Allies and Rivals: Modeling Citation Dynamics among Party-Credentialed Blogs in the 2004 US Electoral Cycle*

Zack W. Almquist† Carter T. Butts‡

Working Paper
Please do not cite
April 22, 2011

*This work was supported in part by ONR award #N00014-08-1-1015.

†Department of Sociology; 3151 Social Science Plaza A; University of California, Irvine; Irvine, CA 92697-5100; almquist@uci.edu

‡Department of Sociology and Institute for Mathematical Behavioral Sciences; University of California, Irvine

Allies and Rivals: Modeling Citation Dynamics among Party-Credentialed Blogs in the 2004 US Electoral Cycle

Abstract

The 2004 US Presidential Election cycle marked the debut of Internet-based media such as blogs and social networking websites as institutionally recognized features of the American political landscape. Particularly significant was the credentialing of selected blogs as officially designated media sources for purposes of covering the major political party conventions, an act which gave particular legitimacy to two contending groups of partisan blogs (one credentialed for the Republican National Convention (RNC) and the other for the Democratic National Convention (DNC)). In the months that followed, these blogs served as significant foci for online journalistic, promotional, fund-raising, and organizing activities relating to the 2004 election.

In this study, we employ a dynamic logistic choice model to study the dynamics of interaction within and between these two groups of political blogs. Using a longitudinal sample of all DNC and RNC-designated blog citation networks (sampled at six hour intervals for approximately four months) from Butts and Cross (2009) we are able to test for the influence of various strategic, institutional, and balance-theoretic mechanisms – as well as exogenous factors such as seasonality and political events – on the propensity of blogs to cite (i.e., hyperlink to) one another over time. Capitalizing on the temporal resolution of our data, we utilize an autoregressive network regression framework to carry out inference for a logistic choice process closely related to the actor-oriented framework of Snijders (2001). Using a combination of deviance-based model selection criteria (e.g. BIC) and simulation-based goodness-of-fit tests akin to Hunter et al. (2008), we identify the combination of processes that best characterizes the choice behavior of the contending blogs. We conclude with a discussion of the potential for autoregressive network regression as a practical way of “scaling up” dynamic choice models for use with high-resolution data sets.

Allies and Rivals: Modeling Citation Dynamics among Party-Credentialed Blogs in the 2004 US Electoral Cycle

1 Introduction

The 2004 US Presidential Election cycle marked the debut of Internet-based media such as blogs and social networking websites as institutionally recognized features of the American political landscape. Particularly significant was the credentialing of selected blogs as officially designated media sources for purposes of covering the major political party conventions, an act which gave particular legitimacy to two contending groups of partisan blogs (one credentialed for the Republican National Convention (RNC) and the other for the Democratic National Convention (DNC)). In the months that followed, these blogs served as significant foci for online journalistic, promotional, fund-raising, and organizing activities relating to the 2004 election.

In this study, we employ a dynamic logistic choice model to study the dynamics of interaction within and between these two groups of political blogs. Using a longitudinal sample of all DNC and RNC-designated blog citation networks (sampled at six hour intervals for approximately four months) from Butts and Cross (2009) we are able to test for the influence of various strategic, institutional, and balance-theoretic mechanisms – as well as exogenous factors such as seasonality and political events – on the propensity of blogs to cite (i.e., hyperlink to) one another over time. Capitalizing on the temporal resolution of our data, we utilize an autoregressive network regression framework to carry out inference for a logistic choice process closely related to the actor-oriented framework of Snijders (2001).

This paper is structured as follows. We begin by providing some general background from the relevant political science and social network literatures, with a particular focus on the role of political blogs during the study period. This is followed by a description of the study data, and an overview of our modeling approach. The latter includes both a discussion of the general assumptions behind the modeling of blog evolution as a dynamic decision process, and a treatment of the factors potentially shaping actors' payoffs. We follow this with a discussion of our implementation and inferential framework, data analysis, and findings. Finally, we conclude with a discussion of the implications of our results for our understanding of the social mechanisms shaping contentious groups in the online environment.

2 Background

In recent years, the online world has generated a diverse array of new media for social interaction (Wellman, 2001), one of the most successful of which is the weblog (or "blog"). While a relatively obscure medium for many years, the growing popularity of blogs as a means for information dissemination, coordination, and political organization through the early to mid 2000s eventually led to their recognition of and adoption by established institutions. A key landmark in this process was the 2004 US Presidential election cycle, in which the DNC and RNC first granted press credentials to selected bloggers for coverage of their national political conventions Adamic and Glance (2005); Butts and Cross (2009); Howard (2005); Rainie et al. (2005). This institutionalized legitimation by the major US political parties constituted a de facto recognition of the role of blogs (and the online community more broadly) as a durable element of the political landscape, and arguably marked the debut of the "new media" as a force in electoral politics.

The impact of blogs first gained institutional attention in the US political sphere in the early phases of the 2004 US electoral cycle, when Democratic presidential candidate and Vermont Governor Howard Dean rose to prominence partially as a result of his extensive use of online organizing to compensate for limited conventional resources in garnering media attention and raising funds (Ammori, 2005; Kerbel and Bloom, 2005). Dean's success in utilizing online interaction to mobilize a widely dispersed base of supporters was quickly noted by political observers, and (despite his loss of the Democratic nomination to Senator John Kerry) paved the way for other politicians to incorporate online media into their political campaigns (Cone, 2003). Indeed, by the the end of the 2004 electoral cycle blogs and other online resources had been adopted by a number of Presidential contenders, and (via actions such as the above-mentioned credentialing of bloggers as members of the press) by the major US political parties themselves. These and and further developments in the historical evolution of the online environment over the past decade have set the stage for academic, governmental, non-profit, and for-profit interest in blogs and other new media, particularly in political contexts (Drezner and Farrell, 2008).

