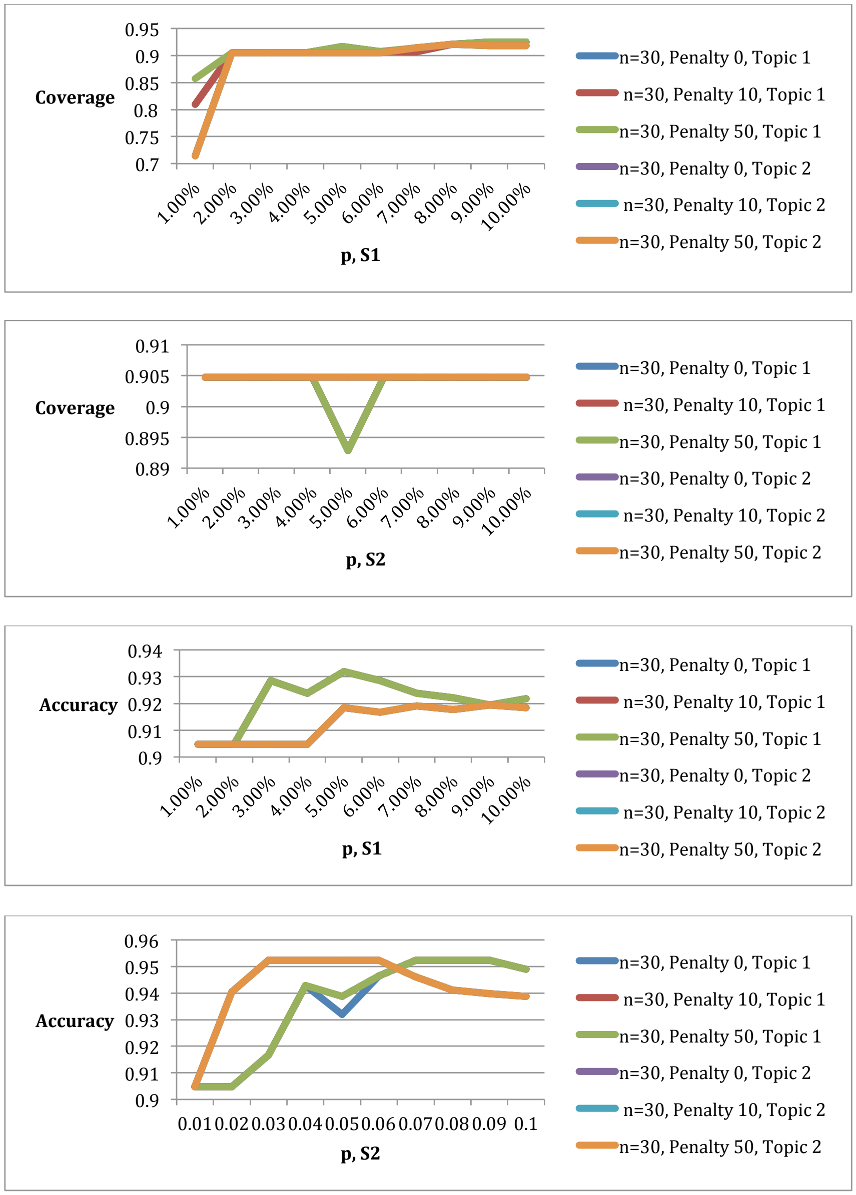

Figure 3.4: Coverage and accuracy, from up to down: coverage of S_1 (first); coverage of S_2(Second); accuracy of S_1 (Third); accuracy of S_2(fourth).

some periods of hot topics, such as in the months surrounding the inauguration of President Obama (January 2013), the debate on raising the debt ceiling of the US government and its temporary suspension around April 2013 and the summer of 2014 (soccer World Cup). We would also like to understand whether the edges' importances are different or not within and outside of these time periods. Given the high computational complexity of the algorithm and the overall high volume of data, we focus on the summer of 2014 (soccer World Cup)

Observations collected from April to June, 2014 form Group 2 of the data, and others form Group 1. Topics sets Γ_1 and Γ_2 then correspond to these two time periods. Although the 2014 summer period is relatively short in duration, it nevertheless contains 55455 actions, which accounts for 40.88% of the tweets count. Given the fast moving landscape of social media, new topics are assigned each week. Combined with pre-assigned topic grouping based on key words, we get 2770 topics in total for the entire data set.

Although not the main focus of this chapter, recall from (3.4) and (3.5), we can still compute the nodes' influences under our proposed model, for sets Γ_1 and Γ_2. To rigorously justify our modeling of the data, we perform a regression analysis to assess how well our influence measure can explain *legislative leadership* in Congress, by comparing to regression results applied with PageRank, on the followers networks (which constitutes the backbone of many ranking algorithms of Twitter accounts). Our response variable is the leadership score, published by *www.govtrack.us* (*GovTrack.us*, 2014). GovTrack creates the leadership score by applying the PageRank algorithm to the adjacency matrix of bill cosponsorship data. Thus, the leadership score for each senator is a number between 0 and 1, where higher values denote greater legislative leadership. The regression model we are interested in is

$$\text{Leadership} = \beta \text{Influence} + \Theta \text{Controls}, \tag{3.15}$$

where Influence contains the two influences computed on Γ_1 and Γ_2, from our proposed model and/or PageRank, and Controls includes party affiliation, gender, age, and number of years in the senate. Seniority endows a number of benefits including preferential assignment to committees. Thus, these control variables likely associate strongly with legislative leadership.

To estimate the parameters in our proposed model and the two influence measures, after preprocessing to get the topics, we apply Algorithm 3 to estimate the α and δ parameters using all the data. The final influence measures, $\hat{\Xi}^{(1)}$ and $\hat{\Xi}^{(2)}$, are constructed by using the average $M_i(\mathcal{T}_m, l)$, at all time points of a retweet or mention happens, in Group 1 and Group 2, respectively.

