Bayesian Dynamic Modeling and Analysis of Streaming Network Data

Xi Chen, Kaoru Irie, David Banks‡, Robert Haslinger, Jewell Thomas & Mike West

July 2015

Abstract

Traffic flow count data in networks arise in many applications, such as automobile or aviation transportation, certain directed social network contexts, and internet studies. Using an example of internet browser traffic flow through domains within a popular international news website, this paper presents Bayesian analyses of two linked models which, in tandem, allow fast, scalable and interpretable Bayesian inference. The first model is a flexible, non-stationary and non-Gaussian state-space model for streaming count data, able to adaptively characterize and quantify network dynamics effectively and efficiently in real-time. The second model is a time-varying gravity model that allows for closer and formal dissection of network dynamics. The former is fast and scalable, and maps to the second in a computationally trivial way to allow and interpret inferences on traffic flow characteristics, and on interactions among network nodes in particular.

KEY WORDS: Bayesian model emulation, Dynamic network flow, Gravity model, Internet traffic flows, On-line advertising, Parallel computing, State-space models

‡ Corresponding author: David Banks (E-mail: banks@stat.duke.edu).

Xi Chen and Kaoru Irie are PhD candidates, David Banks is Professor of Statistical Science, and Mike West is The Arts & Sciences Professor of Statistics & Decision Sciences, in the Department of Statistical Science, Duke University, Durham, NC 27708 (E-mail: xi@stat.duke.edu, kaoru@stat.duke.edu, banks@stat.duke.edu, mw@stat.duke.edu). Robert Haslinger is Senior Data Scientist and Jewell Thomas is Staff Data Scientist at MaxPoint Interactive Inc., Morrisville, NC 27560 (E-mail: Rob.Haslinger@maxpoint.com, Jewell.Thomas@maxpoint.com).
1.  INTRODUCTION

Increasing access to streaming data on large and evolving networks drives interest in formal models to quantify stochasticity and structure of latent processes underlying observable data streams. A key challenge is to define relevant statistical models that yield computationally efficient and scalable methods for streaming data, and that lead to sound statistical methods for monitoring, short-term prediction, single sample inference, and multi-sample comparisons across contexts.

We focus on flows of visitors to a set of domains of webpages in a structurally well-defined but dynamic/evolving network. The network is a subset of domains in the FoxNews website, monitored to generate streaming data on visitors to each domain over time. We use the term “domain” loosely, to refer to collections of webpages that correspond to categories of interest, as defined by Fox News.

On-line advertisers are interested in a host of statistical issues related to traffic flow and domain content. The field has become quite sophisticated, employing complex recommender systems (Koren et al. 2009), sentiment analysis (Pang and Lee 2008), text mining (Soriano et al. 2013), and other methods (Agarwal et al. 2010; Taddy 2013). However, basic questions of understanding and characterizing traffic across domains have not received the attention they require.

Trajectories of users, and groups of users, can give important information about browsing intent and the evolution of purchase interest, and thus ultimately affect ad placement decisions. As pages within a website are updated, questions arise as to whether browsing traffic patterns change as a result. To begin to address this statistically, we need to understand stochastic variation in past browser traffic so that comparisons can be made of incoming traffic streams against recent statistical “norms”, and significant deviations from short-term predictions based on current dynamic patterns can be identified.

We contribute to this area with an applied study that builds on two methodological advances. First, we develop a flexible, adaptive (non-stationary and non-Gaussian) state-space model for streaming count data. The analysis is explicitly designed to be computationally efficient in on-line data analysis; it scales quadratically in the number of network nodes, and is inherently parallelizable so enables distributed implementation for streaming data on increasingly large networks. We achieve this by (a) theoretically-based decoupling of models for individual network node-pair flows, and (b) adapting a computationally trivial univariate stochastic volatility model (e.g. West and Harrison 1997, Section 10.8) to apply to latent rates of underlying flows. Following a discussion of the data and network in Section 2, this Bayesian dynamic flow model (BDFM) is developed in Section 3. Some summaries of the analysis of the MaxPoint FoxNews data highlight its use in Section 4. A detailed Appendix summarizes key aspects of the underlying model theory and Bayesian analysis.