With respect to the role played by blogs per se, Woodly (2008) demonstrates that blogs are actively used in mobilizing opinions, setting agendas, and generally influencing the elite members of the political parties. His work demonstrates that the interactions between political blogs are a particularly important dimension of this phenomenon. Because a distinctive feature of blogs is their combination of commentary on current events with hypertext references to primary or secondary information sources, the constantly evolving network of citations between blogs is at least as significant (e.g., from an information search standpoint) as the content of the individual blogs themselves. Within this network of references, blog authors (or "bloggers") have become a new form of journalist, in some cases with similar information access and responsibilities to practitioners within traditional media outlets (Wall, 2005). As the importance of this medium has continued to increase in recent years, its growth in size and elaboration has made its study both relevant and difficult. We thus focus our attention on the initial "watershed" period of the 2004 US Presidential election, when the role of blogs as legitimated media entities was just beginning to crystalize. In particular, our attention centers on the interactions among the relatively small number of blogs credentialed for the major party political conventions, as they jockeyed to promote their issues, candidates, and arguably themselves in the midst of a rapidly changing political and technological landscape. As players with some institutional recognition but little control from established political actors, these blogs provide an early example of a phenomenon that has become increasingly common throughout the developed world.

3 Data

The data used in this paper is a dynamic inter- and intra-group blog citation network collected by Butts and Cross (2009), consisting of interactions among all blogs creditialed by the DNC or RNC for their respective 2004 conventions. Specifically, the vertex set for this network consists of 34 DNC and 14 RNC credentialed blogs (with one blog credentialled by both groups) providing a combined network of 47 nodes observed over a 121 day period. Network data was obtained by automatically querying the main page of each blog at six hour intervals starting at midnight, Pacific time. The period of observation for this study begins on 7/22/04 (shortly before the DNC convention), and ends 11/19/04 (shortly after the Presidential election), leading to a total of 484 time points. At each time point, the collected data consists of the network of URLs linking the main page of one blog to any page within another; i.e., there is an edge from blog i to blog j at time t if a link to blog j appears on the main page of i at time t. We may conceive of this data as an adjacency array, A, such that $A_{ij,t} = 1$ if i cites (i.e., links to) j at time t, and 0 otherwise. For purposes of this study, we ignore self-citations (i.e., internal links from a blog to itself).

In addition to the evolving blog network, Butts and Cross (2009) provide a timeline of major events during the campaign cycle, dividing the 121 day period into a series of "epochs" based on salient activities such as the RNC and DNC conventions, the televised Presidential debates, and the election itself (Table 1). In an analysis of volatility within the RNC and DNC networks (taken separately), Butts and Cross (2009) find that these campaign events are related to the pace of change within the network (along with daily and weekly seasonal effects). As such, we include these temporal effects as covariates in our analyses (as described below).

4 Network Evolution as a Decision Process

Blogs of the type studied here are the deliberately constructed and maintained products of individuals, or small groups thereof. Moreover, those blogs credentialed during the 2004 electoral cycle represented a small "elite" circle of especially active authors, whose blogs centered on coverage of politics and current events. As such, it is reasonable to consider modeling the evolving blog network as arising from a dynamic decision process, in which blog authors select those to whom they link in response to context and past history. This approach has been most fully developed by Snijders

Epoch	Description	Start Time	End Time	Time Points
PreCon	Start of window to DNC Convention	7/22, 00:00	7/25,18:00	16
DNCCon	DNC Convention	7/26, 00:00	7/29, 18:00	16
InterCon	End of DNC Convention to start of RNC Convention	7/30, 00:00	8/29, 18:00	124
RNCCon	RNC Convention	8/30, 00:00	9/2, 18:00	16
PreDeb	End of RNC Convention to first presidential debate	9/3, 00:00	9/20, 12:00	71
Deb	First presidential debate to last presidential debate	9/20, 18:00	10/14, 18:00	93
PreElec	Post last presidential debate to Election Day	10/14, 00:00	11/1, 18:00	76
Elec	Election Day	11/2, 00:00	11/2, 18:00	4
PostElec	Post election to end of window	11/3, 00:00	11/19, 18:00	67

Table 1: Epochs in the 2004 Election Cycle from Butts and Cross (2009).

(1996, 2001); Snijders and Van Duijn (1997), who posits an “actor-oriented” model in which network members change their relationships via a latent continuous-time choice process. We here employ a somewhat simpler version of this general scheme, which represents network evolution as a discrete time logistic choice process (McFadden, 1974, 1976). Although requiring somewhat stricter assumptions on decision simultaneity, this variant facilitates the accommodation of complex backward-looking behavior, and scales more easily to larger data sets.

Although the inferential aspects of this framework will be described in Section 5, we begin here by presenting the model from a behavioral point of view. First, we review the notion of edge updating as a logistic choice process (Snijders, 2001), with a specific emphasis on its interpretation in the present case. As a revealed preference model, the logistic choice framework requires a parametric utility function; thus, we follow our initial discussion with a consideration of the payoff elements that may be expected to enter into blog authors’ decision-making processes, as they decide to whom they will or will not link. These payoff elements will form the core building blocks for our analysis of the evolving blog network.

4.1 Edge Updating as Logistic Choice

At its crudest level, a blog is a web page with dynamically updated links to other online resources. The core decision facing a blog author, then, is that of the other sites to which he or she should link, and (conversely) the links that can be removed (directly, or by allowing them to “expire” by no longer being shown on the blog’s front page). Such citations can be controlled on an individual basis, and are limited only by attentional and/or energetic costs: there is in principle no effective limit on the number of citations that can be maintained, and no barrier to adding or removing citations when desired. At the same time, adding or removing links requires attention and effort on the part of the author, and is thus the result of deliberate action (as opposed, e.g., to the accidental, incidental, or automatic behaviors that are of considerable importance in face-to-face settings (Goffman, 1959)). Blog authors – particularly active ones, such as those represented in this sample – can and do spend considerable time monitoring their environment, and may thus be expected to be aware of and react to the actions of salient alters; moreover, recent citation history is relatively easily discovered in this environment, potentially facilitating the use of backward-looking strategies. On the other hand, the complexity and dynamic nature of the online environment make prediction difficult, suggesting a very limited capacity for forward-looking behavior.