Since the leadership score provided by GovTrack takes values between 0 and 1, we estimate two models. One model uses the raw leadership scores, and another uses $\log(\frac{\text{leadership}}{1-\text{leadership}})$ for the response variable. In both cases, as shown in Table 3.1, we consistently find that the 2 newly proposed influence measures explains more variation in leadership than our original proposed measure, and PageRank. This observation may serve to suggest that the new more flexible model is more suitable, at the cost of higher computational complexity due to proliferation of parameters from $2n$ for the original models to n^2 for the new model.

In the Senators' social network, we are interested to find the most spawning and jamming edges. As suggested by Figure 3.4, if we use 40 edges with largest (smallest) importances to capture the true 20 most extreme (largest or smallest) edges, we should have enough accuracy. Our estimated 40 edges with the largest importances, on topic Γ_1, are listed in Table 3.2. The list of the edges with the smallest edge importances on the same topic set is given in Table 3.3. Similarly, the 40 estimated edges with the largest importances, on topic Γ_2, are listed in Table 3.4, while those with the smallest edge importances on the second topic set are listed in Table 3.5.

A summary of the main findings is given next:

Table 3.1: Estimated R-squared values for different regression models, where the two new proposed influence measures, the original proposed measure or PageRank is included in the set of independent variables and the influence is computed for the entire data sample. We consistently find that the 2 newly proposed measure is a better indicator of legislative importance.

Response	New Influence on Γ_1	New Influence on Γ_2	Original Proposed	PageRank	R^2
leadership	V				0.363
		V			0.353
			V		0.311
				V	0.276
$\log(\frac{\text{leadership}}{1-\text{leadership}})$	V				0.361
		V			0.350
			V		0.114
				V	0.098

- There is a great deal of agreement between the most important edges for both topics sets. This should be expected due to the fact that the two sets are not separated by a careful topics selection. Obviously, Γ_2 contains many more discussions related to the World Cup, but on the other hand this are not contentious issues that may produce disparate results.

- The results indicate that Jon Tester (Senator from Montana) is least influential, which is consistent with his joining the Twitter platform in March 2012 and overall being a low content producer (740 total tweets since April 15, 2015). His Twitter activity should be juxtaposed with a prolific user like Cory Booker (Senator from New Jersey) who has sent over 47.5K in less than 7 years and a moderate user like John McCain (Senator from Arizona) who has tweeted around 8.5K times since the beginning of 2009.

- It is worth noting that John McCain features prominently in both topics sets, given his foreign policy expertise and immigration views. These discussion topics feature prominently in both sets.

- In general, the strongest influences are between more "senior" Senators (e.g.

Table 3.2: Top forty edges with largest proposed edge importance values from April 16, 2009 - April 30, 2014 (topic set Γ_1).

Rank	Followed	Follower	Rank	Followed	Follower
1	Mike Johanns	Roy Blunt	21	Jeff Flake	Mike Lee
2	The O'Reilly Factor	John McCain	22	Rob Portman	John McCain
3	Brian Schatz	Ron Wyden	23	US Army	John McCain
4	Mike Johanns	Mike Lee	24	Sean Hannity	John McCain
5	Claire McCaskill	John McCain	25	Sheldon Whitehouse	Chuck Schumer
6	Al Franken	Chuck Schumer	26	Bill Nelson	Mike Lee
7	Mike Johanns	Dan Coats	27	Richard Shelby	Lamar Alexander
8	Carl Levin	John McCain	28	Brian Schatz	Kirsten Gillibrand
9	Lindsey Graham	John McCain	29	Mike Johanns	Mark Kirk
10	Bill Nelson	Bob Menendez	30	Susan Collins	John McCain
11	Marco Rubio	John McCain	31	Jerry Moran	Mark Warner
12	John Walsh	Harry Reid	32	Elizabeth Warren	Tom Coburn
13	Jefferson Sessions	Mike Lee	33	Tim Johnson	Jeanne Shaheen
14	Mike Johanns	Mark Begich	34	Mark Begich	Tom Carper
15	Michael F. Bennet	Jay Rockefeller	35	John Barrasso	John McCain
16	Brian Schatz	Dean Heller	36	John Boozman	Mike Lee
17	Al Franken	Barbara Boxer	37	Bill Nelson	Mark Udall
18	Michael F. Bennet	Mark Udall	38	C-SPAN	John McCain
19	Mike Johanns	Rand Paul	39	Debbie Stabenow	Chuck Schumer
20	Dan Coats	John McCain	40	Lisa Murkowski	Mark Begich

Reid, Senate Majority Leader at the time, Levin -36 years in office and Chairman of the powerful Armed Services Committee-, Collins -18 years in office and ranking member of the power Committee on Appropriations, Sessions - 18 years in office and ranking member of the influential Committee on the Judiciary, and so forth).

3.7 Summary

In this chapter, we have proposed a novel measure of the edge importances in large social platform, by considering the amount of influence an account can "borrow" from the follower the edge connects to, with a single *action*. The method is based on using counting processes with exponential hazard rates, to model the time sequences

Table 3.3: Top forty edges with smallest proposed edge importance values from April 16, 2009 - April 30, 2014 (topic set Γ_1).