Second, we use the BDFM as an emulator of a dynamic gravity model (DGM). The DGM represents flows between network nodes with time-varying random processes for node-specific main effects and node-pair interaction effects. This is fundamentally interesting for understanding dynamics in network flows and node-node interactions. While fully Bayesian analyses of specific classes of gravity models have been developed (e.g. Jandarov et al. 2014; West 1994), these involve significant computational demands inherent in the use of Markov chain Monte Carlo and related methods. In any realistic dynamic extension appropriate for scalable, on-line analysis of streaming network flow data, such methods must be avoided.

While the models and analyses developed here represent new applications of Bayesian dynamic modeling ideas, there are at least conceptual intersections with studies of monitoring, inference and forecasting traffic flows in areas. Some of the developments here may well extend to apply to
such areas, including, for example, inference in origin-destination analysis (e.g. Tebaldi and West 1998) and physical traffic studies (e.g. Anacleto et al. 2013a,b; Queen and Albers 2009; Tebaldi et al. 2002).

Emulation analysis uses the unconstrained BDFM as a surrogate model, mapping posterior distributions to those of a coupled DGM. This exploits the statistical flexibility/adaptability and computational efficiency/scalability of the BDFM to create posterior inferences on the more highly structured and substantively focused DGM whose analysis is otherwise far more challenging. This is developed in Section 5, followed by example results and highlights from the MaxPoint FoxNews study in Section 6. Summary comments conclude the paper in Section 7.

2. Context and Data

Modeling internet traffic flow is a Big Data problem. We necessarily focus on smaller defined networks for which there is contextual knowledge to guide interpretation. Our context is traffic flow among “domains” of the FoxNews website. Besides the home/landing page, other domains include Politics, Entertainment, Travel, Science, and similar broad categories of news content. The contents of the domains change, usually on a daily basis at midnight, but more rapidly when some noteworthy event occurs. MaxPoint places ads on pages in these FoxNews domains, and thus can track flows of anonymized users as they move through its pages.

2.1 Data

The data set contains FoxNews website visit data during 09:00-10:00am and 01:00-02:00pm EST on each of six days, February 23-24, March 2-3 and 9-10, 2015. These days are Mondays or Tuesdays. Since the FoxNews website structure changes often, with new pages being added and old pages being archived, the analysis aggregates webpages into groups specified by the host domain www.foxnews.com, and the set of first url paths after the host domain, including examples such as e.g. www.foxnews.com/politics/* and www.foxnews.com/US/*. These classify all pages into 22 domains: Homepage, Politics, US, Opinion, Entertainment, Technology, Science, Health, Travel, Leisure, World News, Sports, Shows, Weather, Category, Latino, Story, On-Air, Video, National News, Magazine, and Other.

The data set includes anonymized users (browsers) from nearly every time zone on the planet. In order to study time-of-day effects, such as, say, a tendency to browse news in the morning and entertainment in the afternoon, it is necessary to stratify by time zone. Here we focus on users in the Eastern North America time zone; those are the most numerous, and the two time windows used in this study were chosen with the expectation that different browsing patterns might occur at those times.

We aggregate data to counts in half-minute intervals to form a time series with 120 time points showing domain occupancy, flows from each domain, and flows into each domain for each period. Several details are relevant. First, in each half minute interval, if the record shows the same user in two or more domains, then each of her/his moves is counted in the flow data into each of these domains. Second, if the user refreshes the same page multiple times spanning more than one time interval, then s/he is counted as simply staying in that domain; this can be done as the web browsing tool performs automatic refresh. Importantly, if a user stays in the same domain for more than five minutes, we assume s/he is no longer active, and is counted as leaving the FoxNews
site. If such a user later appears in one domain, s/he counts as in-flow from outside the FoxNews site. Finally, we cannot track user information either before or after the one-hour observation window, so there is a form of censoring; we thus restrict attention to the period 09:05 – 09:55am and 01:05 – 01:55pm, consisting of uncensored flows, using the first 5 minutes of data informally to define priors. Thus the series runs from $t = 1: T$ with $T = 110$ in each time period.