Taken together, the above considerations suggest the following propositions as a reasonable starting point for modeling the evolution of the blog network. For simplicity of discussion, we will refer to the “blog” as the unit of decision making, and the links or citations from one blog to another as “edges” within the associated network.

1. The state of outgoing edges at each observation of the blog network is assumed to result from the choices of the sending blog;

2. Each blog in the network may send an edge to any number of other blogs in the network at any time;

3. The decision of a given blog regarding the state of a given edge is made myopically, and in isolation (i.e., the decision is considered on its own terms, without factoring in the effects of other decisions that might be made simultaneously);

4. The decision of a given blog regarding the state of a given edge may depend upon the past history of the blog network, or of the current external context (e.g., time of day, electoral cycle events)

Subject to the above, we further presume that blog citation behavior follows a weakly consistent pattern of preferences, in the sense that there exists a *utility function,* u, such that for the two alternative states $A_{ij,t} = 0$ and $A_{ij,t} = 1$, the odds that i will choose $A_{ij,t} = 1$ are strictly increasing in $u_i(A|A_{ij,t} = 1)/u_i(A|A_{ij,t} = 0)$. Such a pattern of behavior is typically referred to as a *stochastic choice* process, and can be viewed as a form of bounded rationality. Although many stochastic choice models exist, we here use the common *logistic choice* model. In the present case, this amounts to the assumption that

$$\Pr(A_{ij,t} = 1) = \frac{\exp\left[u_i\left(A|A_{ij,t} = 1\right)\right]}{\exp\left[u_i\left(A|A_{ij,t} = 1\right)\right] + \exp\left[u_i\left(A|A_{ij,t} = 0\right)\right]}, \tag{1}$$

or, equivalently, that

$$\operatorname{logit}(A_{ij,t}) = \ln \frac{\Pr(A_{ij,t} = 1)}{\Pr(A_{ij,t} = 0)} = u_i\left(A|A_{ij,t} = 1\right) - u_i\left(A|A_{ij,t} = 0\right), \tag{2}$$

i.e., the log-odds that i will choose to cite j at time t is equal to the utility difference associated with sending (versus not sending) an edge. Where the utility of one option is substantially greater than the other, then, actor behavior is nearly deterministic: the utility-increasing choice is selected with very high probability. As the actor approaches indifference, however, choice behavior becomes increasingly random (an effect interpretable either as difficulty in determining the preferable option, or as reflecting the influence of various small, idiosyncratic payoffs). When the actor is entirely indifferent between citing and not citing another, the choice becomes fully arbitrary (i.e., a coin flip).

To put this scheme into practice, we must make some further assumptions regarding the nature of the utility function. From our list of propositions, we have assumed that decisions are made myopically, depending on the past (and on general context), but not on simultaneous or future decisions. As such, we require that u depend upon the network history, A, only through its prior states, and through the conjecturally perturbed state associated with a single decision (i.e., for the $A_{ij,t}$ decision, u_i may depend upon $A_{\cdot\cdot,t-k}$ where $k > 0$, and on $A_{\cdot\cdot,t}$ such that $A_{gh,t} = A_{gh,t-1}$ for all $g, h \neq i, j$). u may also depend upon t, and on exogenous covariates (denoted by X). Finally, we will assume in general that u can be written as a sum of linearly separable payoff elements, s, such that $u_i(A|A_{ij,t}) = \theta^T s(A, A_{ij,t}, i, j, t, X)$. Intuitively, s expresses the factors potentially driving i's behavior, while the parameter vector θ expresses the direction and magnitude of the effect these factors have on the propensity to send or refrain from sending a tie.

As a model of boundedly rational dynamics, the logistic choice framework is quite general: a wide range of factors can potentially enter into the utility function, and the choice of possible candidates must be made based on substantive considerations. With that in mind, we now turn to a consideration of the payoff elements that may plausibly drive behavior within the blog network.

4.2 Potential Payoff Elements

The population of interest for our study consists of two distinct groups (whose membership is common knowledge): those blogs credentialed by the DNC, and those credentialed by the RNC. In general, these credentials constitute an implicitly partisan identification, as reflected in the fact that only one blog (seen as relatively neutral) was credentialed by both organizations. As such, we may view the blog network as consisting of two contentious factions competing for very real and tangible stakes in the US political arena (see Drezner and Farrell, 2008, etc.). In considering these two groups as having a largely adversarial relationship (Hargittai et al., 2008), it is natural to ask whether their behavior can be predicted by the principles of structural balance (Cartwright and Harary, 1956; Heider, 1958). The dynamic nature of our data allows us to consider such effects as motivators for future action, rather than merely as predictors of cross-sectional patterns. Given the nature of the interacting parties, we assume for balance theoretic purposes that blogs credentialed by the same source are linked by a unit relation (in the sense of Heider), and are thus pre-emptively bound by positive ties. Likewise, we treat hypothetical cross-group interactions as presumptively negative. These assumptions (consistent with the content of the blogs and the nature of their interactions during this period) allow us to identify balance-theoretic mechanisms that potentially govern the interactions between blogs; the extent to which actor behavior does or does not reveal preferences consistent with these mechanisms allows us to infer the behavioral factors governing the evolution of the blog network.