Rank	Followed	Follower	Rank	Followed	Follower
1	Debbie Stabenow	Jon Tester	21	Kay Hagan	Jon Tester
2	Michael F. Bennet	Jon Tester	22	Chris Coons	Jon Tester
3	Mark Begich	Jon Tester	23	Chuck Schumer	Jon Tester
4	Claire McCaskill	Jon Tester	24	Bernie Sanders	Jon Tester
5	Mike Johanns	Jon Tester	25	Barbara Mikulski	Jon Tester
6	Al Franken	Jon Tester	26	Richard Blumenthal	Jon Tester
7	Chuck Grassley	Jon Tester	27	Bob Menendez	Jon Tester
8	Tim Johnson	Jon Tester	28	Mark Warner	Jon Tester
9	Jay Rockefeller	Jon Tester	29	Heidi Heitkamp	Jon Tester
10	Mary Landrieu	Jon Tester	30	Joe Manchin	Jon Tester
11	Jeanne Shaheen	Jon Tester	31	WSJ	Jon Tester
12	Jerry Moran	Jon Tester	32	Ben Cardin	Jon Tester
13	Barbara Boxer	Jon Tester	33	Bob Casey	Jon Tester
14	Tom Harkin	Jon Tester	34	Chris Murphy	Jon Tester
15	Mark Pryor	Jon Tester	35	Kirsten Gillibrand	Jon Tester
16	Ezra Klein	Jon Tester	36	Dianne Feinstein	Jon Tester
17	Sheldon Whitehouse	Jon Tester	37	Tom Udall	Jon Tester
18	Harry Reid	Jon Tester	38	Dick Durbin	Jon Tester
19	Mark Udall	Jon Tester	39	Mike Crapo	Jon Tester
20	Carl Levin	Jon Tester	40	Patrick Leahy	Jon Tester

of the *actions* users take on the platform. In the hazard rates, we use independent parameters to model the long term capability for an account to bring on an *action* from each of its followers and the parameters can be different on separate topic sets. The parameters are then estimated by maximizing the log of a partial likelihood function, with lasso penalty included to introduce sparsity on the differences between topic sets. With the estimated parameters, for each edge, we then compute the probability for a follower on the other end of the edge, to take an *action* due to the action the followed account takes, within our observation time. The importance of the edge is then estimated by the probability multiplied by the influence of the follower, which is computed following our influence definition in our previous project. Applications of our new model to the US senators data shows the larger flexibility in the hazard rate model illustrate superior performance on explaining Senators' leadership scores

Table 3.4: Top forty edges with largest proposed edge importance values from May 1, 2014 - July 31, 2014 (topic set Γ_2).

Rank	Followed	Follower	Rank	Followed	Follower
1	Mike Johanns	Roy Blunt	21	Dan Coats	John McCain
2	The O'Reilly Factor	John McCain	22	Rob Portman	John McCain
3	Brian Schatz	Ron Wyden	23	US Army	John McCain
4	Claire McCaskill	John McCain	24	Jeff Flake	Mike Lee
5	Mike Johanns	Mike Lee	25	Sheldon Whitehouse	Chuck Schumer
6	Al Franken	Chuck Schumer	26	Sean Hannity	John McCain
7	Mike Johanns	Dan Coats	27	Brian Schatz	Kirsten Gillibrand
8	Carl Levin	John McCain	28	Richard Shelby	Lamar Alexander
9	Lindsey Graham	John McCain	29	Jerry Moran	Mark Warner
10	Bill Nelson	Bob Menendez	30	Mike Johanns	Mark Kirk
11	John Walsh	Harry Reid	31	Bill Nelson	Mike Lee
12	Marco Rubio	John McCain	32	Susan Collins	John McCain
13	Tim Johnson	Jeanne Shaheen	33	Mark Begich	Tom Carper
14	Jefferson Sessions	Mike Lee	34	Claire McCaskill	Barbara Mikulski
15	Mike Johanns	Mark Begich	35	Elizabeth Warren	Tom Coburn
16	Michael F. Bennet	Mark Udall	36	Bill Nelson	Mark Udall
17	Michael F. Bennet	Jay Rockefeller	37	John Barrasso	John McCain
18	Mike Johanns	Rand Paul	38	Barbara Mikulski	Ben Cardin
19	Brian Schatz	Dean Heller	39	Debbie Stabenow	Chuck Schumer
20	Al Franken	Barbara Boxer	40	C-SPAN	John McCain

in real life. And the estimated edge importances are consistent with our observation.

Recalling the content of Chapter 2, from the computation complexity and application examples given in this thesis, it can be seen that our proposed models in the two chapters are most useful when looking at an small scale ecosystem of related users like the US Senators. They may also be useful for getting better insights into the influence of subsets of users in a bigger network. But Due to scalability issues, they are not yet appropriate to analyze huge network like the entire Twitter space.

Table 3.5: Top forty edges with smallest proposed edge importance values from May 1, 2014 - July 31, 2014 (topic set Γ_2).

Rank	Followed	Follower	Rank	Followed	Follower
1	Debbie Stabenow	Jon Tester	21	Chris Coons	Jon Tester
2	Michael F. Bennet	Jon Tester	22	Chuck Schumer	Jon Tester
3	Mark Begich	Jon Tester	23	Tim Johnson	Jon Tester
4	Claire McCaskill	Jon Tester	24	Bernie Sanders	Jon Tester
5	Mike Johanns	Jon Tester	25	Richard Blumenthal	Jon Tester
6	Al Franken	Jon Tester	26	Bob Menendez	Jon Tester
7	Chuck Grassley	Jon Tester	27	Mark Warner	Jon Tester
8	Jay Rockefeller	Jon Tester	28	Heidi Heitkamp	Jon Tester
9	Mary Landrieu	Jon Tester	29	Joe Manchin	Jon Tester
10	Jeanne Shaheen	Jon Tester	30	WSJ	Jon Tester
11	Jerry Moran	Jon Tester	31	Ben Cardin	Jon Tester
12	Barbara Boxer	Jon Tester	32	Bob Casey	Jon Tester
13	Tom Harkin	Jon Tester	33	Chris Murphy	Jon Tester
14	Mark Pryor	Jon Tester	34	Kirsten Gillibrand	Jon Tester
15	Ezra Klein	Jon Tester	35	Dianne Feinstein	Jon Tester
16	Sheldon Whitehouse	Jon Tester	36	Tom Udall	Jon Tester
17	Harry Reid	Jon Tester	37	Dick Durbin	Jon Tester
18	Mark Udall	Jon Tester	38	Patrick Leahy	Jon Tester
19	Carl Levin	Jon Tester	39	Patty Murray	Jon Tester
20	Kay Hagan	Jon Tester	40	Bob Corker	Jon Tester