2.2 Network Structure and Notation

Referring to sites external to the FoxNews website as node 0, we have 23 network nodes; the $I = 22$ actual domains and “External”, indexed as $i = 0: I$. At each time $t = 1: T$, define $n_{it}$ as the number of occupants of node $i$, and define $x_{ijt}$ as the flow count from node $i$ to $j$, including the in-flows $x_{0it}$ and out-flows $x_{ij0}$ relative to the External domain. Figure 1 provides a visualization of data at the first time interval, and the schematic in Figure 2 reflects our notation.

3. Fast, Flexible Dynamic Modeling

The modeling starting point is conditionally independent Poisson ($Poi$) models for in-flows to the network coupled with conditionally independent multinomials ($Mn$) for flows from each node, all with time-varying parameters. Thus, for all $t = 1: T$, and all network nodes $i = 1: I$,

$$x_{0it} \sim Poi(\phi_{it}) \quad \text{and} \quad x_{i,0:t,t} \sim Mn(n_{i,t-1}, \theta_{i,0:1,t}),$$

and these are conditionally independent across nodes and over time. The conditioning is upon two kinds of (latent) process parameters: $\phi_{it}$, the time-varying rate process governing flows into node $i$ from sites external to FoxNews, and $\theta_{ijt}$, the time-varying probabilities of transitions from node $i$ to node $j$, where the latter includes $j = 0$ to represent leaving the FoxNews network. The BDFM defines the structure for these parameter processes.

3.1 Network In-flows: Poisson Dynamic Models

Begin with the in-flow rates $\phi_{it}$ of eqn. (1) for each node $i$. The BDFM characterizes time variation in $\phi_{it}$ with Markov models for time-varying gamma processes, used for over 30 years as a standard univariate stochastic volatility model (West and Harrison 1997, Section 10.8). The basic model links intimately with “steady” evolutions in non-Gaussian dynamic models (Smith 1979) and yields a closed form Bayesian analysis. The resulting forward filtering analysis is effectively a fully Bayesian framework for one-sided (forward in time) exponential smoothing, with past data being discounted at a per unit rate defined by a specific discount factor. The retrospective analysis—backward simulation and smoothing—updates the results of this forward filtering to generate full posterior distributions for the rates over time, with point estimates generated as fully Bayesian, two-sided exponential smoothers. Full posterior uncertainties are properly characterized.

The Appendix lays out the complete model and analysis details; for $\phi_{it}$, we simply add the suffix $i$ to the Appendix notation and replicate the analysis, decoupled and in parallel, across nodes $i$. The dynamic gamma-beta evolution model is $\phi_{it} = \phi_{i,t-1} \eta_{it} / \delta_i$ where the $\eta_{it}$ are independent random “shocks” with specific beta distributions, and $\delta_i \in (0, 1)$ is a fixed, node-specific discount parameter. The beta innovation distribution depends on $\delta_i$ and this dependence implies, among other things, the random walk nature of the model; this is reflected in the resulting conditional
Figure 1: A snapshot of counts and flows on the FoxNews network at time \( t = 1 \) (09:05.30 on February 23, 2015). The circular numbered nodes represent the domains, with diameters and color proportional to occupancy \( n_{it} \) for node \( i \) at this time point \( t \). Each arrow \( i \rightarrow j \) has width proportional to flow \( x_{ijt} \). The supplemental material at the supporting website includes a video showing the evolution of this dynamic network snapshot over the full time period. The layout used node-node distances based on the Fruchterman-Reingold algorithm in the igraph R package.