4.2.1 Mixing

In the social network literature, patterns of interaction between a priori defined groups are referred to as *mixing patterns*, with systematic biases in such interaction being known as *nonrandom mixing*. Within the context of the blog network, the most obvious prediction of this type is that payoffs for citations to ingroup members will differ from citations to outgroup members; this follows hypotheses of researchers such as Hargittai et al. (2008), who argues that bloggers cluster ideologically and thus exhibit a preference for citing other blogs with the same ideology. If present, this should manifest as a strong "xenophobia" effect (in the sense of Petrescu-Prahova (2007)), with the payoff for citation of outgroup members being lower, ceteris paribus, than the payoff for citing one within one's own group.

A potential alternative hypothesis arises from the observation that the RNC and DNC blogs, while adversarial, are necessarily engaged in a form of rhetorical exchange: to argue with another is nevertheless to interact with them, and we may propose that blogs will in fact actively cross-link for the explicit purpose of refuting claims made by outgroup members. If this is so, then we should see little or no difference (again, ceteris paribus) between the payoffs associated with citations to ingroup members and those to blogs outside one's own group.

To summarize, we propose two contending hypotheses for mixing:

Mixing Hypothesis 1 The partial payoff of an edge from blog i to blog j will be lower if i and j do not belong to the same credentialing faction than if i and j belong to the same faction.

Mixing Hypothesis 2 The partial payoff of an edge from blog i to blog j will be the same or higher if i and j do not belong to the same credentialing faction than if i and j belong to the same faction.

4.2.2 Balance-Theoretic Influences

Cartwright and Harary (1956) introduced the concept of generalized structural balance based on Heider's (1958) cognitive theory of balance, which suggests a number of possible mechanisms for how an actor's evaluative judgments may evolve in response to particular simuli. Heider's theory stems from Gestalt psychology and posits that individuals attempt to bring their complete sets of evaluative judgments into a stable state (known as "balance"). Structural balance generalizes from within-actor impressions to realized relationships, and can be motivated on both cognitve and strategic grounds. The theory of balance suggests a number of different possible mechanisms which might influence an actor's utility function and thus the weights on his or her payoff elements (s).

We propose the four competing hypotheses (BT Hypothesis 1-4), broken up into two competing sets: BT Hypothesis 1 versus BT Hypothesis 2 and BT Hypothesis 3 versus BT Hypothesis 4. The first group tests the influence of extended ingroup and outgroup effects and the second group tests reciprocity effects.

BT Hypothesis 1: "Friend of a friend" (In-Group two paths) We hypothesize that the partial payoff for a link from blog i to blog j is increasing in the number of i, j two-paths contained in the same faction, (for ex. in this case, the RNC or DNC). In other words, the payoff for Ego to form a relationship with a "friendly" is enhanced if that friendly is also connected to Ego via another "friendly" (Figure 1).

$$\text{Ego} \rightarrow \text{Friendly A} \rightarrow \text{Friendly B} \Rightarrow \text{Ego} \rightarrow \text{Friendly B}$$

Figure 1: Graphic of In-Group two path (friend of a friend).

BT Hypothesis 2: "Friend of an enemy" (Cross-Group two paths) We hypothesize that the partial payoff for a link from blog i to blog j is decreasing in the number of two paths through individuals of the same faction as i, where j and i belong to different factions. (Figure 2).

$$\text{Ego} \rightarrow \text{Friendly A} \rightarrow \text{Hostile B} \Rightarrow \text{Ego} \rightarrow \text{Hostile B}$$

Figure 2: Graphic of Cross-Group two path (friend of an enemy).

BT Hypothesis 3: Reciprocity ("friendly") We hypothesize that reciprocity will be accompanied with positive gains for ingroup edge creation and low or negative gains for across-group citation. In this case citations are primarily a positive relation such that utility gain comes from increasing the prominence of ones friends and not through competition with ones enemies.

$$\text{Friendly} \rightarrow \text{Ego} \Rightarrow \text{Ego} \rightarrow \text{Friendly}$$

Figure 3: Graphic of Reciprocity between friends.

BT Hypothesis 4: Reciprocity ("hostile") Conversely we hypothesize that reciprocity will be accompanied with positive gains for outgroup edge creation and low or negative gains for ingroup citation. In this case citations are primarily a negative relation such that utility gain comes from refuting enemies' accusations.

Hostile → Ego ⇒ Ego → Hostile

Figure 4: Graphic of Reciprocity between enemies.

4.2.3 Context and Seasonality

We know that networks tend to have certain baseline characteristics which need to be accounted for any network analysis (e.g., in-degree and out-degree effects; Wasserman and Faust (1994)).

Likewise, in a time-series context it is known that there are certain seasonal and period effects that occur in most temporally collected data (Shumway and Stoffer, 2006). Common seasonal effects in behavior data include daily and hourly effects (e.g., Monday, Tuesday, etc. and midnight versus midday).

Seasonality Hypothesis 1 Butts and Cross (2009) found that the volatility of the blog networks changes with time of day, day of week, and period in the electoral cycle. One way in which "volatility" might manifest in this case is through an increase or decrease degree of inertia in the network structure. Thus, the partial payoff to continuing a previous action is predicted to follow seasonal and periodic patterns.

Seasonality Hypothesis 2 We suspect that overall propensity to send links will vary over time. We argue that ego's linking to others involves a search process, and is consumptive of attentional/energetic resources. Resource availability varies seasonally, and with it the partial payoff for sending links per se.

Seasonality Hypothesis 3 We propose that behavioral factors might change with time and context:

Selective salience We might expect that there would be larger propensity to create ties across or within group during important events in the election cycle (see Table 1).

5 Methodology

This work employs the methodology – Dynamic Lagged-Logistic Network Regression – recommended by Almquist and Butts (2010) for large dynamic data-sets, which builds on the Exponential Random Graph (Butts, 2008; Holland and Leinhardt, 1981a,b; Snijders et al., 2006; Strauss and Ikeda, 1990) and Network Regression literatures (Krackhardt, 1987a,b, 1988). This model is particularly appealing in this context because it is very natural to model citation linking as a binary choice, and this framework allows us to explore the mechanisms that predict whether one blogger chooses to cite another blogger.