3.8 Estimation Algorithm and Proofs

3.8.1 Computation equations for Newton's update to maximize LL

Note that $LL = LL_0(t_0) - \gamma \sum_{i=1} |\delta_i|$. To simplify notations, we will use LL_0 to represent $LL_0(t_0)$ in the rest of the paper. We will first give expressions for the gradient vector and Hessian matrix of the LL_0 function and modify them to maximize LL. Some rather straightforward algebra yields the following expressions for the elements of the gradient vector $G_0 \equiv \nabla_\Omega LL_0$:

$$
\begin{aligned}
&\frac{\partial LL_0}{\partial \alpha_{ij_1}} \\
&= \sum_{l\in\Gamma_1} \Bigg\{ \sum_{1\le k\le n^l_{j_1}} L_{ij_1} \log(M_i(T^l_{j_1,k}, l)+1) \\
&-\sum_{j=1}^{n} \sum_{1\le k\le n^l_j} \frac{L_{ij_1}\log(M_i(T^l_{j,k}, l)+1)\exp\left(\sum_{u\neq j_1} L_{uj_1}\alpha_{ij}\log(M_u(T^l_{j,k}, l)+1)\right)}{\sum_{1\le v\le n}\exp\left(\sum_{u\neq v} L_{uv}\alpha_{uv}\log(M_u(T^l_{j,k}, l)+1)\right)} \Bigg\} \\
&+\sum_{l\in\Gamma_2} \Bigg\{ \sum_{1\le k\le n^l_{j_1}} L_{ij_1} \log(M_i(T^l_{j_1,k}, l)+1) \\
&-\sum_{j=1}^{n} \sum_{1\le k\le n^l_j} \frac{L_{ij_1}\log(M_i(T^l_{j,k}, l)+1)\exp\left(\sum_{u\neq j} L_{uj_1}(\alpha_{uj}+\delta_u)\log(M_u(T^l_{j,k}, l)+1)\right)}{\sum_{1\le v\le n}\exp\left(\sum_{u\neq v} L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T^l_{j,k}, l)+1)\right)} \Bigg\}
\end{aligned}
\tag{3.16}
$$

for $1 \le i, j_1 \le n, i \neq j_1$, and

$$
\begin{aligned}
\frac{\partial LL_0}{\partial \delta_i} &= \sum_{l\in\Gamma_2} \Bigg\{ \sum_{j=1}^{n} \sum_{1\le k\le n^l_j} L_{ij}\log(M_i(T^l_{j,k}, l)+1) \\
&-\sum_{j=1}^{n} \sum_{1\le k\le n^l_j} \frac{\sum_{v\neq i} L_{iv}\log(M_i(T^l_{j,k}, l)+1)}{\sum_{1\le v\le n}\exp\left(\sum_{u\neq v} L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T^l_{j,k}, l)+1)\right)} \\
&\cdot \exp\left(\sum_{u\neq v} L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T^l_{j,k}, l)+1)\right)\Bigg\}
\end{aligned}
\tag{3.17}
$$

for $1 \le i \le n$.

Next, we obtain the necessary expressions for the Hessian matrix $H(LL_0)$. We start by computing the sub-matrix of H that includes the second partial derivatives of LL_0 with respect to the α parameters and obtain

$$
\begin{aligned}
\frac{\partial^2 LL_0}{\partial \alpha_{i_1 j_1} \partial \alpha_{i_2 j_1}} = \sum_{l \in \Gamma_1} \Bigg\{ & -\sum_{j=1}^{n} \sum_{1 \le k \le n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_1} \log(M_{i_1}(T_{j,k}^l, l) + 1) \log(M_{i_2}(T_{j,k}^l, l) + 1)}{\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv} \alpha_{uv} \log(M_u(T_{j,k}^l, l) + 1)\right)} \\
& \cdot \exp\left(\sum_{u \ne v} L_{uj_1}(T_{j,k}^l) \alpha_{uj_1} \log(M_u(T_{j,k}^l, l) + 1)\right) \\
& + \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_1} \log(M_{i_1}(T_{j,k}^l, l) + 1) \log(M_{i_2}(T_{j,k}^l, l) + 1)}{\left[\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv} \alpha_{uv} \log(M_u(T_{j,k}^l, l) + 1)\right)\right]^2} \\
& \cdot \exp\left(2 \sum_{u \ne j_1} L_{uj_1} \alpha_{uj_1} \log(M_u(T_{j,k}^l, l) + 1)\right) \Bigg\} \\
+ \sum_{l \in \Gamma_2} \Bigg\{ & -\sum_{j=1}^{n} \sum_{1 \le k \le n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_1} \log(M_{i_1}(T_{j,k}^l, l) + 1) \log(M_{i_2}(T_{j,k}^l, l) + 1)}{\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv} \alpha_{uv} \log(M_u(T_{j,k}^l, l) + 1)\right)} \\
& \cdot \exp\left(\sum_{u \ne v} L_{uj_1}(T_{j,k}^l) (\alpha_{uj_1} + \delta_u) \log(M_u(T_{j,k}^l, l) + 1)\right) \\
& + \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_1} \log(M_{i_1}(T_{j,k}^l, l) + 1) \log(M_{i_2}(T_{j,k}^l, l) + 1)}{\left[\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv} (\alpha_{uv} + \delta_u) \log(M_u(T_{j,k}^l, l) + 1)\right)\right]^2} \\
& \cdot \exp\left(2 \sum_{u \ne j_1} L_{uj_1} (\alpha_{uj_1} + \delta_u) \log(M_u(T_{j,k}^l, l) + 1)\right) \Bigg\}
\end{aligned} \tag{3.18}
$$