Figure 2: Network schematic and notation for flows at time \( t \).
mean \( E(\phi_{it}|\phi_{i,t-1}) = \phi_{i,t-1} \); i.e., change is allowed though directional change is not anticipated. As detailed in the Appendix, analysis of the in-flow data \( x_{0it} \) over time is now performed by analytic updating/filtering, followed by retrospective simulation, of the gamma prior and posterior distributions.

Now extend the in-flow model notation to include a latent prior rate \( \phi_{i0} \), and correspondingly extend the in-flow data notation so that \( x_{0it} \) represents information at \( t = 0 \) used to define a gamma prior for \( \phi_{i0} \). The following simply summarizes key details, as given in the Appendix.

1. **Forward Filtering (FF):** For each \( t = 1:T \) and independently across nodes \( i = 1:I \), the online, time \( t \) posterior \( p(\phi_{it}|x_{0i,1:t-1}) \) is gamma with defining parameters evaluated from past information \( x_{0i,1:t-1} \), and trivially updated over time.

2. **One-Step Forecasts:** The one-step ahead forecast distribution made at time \( t - 1 \) to predict time \( t \) in-flow is negative binomial with p.d.f. \( p(x_{0it}|x_{0i,1:t-1}, \delta_i) \) of Appendix eqn. (8).

3. **Model Marginal Likelihood (MML):** The node \( i \) model marginal likelihood at time \( t \), namely \( p(x_{0i,t}|x_{0i0}, \delta_i) \), is available and trivially updated over time, from Appendix eqn. (9).

4. **Backward Sampling (BS):** At end time \( T \), trivial recursive simulation generates time trajectories \( \phi_{i,1:T} \) of the rate process under its full posterior \( p(\phi_{i,1:T}|x_{0i,1:T}) \).

The node-specific MML measures at time \( T \) feed into model assessment including inference on, and selection of, the discount factor \( \delta_i \). They can also be exploited over time to sequentially monitor the analysis as data streams into the study, providing on-line tracking of model performance, with potential uses in flagging anomalous, node-specific in-flow patterns, such as those associated with some news event.

### 3.2 Transitions from Network Nodes: Multinomial Dynamic Models

Consider now the transition probabilities \( \theta_{i,0:I,t} \) of eqn. (1) for each node \( i \). A model in which the prior for \( (\theta_{i,0:I,t}|x_{i,0:I,1:t-1}) \) is a conjugate Dirichlet is the key to applying the same univariate volatility models here: we construct Dirichlet random variates from sets of independent gamma variates, and this opens the path to the model extension. Inductively, suppose at time \( t \) and independently across nodes \( i \) that the posteriors for transition probabilities are Dirichlet, \( \theta_{i,0:I,t}|x_{i,0:I,1:t} \sim Dir(a_{ij},q_{i0:I,t}) \) where the probability vector \( q_{i0:I,t} \) is the posterior mean and \( a_{ij} > 0 \) is the total mass, or precision. Dirichlet distribution theory implies that there are \( I + 1 \) underlying latent rates \( \phi_{i,0:I,t} \) such that: (a) for each \( j = 0:I \), \( \theta_{ijt} = \phi_{ijt}/\sum_{k=0:I} \phi_{ikt} \), and (b) \( \phi_{ijt} \sim \text{Ga}(r_{ijt},c_{ijt}) \) independently over \( j = 0:I \), where \( r_{ijt} = a_{ij}q_{ijt} \) and \( c_{ijt} > 0 \). Note that the \( a_{ij}, c_{ij} \) depend on the node \( i \) from which the flow originates, but not upon the receiving node \( j \).