We begin by discussing the necessary mathematical details and then follow this up with how we operationalize the mechanisms discussed in Section 4.2. Next, we discuss certain necessary control parameters (i.e., seasonality and other network effects). Lastly, this is followed by a brief discussion on the computation of the models.

5.1 Dynamic Lagged-Logistic Network Regression

A standard inferential framework for network analysis is that of the Exponential Random Graph Model (ERGM) (see, e.g. Butts, 2008; Holland and Leinhardt, 1981a, etc.). This family of models may be written in the following fashion,

$$P_\theta(A = a) = \frac{\exp\{\theta^t t(a)\}}{c(\theta)} \tag{3}$$

where A is a random network (represented by random adjacency matrix A) on n nodes (directed or undirected). θ is a vector of parameters, $t(y)$ is a known vector of graph statistics on a. $c(\theta) = \sum_{a \in \mathcal{A}} \exp\{\theta^t t(A)\}$ (i.e. the sum over all possible graphs a).

Almquist and Butts (2010) demonstrate that if one assumes the network only depends upon the past history and/or on exogenous factors (i.e., covariates X) one can show that this leads to *lagged-logistic network regression*. Specifically, these assumptions lead to conditionally independent edge realizations at each time step, with the conditional log-odds of an i, j edge given by:

$$\text{logit}(A_{ij,t}) = \theta^t \left[t_{ij}(A|A_{ij,t} = 1, X) - t_{ij}(A|A_{ij,t} = 0, X) \right] \tag{4}$$

In Section 4 we introduced the notation and theory of a dynamic logistic-choice model. Under the assumptions of a logistic-choice framework (Eq. 1, 2) and the independence assumptions of Section 4.1 it is clear that in our case the logistic-choice model is indeed equivalent to lagged-logistic network regression where the inferred parameters represent the weights the blogger places on utility differential. Thus, we may directly employ the framework of Almquist and Butts (2010) to estimate utilities under the behavioral assumptions stated in Section 4.1.

5.2 The Model

5.2.1 Mixing Terms

Methods for modeling nonrandom mixing have been known for some time; for our purposes we employ a method similar to Morris (1991), and use a type of block model to represent the two groups (see Wasserman and Faust, 1994, for a full review of the literature on block modeling). In doing so we impose four parameters on the model: one parameter for each group and one parameter for each group's interaction. (Note that since this is a directed network, RNC→DNC is different than DNC→RNC.) We assume these two groups represent competing organizations, and thus expect to see a higher propensity for within group citation than between group citation.

The mixing terms are modeled as a block matrix such that the model contains an indicator variable for within group edges and between group edges, introducing a total of four parameters into the model: one for the DNC, one for the RNC, one for the DNC→RNC, and one for the RNC→DNC interaction. This model is identifiable and replaces the standard edge (count of the number of edges in the model) or density (number of edges divided by the total number of possible edges) effects typically used in ERG models (Morris, 1991; Wasserman and Faust, 1994).

Mixing Hypothesis 1 The weights of the ingroup effects will be large and positive and cross-group effects will be small or negative; i.e. the model will favor ingroup, but not cross-group mixing.

Mixing Hypothesis 2 The weights of the ingroup effects will be smaller than the cross-group effects; i.e. the model will favor cross-group mixing.

5.2.2 Heiderian Terms

BT Hypothesis 1: Friend of a friend (In-Group two paths) A count of the number of ingroup two paths a given edge is involved. We expect the weight on this term to be positive and significant.

BT Hypothesis 2: Friend of an enemy (Cross-Group two paths) A count of the number of cross-group two paths a given edge is involved. We expect the weight on this term to be negative and significant.

BT Hypothesis 3 and : Reciprocity (friendly/hostile) An indicator if a relation is reciprocal and between ingroup members and an indicator if a relation is reciprocal and between group members. We expect these terms to have opposite signs (positive and negative for BT 3 and negative and positive for BT 4).

5.3 Controls

5.3.1 Network Effects

We expect to find baseline network effects, the most important of which is that of In-Degree and Out-Degree, which we model with a simple count of the In and Out-degree of each edge.

5.3.2 Seasonality

Seasonality Hypothesis 1 There will be substantive and large inertial effect, i.e. the lag term will be large and significant.

Seasonality Hypothesis 2 We test the hypothesis that the overall propensity to send links will vary over time via an interaction between hourly fixed effects and In-degree and Out-degree effects. We expect these to be important terms in the model, to be large and significant.

Seasonality Hypothesis 3 Selective salience We test the "selective salience" hypothesis with nine period effects, we expect there to be increase in activity during *PreCon, DNCCon, RNCCon, Deb, Elec* and decrease in activity during *InterCon, PreDeb, PreElec,* and *PostElec* (See Table 1 for details).

To control for hourly seasonality we employ three parameters for hour 6, 12, and 18 in the day with hour 0 as the reference group (ϕ_6, ϕ_{12}, ϕ_{18}); note that this is normalized between 0 and 1 (e.g. $\frac{6}{24}$).

To control for weekly seasonality we employ some ideas from Harmonic Regression (Shumway and Stoffer, 2006). Equation 5 is used to model daily seasonality (days within the week). We assume a classic "signal in noise" with a hidden periodic signal, that is a single sinusoid, which can be modeled as follows,

$$R\cos(2\pi\omega_d t + \Phi)$$

Using the classic trigonometric formula $\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b)$ we can derive the terms we place in the model (Eq. 5).

$$\begin{aligned} R\cos(\Phi)\cos(2\pi\omega_d t) + -R\sin(\Phi)\sin(2\pi\omega_d t) \\ \theta_1\cos(2\pi\omega_d t) + \theta_2\sin(2\pi\omega_d t) \end{aligned} \tag{5}$$

5.4 Computation

All computation for this article was written and executed in the `R` environment[1]. We employ a modified form of the code used in Almquist and Butts (2010).