when $1 \le i_1, j_1, i_2 \le n, i_1 \ne j_1, i_2 \ne j_1$, and also

$$
\begin{aligned}
&\frac{\partial^2 LL_0}{\partial \alpha_{i_1 j_1} \partial \alpha_{i_2 j_2}} \\
&= \sum_{l \in \Gamma_1} \left\{ \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_2} \log(M_{i_1}(T_{j,k}^l, l) + 1) \log(M_{i_2}(T_{j,k}^l, l) + 1)}{\left[\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv} \alpha_{uv} \log(M_u(T_{j,k}^l, l) + 1)\right)\right]^2} \right. \\
&\qquad \left. \cdot \exp\left(\sum_{u \ne j_1} L_{u j_1} \alpha_{u j_1} \log(M_u(T_{j,k}^l, l) + 1)\right) \exp\left(\sum_{u \ne j_2} L_{u j_2} \alpha_{u j_2} \log(M_u(T_{j,k}^l, l) + 1)\right) \right\} \\
&+ \sum_{l \in \Gamma_2} \left\{ \sum_{1 \le j \le n} \sum_{1 \le k \le n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_2} \log(M_{i_1}(T_{j,k}^l, l) + 1) \log(M_{i_2}(T_{j,k}^l, l) + 1)}{\left[\sum_{1 \le v \le n} \exp\left(\sum_{u \ne v} L_{uv} (\alpha_{uv} + \delta_u) \log(M_u(T_{j,k}^l, l) + 1)\right)\right]^2} \right. \\
&\qquad \cdot \exp\left(2 \sum_{u \ne j_1} L_{u j_1} (\alpha_{u j_1} + \delta_u) \log(M_u(T_{j,k}^l, l) + 1)\right) \\
&\qquad \left. \cdot \exp\left(2 \sum_{u \ne j_2} L_{u j_2} (\alpha_{u j_2} + \delta_u) \log(M_u(T_{j,k}^l, l) + 1)\right) \right\}
\end{aligned}
\tag{3.19}
$$

when $1 \le i_1, j_1, i_2 \le n, i_1 \ne j_1, i_2 \ne j_1$ and especially $j_1 \ne j_2$.

Next, we obtain the sub-matrix of H that includes the second partial derivatives of LL with respect to the δ parameters and get

$$\begin{aligned}\frac{\partial^2 LL_0}{\partial\delta_{i_1}\partial\delta_{i_2}} = \sum_{l\in\Gamma_2}\Bigg\{ &-\sum_{1\le j\le n}\sum_{1\le k\le n_j^l}\frac{\sum_{v\neq i_1,i_2} L_{i_1v}L_{i_2v}\log(M_{i_1}(T_{j,k}^l,l)+1)\log(M_{i_2}(T_{j,k}^l,l)+1)}{\sum_{1\le v\le n}\exp\left(\sum_{u\neq v}L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)} \\ &\cdot\exp\left(\sum_{u\neq v}L_{uv}(T_{j,k}^l)(\alpha_{uj_1}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right) \\ &+\sum_{1\le j\le n}\sum_{1\le k\le n_j^l}\frac{\sum_{v\neq i_1} L_{iv}\log(M_{i_1}(T_{j,k}^l,l)+1)}{\left[\sum_{1\le v\le n}\exp\left(\sum_{u\neq v}L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)\right]^2} \\ &\cdot\exp\left(\sum_{u\neq v}L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right) \\ &\cdot\sum_{v\neq i_2}L_{iv}\log(M_{i_1}(T_{j,k}^l,l)+1)\exp\left(\sum_{u\neq v}L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)\Bigg\},\end{aligned} \tag{3.20}$$

when $1 \le i_1, i_2 \le n$.

Finally, we provide expressions for the cross-partials

$$\begin{aligned}&\frac{\partial^2 LL_0}{\partial\alpha_{i_1,j_1}\partial\delta_{i_2}} \\ &= \sum_{l\in\Gamma_2}\Bigg\{-\sum_{1\le j\le n}\sum_{1\le k\le n_j^l}\frac{L_{i_1j_1}L_{i_2j_1}\log(M_{i_1}(T_{j,k}^l,l)+1)\log(M_{i_2}(T_{j,k}^l,l)+1)}{\sum_{1\le v\le n}\exp\left(\sum_{u\neq j_1}L_{uj_1}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)} \\ &\cdot\exp\left(\sum_{u\neq j_1}L_{uj_1}(T_{j,k}^l)(\alpha_{uj_1}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right) \\ &+\sum_{1\le j\le n}\sum_{1\le k\le n_j^l}\frac{L_{i_1j_1}\log(M_{i_1}(T_{j,k}^l,l)+1)\exp\left(\sum_{u\neq j_1}L_{uj_1}(\alpha_{uj_1}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)}{\left[\sum_{1\le v\le n}\exp\left(\sum_{u\neq v}L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)\right]^2} \\ &\cdot\sum_{v\neq i_2}L_{i_2v}\log(M_{i_1}(T_{j,k}^l,l)+1)\exp\left(\sum_{u\neq v}L_{uv}(\alpha_{uv}+\delta_u)\log(M_u(T_{j,k}^l,l)+1)\right)\Bigg\},\end{aligned} \tag{3.21}$$

when $1 \le i_1, i_2, j_1 \le n, i_1 \neq j_1$.

After getting G_0 and H, we let

$$G = G_0 - \gamma \text{sign}(\Delta) \tag{3.22}$$

and use G and H in Newton's update. The function sign$(\cdot)$ gives a (vector) of the signs of the values.

3.8.2 Implementation Issues

As outlined above, the maximum likelihood estimator is also obtained by Newton's algorithm and detailed expressions for the respective gradient and Hessian are given in the previous subsection. However, luckily enough, the structure of the problem again allows us to precompute and store several quantities for repeated use, thus saving on computational time in practice. Note that the data containing the actions are stored according to their time stamps. We again start by computing four groups of quantities introduced by an action, labeled respectively by indices j, l, k and possibly some other parameters, where j indicates the node that takes the activity, l is the topic label and k represents the relative sequence number of the action, in all the actions that node j has taken under topic l.