Extend the model notation to include a latent prior set \( \theta_{i,0:I,0} \), and correspondingly extend the transition data notation so that \( x_{i,0:I,0} \) represents information at \( t = 0 \) used to define the prior for \( \theta_{i,0:I,0} \). Then, over time \( t \) we can apply the gamma-beta process model just as used for the in-flow rates. That is, based on a node-specific discount factor \( \rho_i \) (the transition probability analogue of the in-flow discount factor \( \delta_i \)), we model the latent processes \( \phi_{ijt} \) underlying the transition probabilities via individual gamma-beta BDFMs. All of the theory and resulting algorithms detailed in the Appendix then apply, simply by adding the double subscripts \( ij \) to the notation of the Appendix. Details of the resulting FFBS analyses are left to the reader as they replicate those of the Appendix.
Within this analysis, note that we have access to node $i \to j$ specific MML measures that can be exploited in sequential monitoring and anomaly detection, as noted for the in-flow models.

The one key feature requiring additional discussion relates to mapping back to the inherent Dirichlet/Multinomial structure for forecasting and computation of marginal model likelihoods. From the gamma-beta BDFMs at time $t - 1$, the gamma priors for each $\phi_{ij}$ imply a Dirichlet prior $\theta_{i,0,1,t}|x_{i,0,1,t-1} \sim \text{Dir}(\rho_i r_{i,0,1,t-1})$ with $r_{ij,t-1} = a_{i,t-1} q_{ij,t-1}$ for each $j = 0:I$. Coupled with the multinomial model in of eqn. (1), this yields a one-step ahead forecast p.d.f. of the multivariate Polya form, namely

$$p(x_{i,0:1,t}|x_{i,0:1,t-1}, \rho_t) = \frac{\left(\sum_{j=0:I} x_{ij,t}\right)! \Gamma\left(\sum_{j=0:I} \rho_i r_{ij,t-1}\right)}{\sum_{j=0:I} x_{ij,t}! \Gamma\left(\sum_{j=0:I} (\rho_i r_{ij,t-1} + x_{ij,t})\right)} \prod_{j=0:I} \frac{\Gamma(\rho_i r_{ij,t-1} + x_{ij,t})}{\Gamma(\rho_i r_{ij,t-1})}.$$ (2)

These node- and time-specific measures can be monitored, and then combined in the overall MML for node $i$ at time $t$, i.e.,

$$p(x_{i,0:1,t}|x_{i,0:1,0}, \rho_t) = \prod_{s=1:t} p(x_{i,0:1,s}|x_{i,0:1,s-1}, \rho_t).$$ (3)

At end time $t = T$, this provides the value of the marginal likelihood at the specific chosen value of $\rho_t$. Repeating the analysis over a grid of values coupled with a specified prior for $\rho_t$ then yields a posterior $p(\rho_t|x_{i,0:1:T})$ for assessment of relevant values of the discount factor specific to node $i$.

4. BDFM Analysis of FoxNews Data

The analysis was applied separately to data from each of the six days. We focus on the first day (February 23 2015) for initial summaries, and then make some comparisons across days. Each analysis used priors as follows. The gamma priors for the in-flow rates are $\phi_{i0}|x_{i0} \sim Ga(r_{i0}, c_{i0})$ with $c_{i0} = 1$ and $r_{i0} = c_{i0} z_i$ where $z_i$ is a point estimate based on in-flows in the first 5 minutes of data prior to $t = 1$. The Dirichlet prior for the transition probabilities $\theta_{i,0,1,0}|x_{i,0,1,0} \sim \text{Dir}(a_{i0} q_{i0,1,0})$ where the prior means $a_{i0} q_{i0,0}$ are simple point estimates based on in-flows in the first 5 minutes of data prior to $t = 1$, and the $a_{i0}$ are determined so that $q_{i0,1} = x_{i0,1}$ is a probability vector. The priors for the underlying, unconstrained node-node flow rates are then $\phi_{ij0}|x_{ij0} \sim Ga(r_{ij0}, c_{ij0})$ with $c_{ij0} = 1$ and $r_{ij0} = a_{i0} q_{ij0}$.

Regarding sensitivity to this prior specification, we note that the shape parameter of Gamma distribution, $c_{ij0}$, could be smaller, so that the prior variance for the flow counts is larger. However, this choice has little impact upon the posterior analysis, especially when this parameter is less than one. As seen in the posterior updating rule in Section 3, even a single observation “washes out” the effect of the hyperparameter from the posterior.