6 Analysis

The first model we employ contains just the Mixing terms and seasonal effects (Model 1). We then add a single lag term (Model 2), followed by adding network control

[1] `http://www.r-project.org/`

effects (Model 3). Next, we add in the Heiderian mechanisms (Model 4); lastly we add in period effects (Model 5 and Model 6), and hourly interaction terms with In-degree and Out-degree (Model 6).

6.1 Results

We start by employing the BIC model selection (Schwarz, 1978), where the model with the lowest BIC is chosen as the best one. We then perform model goodness-of-fit via simulation analysis of the one-step predictor (Almquist and Butts, 2010; Hunter et al., 2008). This method is based upon analyzing so-called Graph Level Indices (GLI) (Anderson et al., 1999) in order to analyze how well the model captures properties that are thought to be important.

In the goodness-of-fit simulation we treat the model as though it is producing an inhomgenious Bernoulli graph, which we then take repeated draws from in order to represent the predicted distribution. We then plot the bar-graphs of each time prediction and collect a number of given GLIs at each time point. We analyze the resulting temporal plots to see if the predicted distribution covers the realized GLI and trend (Figure 5).

We see that that Model 6 captures the basic trend for Density, Average In and Out-Degree, Triad Census 300, and Krackhard's Connectedness measure (Figure 2); however it appears this model under-predicts the Triad Census 003 statistic. Overall, it appears that this model captures many of the GLIs that would be important in our logistic-choice model of blog-to-blog citations.

It is important when interpreting the parameters of Model 6 to realize that a number of the parameters can not be interpreted individually. Take for example the mixing terms; while two of the mixing terms are negative they cannot be interpreted without taking into account the lag term which is larger than any one of the mixing terms and positive. We see that the base propensity for within group blog-to-blog citation is higher for the RNC than DNC and that the two groups are about equally likely to send ties across group.

6.2 Findings

We confirm Mixing Hypotheses 1 that within group linking is more likely than between group linking and refute Mixing Hypothesis 2. We see that the two groups have similar levels of between group linking. In practice this means the RNC group spends more time recommending other RNC blogs and about equal amount of time refuting DNC blogs.

We confirm our balance-theoretic In-Group Two-Path (friend of a friend) and Cross-Group Two-Path (Friend of my Enemy) hypotheses. However, the question remains, what are the practical implications of this result? Well, for one, citations are more likely to flow through friends or enemies than across. We see that In-Group

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
BIC	598094.826	42209.8184	41863.9425	41246.5485	38484.5318	38408.6416
DNC	-1.796*	-5.8729*	-5.9459*	-6.0266*	0.2001*	0.2003*
RNC	-0.9401*	-4.0849*	-4.1857*	-4.432*	0.9405*	0.9403*
DNC→RNC	-3.9527*	-6.9714*	-6.996*	-6.9992*	-0.5548*	-0.5548*
RNC→DNC	-3.3602*	-5.2977*	-5.496*	-5.7088*	-0.5605*	-0.5603*
LAG		10.8933*	10.739*	10.4363*	9.9406*	9.938*
In-Degree			0.0633*	0.0353*	0.0784*	0.0637*
Out-Degree			-0.0523*	-0.0321*	0.014*	0.0135*
In-Group 2-Path				0.2165*	0.194*	0.1941*
Cross-Group 2-Path				0.5393*	-0.1069*	-0.1071*
In-Group Reciprocity				-0.5401*	-0.3479*	-0.3465*
Between-Group Reciprocity				0.3752	0.9911*	0.9918*
θ_1	0.0016	0.0604*	0.058*	0.0525*	0.0422	0.0422
θ_2	0.0363*	0.1593*	0.145*	0.1228*	-0.094*	-0.0942*
ϕ_6	-8.9597*	-52.952*	-49.2272*	-43.085*	22.5189*	10.1655*
ϕ_{12}	-4.428*	-28.3153*	-26.4316*	-23.2371*	9.3687*	3.0762
ϕ_{18}	-2.9948*	-22.0383*	-20.7868*	-18.5321*	1.3562	-1.2952
In-Degree$\times\phi_6$					-2.047*	
In-Degree$\times\phi_{12}$					-0.7563*	
In-Degree$\times\phi_{18}$					-0.6846*	
Out-Degree$\times\phi_6$					-0.1363	
Out-Degree$\times\phi_{12}$					-0.3967	
Out-Degree$\times\phi_{18}$					0.2572	
PreCon					-6.5823*	-6.4952*
DNCCon					-6.7895*	-6.7017*
InterCon					-7.0049*	-6.9177*
RNCCon					-6.9668*	-6.8796*
PreDeb					-6.9792*	-6.8918*
Deb					-7.001*	-6.9134*
PreElec					-6.9764*	-6.8891*
Elec					-7.5576*	-7.47*
PostElec					-7.0231*	-6.9357*

Table 2: Six models for dynamic logistic choice of Inter and Intra-group Blog Citation networks in the 2004 US Presidential Election ordered by BIC. Significant at "*" 0.05 p-value level under a z-test unless otherwise specified.