First, we define

$$E^{(1)}_{j,v,k,l} = \exp\left(\sum_{u \neq v} L_{uv} \alpha_{uv} \log(M_u(T^l_{j,k}, l) + 1)\right).$$

and

$$E^{(2)}_{j,v,k,l} = \exp\left(\sum_{u \neq v} L_{uv} (\alpha_{uv} + \delta_u) \log(M_u(T^l_{j,k}, l) + 1)\right).$$

We also define

$$O^{(c)}_{j,i,k,l} = L_{ij} \log(M_i(T^l_{j,k}, l) + 1) E^{(c)}_{j,v,k,l}, c = 1, 2.$$

Then, we compute

$$SE^{(c)}_{j,k,l} = \sum_{1\le v\le n} E^{(c)}_{j,v,k,l},$$

and

$$ME^{(c)}_{j,i,k,l} = \sum_{v\neq i} L_{iv}\log(M_i(T^l_{j,k}, l) + 1)E^{(c)}_{j,v,k},$$

$c = 1, 2$. Also, we have

$$LM_{j,s,k,l} = \sum_{u\neq j} L_{uj}\log(M_u(T^l_{s,k}, l) + 1).$$

Then, based on the precomputed components values, the elements of the gradient vector $G \equiv \nabla_\Omega LL$ are obtained as follows:

$$\frac{\partial LL_0}{\partial \alpha_{ij_1}} = \sum_{l\in\Gamma_1}\left\{\sum_{1\le k\le n^l_{j_1}} L_{ij_1}\log(M_i(T^l_{j_1,k}, l) + 1) - \sum_{1\le j\le n}\sum_{1\le k\le n^l_j}\frac{O^{(1)}_{j_1,i,k,l}}{SE^{(1)}_{j,k,l}}\right\}$$
$$\sum_{l\in\Gamma_2}\left\{\sum_{1\le k\le n^l_{j_1}} L_{ij_1}\log(M_i(T^l_{j_1,k}, l) + 1) - \sum_{1\le j\le n}\sum_{1\le k\le n^l_j}\frac{O^{(2)}_{j,i,k,l}}{SE^{(2)}_{j,k,l}}\right\}$$

for $2 \le i \le n$, and

$$\frac{\partial LL_0}{\partial \delta_i} = \sum_{l\in\Gamma_1}\left\{\sum_{1\le j\le n, j\neq i}\sum_{1\le k\le n^l_j} L_{ij}\log(M_i(T^l_{j,k}, l) + 1) - \sum_{1\le j\le n}\sum_{1\le k\le n^l_j}\frac{ME^{(1)}_{j,i,k,l}}{SE^{(1)}_{j,k,l}}\right\}$$
$$+\sum_{l\in\Gamma_2}\left\{\sum_{1\le j\le n, j\neq i}\sum_{1\le k\le n^l_j} L_{ij}\log(M_i(T^l_{j,k}, l) + 1) - \sum_{1\le j\le n}\sum_{1\le k\le n^l_j}\frac{ME^{(2)}_{j,i,k,l}}{SE^{(2)}_{j,k,l}}\right\}$$

for $1 \le i, j_1 \le n, i \neq j_1$.

Regarding the Hessian, based on the four precomputed groups of quantities, we start by computing the sub-matrix of H that includes the second partial derivatives of

LL with respect to the α parameters. We get when $1 \leq i_1, j_1, i_2 \leq n, i_1 \neq j_1, i_2 \neq j_1$,

$$\frac{\partial^2 LL_0}{\partial \alpha_{i_1 j_1} \partial \alpha_{i_2 j_1}} = \sum_{l \in \Gamma_1} \left\{ - \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(1)}_{j_1,i_1,k,l} O^{(1)}_{j_1,i_2,k,l}}{SE^{(1)}_{j,k,l}} + \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(1)}_{j_1,i_1,k,l} O^{(1)}_{j_1,i_2,k,l}}{\left(SE^{1}_{j,k,l}\right)^2} \right\}$$
$$+ \sum_{l \in \Gamma_2} \left\{ - \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(2)}_{j_1,i_1,k,l} O^{(2)}_{j_1,i_2,k,l}}{SE^{(1)}_{j,k,l}} + \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(2)}_{j_1,i_1,k,l} O^{(2)}_{j_1,i_2,k,l}}{\left(SE^{2}_{j,k,l}\right)^2} \right\}$$

When $1 \leq i_1, j_1, i_2 \leq n, i_1 \neq j_1, i_2 \neq j_1$ and especially $j_1 \neq j_2$, we similarly have

$$\frac{\partial^2 LL_0}{\partial \alpha_{i_1 j_1} \partial \alpha_{i_2 j_2}} = \sum_{l \in \Gamma_1} \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(1)}_{j_1,i_1,k,l} O^{(1)}_{j_2,i_2,k,l}}{\left(SE^{1}_{j,k,l}\right)^2} + \sum_{l \in \Gamma_2} \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(2)}_{j_1,i_1,k,l} O^{(2)}_{j_2,i_2,k,l}}{\left(SE^{2}_{j,k,l}\right)^2}$$

Next, we obtain the sub-matrix of H that includes the second partial derivatives of LL with respect to the δ parameters and get

$$\frac{\partial^2 LL_0}{\partial \delta_{i_1} \partial \delta_{i_2}} = \sum_{l \in \Gamma_2} \left\{ - \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{\sum_{v \neq i_1, i_2} L_{i_1 v} L_{i_2 v} \log(M_{i_1}(T^l_{j,k}, l) + 1) \log(M_{i_2}(T^l_{j,k}, l) + 1) E^{(1)}_{j,v,k,l}}{SE^{(2)}_{j,k,l}} \right.$$
$$\left. + \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{ME^{(1)}_{j,i_1,k,l} ME^{(2)}_{j,i_2,k,l}}{\left[SE^{(2)}_{j,k,l}\right]^2} \right.$$

when $1 \leq i_1, i_2 \leq n$.