Priors for each discount factor were $Be(19, 1)$ distributions truncated to $(0.9, 1)$; reanalysis using priors on this range led to little in the way of noticeable differences, as the marginal likelihoods at $t = T$ dominate. The prior truncation ensures some degree of smoothness. Running the models in parallel across discrete grids of discount factors and evaluating the MML measures at $t = T$ gave marginal likelihoods which were mapped to posteriors. Across all in-flows, this suggested $\delta_i \approx 0.9$ as a posterior mode. Figure 3 plots posteriors truncated at 0.9 for the discounts $\rho_i$ in the transition flow models. Some nodes exhibit higher volatility in flows to other nodes over time than others, requiring and hence favoring smaller discount factors; these are particularly
associated with nodes representing domains with high flow counts (e.g., in-flows from External to Homepage).

Some summary inferences on selected model components are reported, based on models with discount factors fixed at their posterior modes. Figure 4 gives one example of learning about in-flow rates, in this case the flow from all nodes to node 10 (the Leisure domain). The figure exemplifies sequential learning about the flow rate together with its retrospectively updated trajectory and a visual assessment of one-step ahead forecasting aligned with the data.

A similar display in Figure 5 highlights the same aspects of the analysis, now with an example of flow between two network nodes (from Homepage to the Politics domain). It shows the rate between two nodes together with the retrospective smoothing for full inference on the trajectory and one-step ahead forecast summaries.

Homepage is the most popular single domain on FoxNews, so the transition probabilities from Homepage to other domains are of particular interest. Figure 6 shows that most Homepage visitors stay on Homepage for a while. Of those that leave, many exit the FoxNews site entirely. Across all six days, the probability of staying on the Homepage decreases over the course of the 50 minute morning period.

On examples of transition probabilities, Figure 7 shows that the probability of people leaving the FoxNews website from Homepage increases in this 50 minute window for each of the six mornings. Note that there are significant day effects; e.g., visitors are more likely to leave FoxNews on the morning of March 9th compared to the other mornings. More detailed insights, based on the gravity model, are noted in the next section.

As an illustration of a more detailed analysis of a very specific flow, consider Figure 8. Among the visitors who leave the Homepage for other FoxNews domains, Entertainment is generally the
Figure 4: BDFM-based inference over time $t$ on in-flows to domain $i = 10$ (Leisure). Upper: data $x_{0,t}$ (circles) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{0,t} | x_{0,1:t})$ plotted against $t$. Lower: revised trajectory under full retrospective posterior $p(\phi_{0,t} | x_{0,1:T})$.

Figure 5: BDFM-based inference over time $t$ on transitions from domain $i = 1$ (Homepage) to $j = 2$ (Politics). Upper: data $x_{12,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{12,t} | x_{12,1:t})$ plotted against $t$. Lower: revised trajectory under full retrospective posterior $p(\phi_{12,t} | x_{12,1:T})$. 
Figure 6: Retrospective mean and 95% CI of trajectories of transition probability \( \theta_{11t} \) (staying at Homepage) from analysis on data collected from each of the six mornings.

Figure 7: Retrospective mean and 95% CI of trajectories of transition probability \( \theta_{10t} \) (Homepage \( \rightarrow \) External) from analysis on each of the six mornings.
most popular destination. For the six datasets collected during the morning, we see large differences in transition probabilities; in particular February 23 and 24 have larger rates than the other days. It is noteworthy that the Academy Awards ceremony was held on the night of February 22, which may have driven this uptick.

![Graph showing transition probabilities for different dates](image)

Figure 8: Retrospective mean and 95% CI of trajectories of transition probabilities \( \theta_{1st} \) (Homepage → Entertainment) for each of the six mornings.