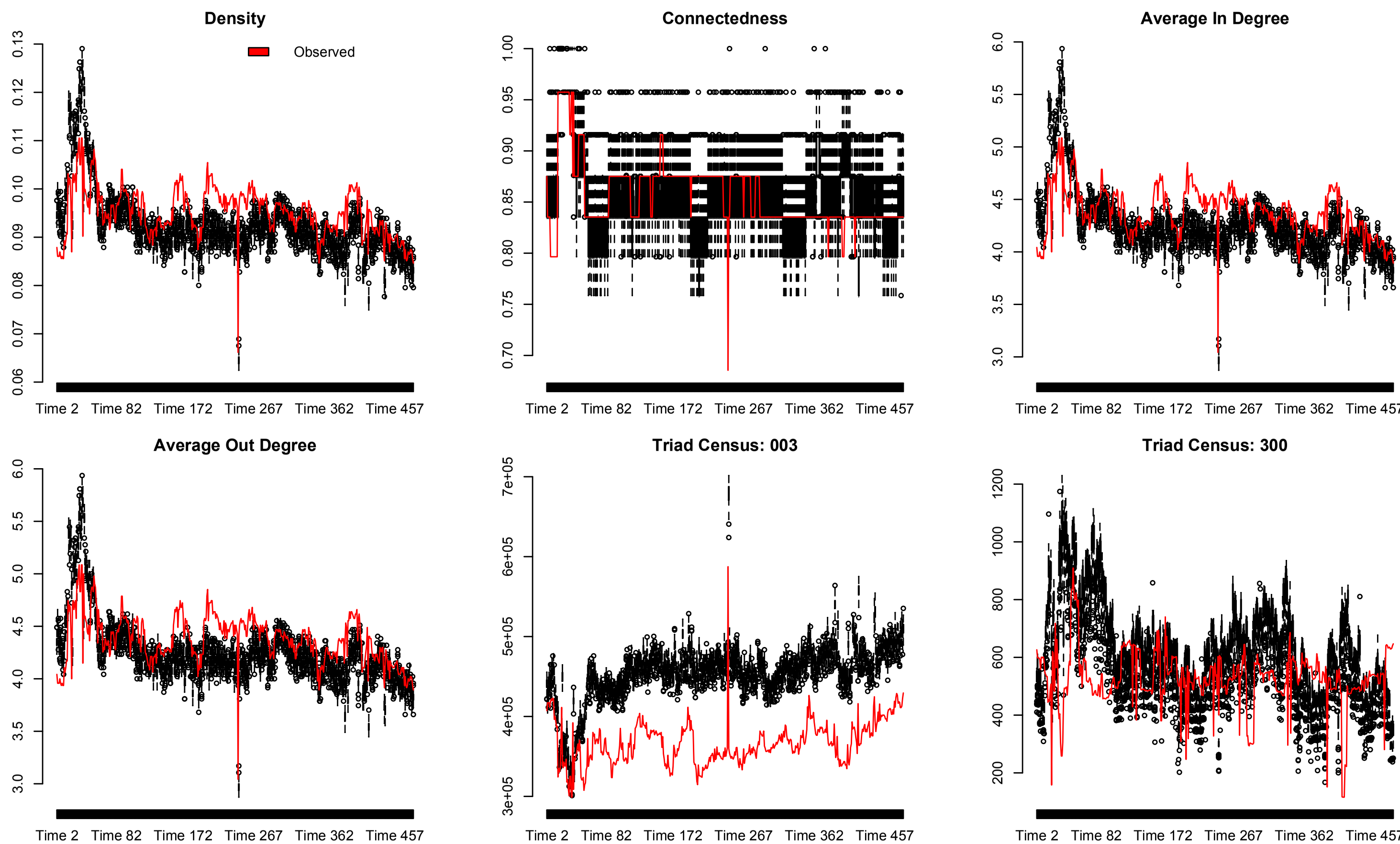

Figure 5: Graph Level index comparison for one-step one-lag network logistic regression for Model 4 with 100 simulations at each time point.

Reciprocity is less likely than Between-Group Reciprocity. We interpret this as a type of warring between the two groups, i.e. a citation from one group needs to be acted upon with reciprocal discussion.

We confirm our Seasonality Hypothesis 1 that inertia is strong and persistent effect in this network. We reject Seasonality Hypothesis 2 for hourly effects, since Model 5 has a lower BIC than Model 6 and thus is not informative and by assumption influential in the decision making process. Lastly, we obtain mixed results for our Selective salience hypothesis (remembering that we have to interpret these terms in relation to the lag term). We see the largest propensity for edges creation during PreCon and the DNCCon and the rest of the time there is rather stable effect for the periodicity terms except during the election when there is a noticeable decrease in edge creation, which largely disagrees with our hypothesis. It appears that instead of increased payoff for "important" events there is a noticeable decrease if the event is large enough (i.e., the election itself).

7 Discussion and Conclusion

In this work we have modeled the evolution of the network as a logistic-choice process which views actors as able to choose whether to cite a given blog or not (similar in concept to Snijders, 2001). To do this we employ a type of lagged logistic network regression (see Almquist and Butts, 2010) to handle a large dynamic network of blog-blog citations between RNC and DNC designated blogs. We assume that these two groups should be viewed as competing factions, and derive a number of balance-theoretic mechanisms to inform our logistic-choice model. Finally, we propose a series of basic hypotheses, based on group mixing and balance-theory, which we test exhaustively through model selection and parameter estimation.

First, we proposed a series of potential mechanisms which might inform our actors choices in their binary choice and then we fit the models and perform model selection via BIC and model validation via simulation analysis. This provided us with the model we used to test our hypotheses. This model provides the basic structure for our logistic-choice model and provides us with a measurement of our mechanisms effects on this choice process.

In conclusion, we find that the RNC designated blogs have a larger propensity for ingroup citation than DNC designated blogs. This suggests that the RNC blogs are more cohesive than DNC blogs, however both groups have a large propensity for within group mixing and lower propensity for across group mixing which is similar to findings in Hargittai et al. (2008) . We also confirmed the balance-theoretic In-Group Two-Path (friend of a friend) hypothesis, and Cross-Group Two-Path (friend of my enemy) hypothesis for this data-set.We find that reciprocity is more likely to occur across-groups then within groups suggestion that cross citation is a form of inter-blog warring. Lastly, we find that there are strong periodicity in this data-set centered around key-events in the election cycle; however, we find that the key-events are a time of low blogging/citation and in-between events are a time with increased propensity for linking.