Finally, we provide expressions for the cross-partials

$$\frac{\partial^2 LL_0}{\partial \alpha_{i_1,j_1} \partial \delta_{i_2}} = \sum_{l \in \Gamma_2} \left\{ - \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{L_{i_1 j_1} L_{i_2 j_1} \log(M_{i_1}(T^l_{j,k}, l) + 1) \log(M_{i_2}(T^l_{j,k}, l) + 1) E^{(2)}_{j,j_1,k,l}}{SE^{(2)}_{j,k,l}} \right.$$
$$\left. + \sum_{1 \leq j \leq n} \sum_{1 \leq k \leq n_j^l} \frac{O^{(2)}_{j_1,i_1,k,l} ME^{(2)}_{j,i_2,k,l}}{\left[SE^{(2)}_{j,k,l}\right]^2} \right.$$

when $1 \leq i_1, i_2, j_1 \leq n, i_1 \neq j_1$.

3.8.3 Proof of Theorem 2

To simplify the notations, similar to what we have done in the proof of Theorem 1 in Chapter 1, we first define the following notations:

$$\begin{aligned}
E_{1,l}(t,\Omega') &= \sum_{j=1}^{n} \lambda_{0,l}(t) \exp\left(\sum_{i,i\neq j} L_{ij}(t)\alpha'_{ij} \log(M_i(t,l)+1)\right) \\
E_{2,l}(t,\Omega') &= \sum_{j=1}^{n} \lambda_{0,l}(t) \exp\left(\sum_{i,i\neq j} L_{ij}(t)(\alpha'_{ij}+\delta'_i) \log(M_i(t,l)+1)\right) \\
\Phi'_{1,j} &= (\phi'_{1,1j},\ldots,\phi'_{1,nj}) \equiv (\alpha'_{1j},\ldots,\alpha'_{nj})' \\
\Phi'_{2,j} &= (\phi'_{2,1j},\ldots,\phi'_{2,nj}) \equiv (\alpha'_{1j}+\delta'_i,\ldots,\alpha'_{nj}+\delta'_i)' \\
E^{(1)}_{k,lj}(t,\Omega') &= \left(\frac{\partial E_{k,l}(t,\Omega')}{\partial \phi'_{k,1j}},\ldots,\frac{\partial E_{k,l}(t,\Omega')}{\partial \phi'_{k,nj}}\right), k=1,2 \\
E^{(2)}_{k,lj}(t,\Omega') &= \left(\frac{\partial^2 E_{k,l}(t,\Omega')}{\partial \phi'_{k,ij}\phi'_{k,qj}}\right)_{1\leq i,q\leq n}, k=1,2
\end{aligned} \tag{3.23}$$

Let we let

$$\begin{aligned}
e_{k,l}(t,\Omega') &\equiv E[E_{k,l}(t,\Omega')], \\
e^{(1)}_{k,lj}(t,\Omega') &\equiv \left(\frac{\partial e_{k,l}(t,\Omega')}{\partial \phi'_{k,1j}},\ldots,\frac{\partial e_{k,l}(t,\Omega')}{\partial \phi'_{k,nj}}\right) \\
e^{(2)}_{k,lj}(t,\Omega') &\equiv \left(\frac{\partial^2 e_{k,l}(t,\Omega')}{\partial \phi'_{k,ij}\partial \phi'_{k,i'j'}}\right)_{1\leq i,j,i',j'\leq n}
\end{aligned} \tag{3.24}$$

By Condition D of Theorem 2, we can let

$$X(\Omega',t) = \frac{1}{\|\Gamma_1\|_0}(LL(\Omega',t) - LL(\Omega,t))$$

Following the Step 1 to 3 in the proof of Theorem 1 in Chapter 1, and put the $\frac{\gamma}{\|\Gamma_1\|_0}\sum_{i=1}^{n}|\delta_i|$ into $o_P(1)$, it can be shown that

$$|X(\Omega',t) - P(\Omega',t)| \to_P 0, \text{ as } \|\Gamma_1\|_0, \|\Gamma_2\|_0 \to \infty,$$

where

$$P(\Omega', t_0) = \int_0^{t_0} \frac{1}{\|\Gamma_1\|_0} \sum_{l \in \Gamma_1} \left[\sum_{j=1}^{n} (\Phi_j' - \Phi_j)' e_{1,lj}^{(1)}(u, \Omega) - \log \left\{ \frac{e_{1,l}(u, \Omega')}{e_{1,l}(u, \Omega)} \right\} e_{1,l}(u, \Omega) \right] du$$
$$\int_0^{t_0} \frac{1}{\|\Gamma_1\|_0} \sum_{l \in \Gamma_2} \left[\sum_{j=1}^{n} (\Phi_j' - \Phi_j)' e_{2,lj}^{(1)}(u, \Omega) - \log \left\{ \frac{e_{2,l}(u, \Omega')}{e_{2,l}(u, \Omega)} \right\} e_{2,l}(u, \Omega) \right] du$$

Similar to our argument in Step 1 of the proof of Theorem 1, it can be shown that $P(\Omega', t_0)$ is strictly concave and uniquely maxiimized at $\Omega' = \Omega$. Further by Condition E, the smallest eigenvalue of $-\nabla_{\Omega'} \nabla_{\Omega'} P(\Omega', t_0)$ is not smaller than C_4, in probability. Following the argument of Step 3 of the proof of Theorem 1, it can be shown that $\hat{\Omega}$, the maximizer of $X(\Omega', t)$ converges in probability to Ω.

BIBLIOGRAPHY

BIBLIOGRAPHY

Andersen, P. K., and R. D. Gill (1982), Cox's regression model for counting processes: a large sample study, *The annals of statistics*, pp. 1100–1120.

Barbieri, N., F. Bonchi, and G. Manco (2013), Topic-aware social influence propagation models, *Knowledge and information systems*, *37*(3), 555–584.

Blei, D. M. (2012), Probabilistic topic models, *Communications of the ACM*, *55*(4), 77–84.

Blei, D. M., and J. D. Lafferty (2007), A correlated topic model of science, *The Annals of Applied Statistics*, pp. 17–35.