5. Emulation-Based Mapping to Dynamic Gravity Models

5.1 Gravity Models

We now consider a more nuanced model that structures flow rates in terms node-specific main effects and node-node interaction terms. For each within-network node \( i = 1:I \) and all \( j = 0:I \), suppose that the rate process parameter in the BDFM is now

\[
\phi_{ijt} = \mu_t \alpha_{it} \beta_{jt} \gamma_{ijt} \tag{4}
\]

with the following terms: an baseline network flow level \( \mu_t \); an “origin” node \( i \) main effect \( \alpha_{it} \), representing the adjustment to baseline in the intensity of flows from node \( i \) at time \( t \); a “destination” node \( j \) main effect \( \beta_{jt} \), measuring the additional “attractiveness” of node \( j \) at time \( t \) relative to baseline; and an interaction term \( \gamma_{ijt} \), representing the directional “affinity” of node \( i \) for \( j \) at time \( t \) relative to the combined contributions of baseline and main effects.

Models of this and more elaborate forms have been popular in transportation studies (e.g. Sen and Smith 1995; West 1994) where the interaction term is typically structured as a function of physical distance between nodes; there the “gravity model” terminology relates to the role of small distances in defining large interactions and hence “attraction” of traffic from node \( i \) to node \( j \). We
refer to the $\gamma_{ijt}$ interactions as “affinities”. In dissecting the network flow activity, we are most interested in questions about which affinities are greater than one ($j$ attracts flow from $i$ over and above the main effects), or less than one ($j$ is relatively unattractive to $i$), or not significantly different to one (neutral). Critically, affinities are time-varying, and any identified patterns of variation over time may be related to interpretable events or network changes.

In the first fully Bayesian approach to gravity models using MCMC methods, West (1994) developed such models in the static case; i.e., with no dynamics in the model parameters, and applied the model to a large transportation flow network. Congdon (2000) explored a similar approach in studies of patient flows to hospitals. Analysis via MCMC is computationally very demanding, and the burden increases quadratically in $I$, and inherently non-sequentially. More recently, Jandarov et al. (2014) studied such models for spread of infectious diseases, and used Gaussian process approximations for approximate inference rather than full MCMC or other computational methods.

We share the spirit of this latter work, in using the simply and efficiently implemented BDFM as a path to fitting the gravity model—now extended to time-varying effect parameter processes. However, we do not constrain the affinity parameters $\gamma_{ijt}$ as a function of covariates of any kind, simply treating the DGM as a dynamic, random effects model. This leads to a direct parameter mapping between the BDFM to the DGM; as a result, the trivially generated simulations from the full posterior of the BDFM are mapped directly to full posterior samples from the DGM, providing immediate access to inference on main effect and affinity processes over time.

### 5.2 Model Mapping

Given a set of flow rates $\phi_{ijt}$ for all $i = 1:I, j = 0:I$, at each time $t = 1:T$, the mapping to DGM parameters in eqn. (4) requires aliasing constraints to match dimensions. We adopt the common zero-sum constraint on logged values. Define the log parameters $m_t = \log(\mu_t), a_{it} = \log(\alpha_{it}), b_{jt} = \log(\beta_{jt})$ and $g_{ijt} = \log(\gamma_{ijt})$. Using the $+$ notation to denote summation over the range of identified indices, we constrain such that $a_{i+t} = b_{+t} = 0, g_{+jt} = g_{i+t} = 0$ for all $i, j, t$. We then have a bijective map between BDFM and DGM parameters and, given the $\phi_{ijt}$ in the former, can directly compute the DGM parameters.

These constraints ensure identifiability, but even so, the model is saturated. There are exactly as many parameters in the DGM as there are observations in the data set. But the forward filtering and backwards sampling enforce smoothness over time that substantially reduce the effective degrees of freedom required for the model. This is another advantage of using the BDFM emulator to represent the DGM.