References

Adamic, Lada and Natalie Glance. 2005. “The Poitcal Blogosphere and the 2004 US Election: Divided They Blog.” *LinkKDD 2005: Proceedings of the 3rd International Workshop on Link Discovery* pp. 36–43.

Almquist, Zack W. and Carter T. Butts. 2010. “Logistic Network Regression for Scalable Analysis of Dynamic Relational Data: an Overview and Case Study.” Working Paper: University of California, Irvine.

Ammori, Marvin. 2005. “A Shadow Government: Private Regulation, Free Speech, and Lessons from the Sinclair Blogstorm.” `http://www.mttlr.org/voltwelve/ammori.pdf`.

Anderson, Brigham S., Carter T. Butts, and Kathleen Carley. 1999. “The interaction of size and density with graph-level indices.” *Social Networks* 21:239–267.

Butts, Carter T. 2008. “Social network analysis: A methodological introduction.” *Asian Journal of Social Psychology* pp. 13–41.

Butts, Carter T., Ryan M. Acton, and Christopher Steven Marcum. 2010. “Interorganizational Collaboration In the Hurricane Katrina Response.” *Journal of Social Structure* .

Butts, Carter T. and B. Remy Cross. 2009. “Change and External Events in Computer-Mediated Citation Networks: English Language Weblogs and the 2004 U.S. Electoral Cycle.” *The Journal of Social Structure* 10.

Cartwright, Dorwin and Frank Harary. 1956. “Structural Balance: A Generalization of Heider’s Theory.” *Psychological Review* 63:277–293.

Cone, Edward. 2003. “The marketing of the PResident 2004.” *Baseline Magazine* .

Drezner, Daniel W. and Henry Farrell. 2008. “Blogs, politics and power: a special issue of Public Choice.” *Public Choice* 134:1–13.

Goffman, Erving. 1959. *Presentation of Self in Everyday Life*. Garden City, NY: Doubleday Anchor Books.

Hargittai, Eszter, Jason Gallo, and Mathew Kane. 2008. “Cross-ideological discussions among conservative and liberal bloggers.” *Public Choice* pp. 67–86.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. Hillsdale, London: Lawernce Erlbaum Associates, Publishers.

Holland, Paul W. and Samuel Leinhardt. 1981a. "An Exponential Family of Probability Distributions for Directed Graphs." *Journal of the American Statistical Association* 76:33–50.

Holland, Paul W. and Samuel Leinhardt. 1981b. "An Exponential Family of Probability Distributions for Directed Graphs: Rejoinder." *Journal of the American Statistical Association* 76:62–65.

Howard, Philip N. 2005. "Deep Democracy, Thin Citizenship: The Impact of Digital Media in Political Campaign Strategy." *The Annals of the America Academy of Political and Social Sciences* 597.

Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. 2008. "Goodness of fit of social network models." *Journal of the American Statistical Association* 103:248–258.

Kerbel, Mathew R. and Joel David Bloom. 2005. "Blog for America and Civic Involvement." *Harvard International Journal of Press/Politics* 10:3–27.

Krackhardt, David. 1987a. "Cognitive Social Structures." *Social Networks* 9:109–134.

Krackhardt, David. 1987b. "QAP Partialling as a Test of Spuriousness." *Social Networks* 9:171–186.

Krackhardt, David. 1988. "Predicting With Networks: Nonparametric Multiple Regression Analyses of Dyadic Data." *Social Networks* 10:359–382.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In *Frontiers in Econometrics*, edited by Paul Zarembka, pp. 105–142. New York, NY: Academic Press.

McFadden, Daniel. 1976. "The Mathematical Theory of Demand Models." In *Behavioral Travel-Demand Models*, edited by Peter R. Stopher and Arnim H. Meyburg. Lexington, MA: Lexington Books.

Morris, Martina. 1991. "A Log-Linear Modeling Framework for Selective Mixing." *Mathematical Biosciences* 107:349–377.

Petrescu-Prahova, Miruna. 2007. "Discrete Exponential Family Models for the Study of Ethnic Residential Segregation." In *Constructal Theory of Social Dynamics*, edited by Adrian Bejan and Gilbert Merkx. New York: Springer.

Rainie, Lee, Michael Cornfield, and John Horrigan. 2005. "The Internet and Campaign 2004." Technical report, The Pew Research Center for the People and the Press: Internet and American Life Project, Washington, DC.

Schwarz, Gideon E. 1978. "Estimating the dimension of a model." *Annals of Statistics* 6:461–464.

Shumway, Robert H. and David S. Stoffer. 2006. *Time Series Analysis and Its Applications*. New York, NY: Springer, 2nd edition.

Snijders, Tom A.B. 1996. "Stochastic Actor-Oriented Models for Network Change." *Journal of Mathematical Sociology* 21:149–172.

Snijders, Tom A.B. 2001. "The statistical evaluation of social network dynamics." In *Sociological Methodology*, edited by M.E. Sobel and M.P. Becker. Boston, pp. 361–395. London: Basil Blackwell.

Snijders, Tom A.B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36:99–153.

Snijders, Tom A.B. and M.A.J. Van Duijn. 1997. "Simulation for statistical inference in dynamic network models." In *Simulating Social Phenomena*, edited by R. Conte, R. Hegselmann, and P. Terna, pp. 493–512. Berlin: Springer.

Strauss, David and Michael Ikeda. 1990. "Psuedolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85:204–212.

Wall, Melissa. 2005. "Blogs of war." *Journalism* 6:153–172.

Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York: Cambridge University Press.

Wellman, Barry. 2001. "Computer networks as social networks." *Science* 293:2031–2034.

Woodly, Deva. 2008. "New competencies in democratic communication? Blogs, agenda setting and political participation." *Public Choice* 134.