Blei, D. M., A. Y. Ng, and M. I. Jordan (2003), Latent dirichlet allocation, *the Journal of machine Learning research*, *3*, 993–1022.

Bulearca, M., and S. Bulearca (2010), Twitter: a viable marketing tool for smes?, *Global Business and Management Research: An International Journal*, *2*(4), 296–309.

Cha, M., H. Haddadi, F. Benevenuto, and P. K. Gummadi (2010), Measuring user influence in twitter: The million follower fallacy., *ICWSM*, *10*, 10–17.

Cox, D. R. (1972), Regression models and life tables (with discussion), *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *34*(2), 187–220.

Dave, P. (2015), Firms turn to online 'influencers' to spread the word on social media, *The Los Angeles Times*, accessed: 2015-01-05.

Dorogovtsev, S. N., and J. F. F. Mendes (2002), Evolution of networks, *Advances in Physics*, *51*(4), 1079–1187, doi:10.1080/00018730110112519.

Du, N., L. Song, M. Yuan, and A. J. Smola (2012), Learning networks of heterogeneous influence, in *Advances in Neural Information Processing Systems 25*, edited by F. Pereira, C. Burges, L. Bottou, and K. Weinberger, pp. 2780–2788, Curran Associates, Inc.

Dubois, E., and D. Gaffney (2014), The multiple facets of influence: Identifying political influentials and opinion leaders on twitter, *American Behavioral Scientist*, doi:10.1177/0002764214527088.

Fienberg, S. E. (2012), A brief history of statistical models for network analysis and open challenges, *Journal of Computational and Graphical Statistics*, *21*(4), 825–839, doi:10.1080/10618600.2012.738106.

Fleming, T., and D. Harrington (2013), *Counting Processes and Survival Analysis*, Wiley Series in Probability and Statistics, Wiley.

Gayo-Avello, D., P. T. Metaxas, and E. Mustafaraj (2011), Limits of electoral predictions using twitter., in *ICWSM*.

Golbeck, J., J. M. Grimes, and A. Rogers (2010), Twitter use by the u.s. congress, *Journal of the American Society for Information Science and Technology*, *61*(8), 1612–1621, doi:10.1002/asi.21344.

Gomez-Rodriguez, M., J. Leskovec, and B. Schölkopf (2013), Modeling information propagation with survival theory, *CoRR*, *abs/1305.3616*.

GovTrack.us (2014), Govtrack.us: Tracking the united states congress, accessed: 2014-10-26.

Haveliwala, T. H. (2003), Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search, *Knowledge and Data Engineering, IEEE Transactions on*, *15*(4), 784–796.

Hong, L., and B. D. Davison (2010), Empirical study of topic modeling in twitter, in *Proceedings of the First Workshop on Social Media Analytics*, pp. 80–88, ACM.

Kleinberg, J. M. (1999), Authoritative sources in a hyperlinked environment, *J. ACM*, *46*(5), 604–632, doi:10.1145/324133.324140.

Klout (2014), Klout — be known for what you love, accessed: 2014-11-01.

Kolaczyk, E. D. (2009), *Statistical Analysis of Network Data: Methods and Models*, Springer Series in Statistics, Springer Publishing Company, Incorporated, New York, NY, USA, doi:10.1007/978-0-387-88146-1.

Kwak, H., C. Lee, H. Park, and S. Moon (2010), What is twitter, a social network or a news media?, in *Proceedings of the 19th international conference on World wide web*, pp. 591–600, ACM.

Miritello, G., E. Moro, and R. Lara (2011), Dynamical strength of social ties in information spreading, *Phys. Rev. E*, *83*, 045,102, doi:10.1103/PhysRevE.83.045102.

Newman, M. E. J. (2003), The structure and function of complex networks, *SIAM Review*, *45*(2), 167–256, doi:10.1137/S003614450342480.

Page, L., S. Brin, R. Motwani, and T. Winograd (1999), The pagerank citation ranking: Bringing order to the web.

PeerIndex (2014), Peerindex – think people, accessed: 2014-11-01.

Perry, P. O., and P. J. Wolfe (2013), Point process modelling for directed interaction networks, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *75*(5), 821–849.

Probst, F., L. Grosswiele, and R. Pfleger (2013), Who will lead and who will follow: Identifying influential users in online social networks, *Business & Information Systems Engineering*, *5*(3), 179–193, doi:10.1007/s12599-013-0263-7.

Ramage, D., S. Dumais, and D. Liebling (2010), Characterizing microblogs with topic models, in *Proc. ICWSM 2010*, American Association for Artificial Intelligence.

Salter-Townshend, M., A. White, I. Gollini, and T. B. Murphy (2012), Review of statistical network analysis: models, algorithms, and software, *Statistical Analysis and Data Mining*, *5*(4), 243–264, doi:10.1002/sam.11146.

SAS (2015), Sas social media analytics, accessed: 2015-01-05.

Taddy, M. (2013), Measuring political sentiment on twitter: Factor optimal design for multinomial inverse regression, *Technometrics*, *55*(4), 415–425, doi:10.1080/00401706.2013.778791.

Toivonen, R., J. M. Kumpula, and J. Saramaki (2007), The role of edge weights in social networks: modelling structure and dynamics, in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*.

Tong, H., B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos (2012), Gelling, and melting, large graphs by edge manipulation, in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM '12, pp. 245–254, ACM, New York, NY, USA, doi:10.1145/2396761.2396795.

Trusov, M., A. V. Bodapati, and R. E. Bucklin (2010), Determining influential users in internet social networks, *Journal of Marketing Research*, *47*(4), 643–658.

Twitter, I. (2014), About twitter, inc., accessed: 2014-09-19.

Weng, J., E.-P. Lim, J. Jiang, and Q. He (2010), Twitterrank: finding topic-sensitive influential twitterers, in *Proceedings of the third ACM international conference on Web search and data mining*, pp. 261–270, ACM.