To compute the DGM parameters, define $f_{ijt} = \log(\phi_{ijt})$ for each $i = 1:I, j = 0:I$, at each time $t = 1:T$. Then at each time $t$, compute the following in the order given:

1. the baseline level $\mu_t = \exp(m_t)$ where $m_t = f_{i+t}/(I + 1)$;
2. for each $i = 1:I$, the origin node main effect $\alpha_{it} = \exp(a_{it})$ where $a_{it} = f_{i+t}/(I + 1) - m_t$;
3. for each $j = 0:I$, the destination node main effect $\beta_{jt} = \exp(b_{jt})$ where $b_{jt} = f_{j+t}/I - m_t$;
4. for each $i = 1:I$ and $j = 0:I$, the affinity $\gamma_{ijt} = \exp(g_{ijt})$ where $g_{it} = f_{ijt} - m_t - a_{it} - b_{jt}$.

In our data analysis below, we apply this to all simulated $\phi_{ijt}$ from the full posterior analysis under the BDFM to map to posteriors for the DGM parameter processes.
A technical problem with this mapping arises in cases of sparse flows, i.e., when multiple \( x_{ijt} \) counts are zero or very small for multiple node pairs. In such cases the posterior for \( \phi_{ijt} \) favors very small values and the log transforms are large and negative, which unduly impacts the resulting overall mean and/or origin or destination means. While one can imagine model extensions to address this, at a practical level it suffices to adjust the mapping as is typically done in related problems of log-linear models of contingency tables with structural zeros (Bishop et al. 1975, chap. 5). This is implemented by simply restricting the summations in the identifiability constraints to those node pairs for which \( x_{ijt} > d \), for some small \( d \), and adjusting the divisors to count the numbers of terms in each summation. For this study, we use \( d = 3 \). With this adjustment, very small \( \phi_{ijt} \) appropriately lead to very small affinities \( \gamma_{ijt} \), so the latter then represent very sparse flows.

6. DGM Analysis of FoxNews Data

6.1 February 23 2015, 09:00-10:00am

We first apply the gravity model decomposition to the morning data on February 23rd. Each flow is analyzed by the dynamic Poisson-Gamma and Multinomial-Dirichlet models independently, under the same settings used in Section 4. The particles of Poisson rate parameters sampled from their posteriors are decomposed into the gravity model components, which allows us to construct the posteriors for the gravity model parameters.

Figure 9 shows the retrospective posterior of the baseline level \( \mu_t \). Its posterior mean is almost stable around 0.66 throughout the fifty minute period and its fluctuation is at most 0.02, or a 2% increase of its effect to the total access counts.

![Figure 9: DGM-based smoothed trajectory of baseline level process \( \mu_{1:T} \). In this and following figures, the dashed lines indicate 95% intervals about the displayed posterior mean trajectory.](image)

The origin and destination effects are shown in Figures 10 and 11. The posteriors for origin effects show that large-scale domains, such Homepage (domain 1), have higher values of \( \alpha_{it} \), while
Figure 10: DGM-based smoothed trajectories of node-specific outflows $\alpha_{i,1:T}$.

Figure 11: DGM-based smoothed trajectories of node-specific inflows $\beta_{j,1:T}$.
domains with low or zero flows, such as Shows (domain 7), naturally have lower values. The two graphs show similar patterns in many domains, but differences are also apparent. In particular, the posterior analysis for several domains, such as Science (domain 7), Health (domain 8) and Video (domain 19), shows “significance” in their origin effects but “insignificance” in their destination effects (i.e., their 95% Bayesian credible intervals of the destination effects contain one, but this is not case for the origin effects). These distinctions between the two effects show the roles of $\alpha_{it}$ and $\beta_{jt}$; they represent the common factors across the origin and destination of the flows, which is confirmed by the differences of the two results. These effects are also essential in order to capture the scale of the domains, by having similar patterns for $\alpha_{it}$ and $\beta_{jt}$ when $i = j$.

For the affinity effects $\gamma_{ijt}$, we have $(I + 1)^2 - 1$ parameters (one for each pair of nodes except the unobserved External $\rightarrow$ External flow) at each time $t$. The number of effects becomes massive for large $I$. Even in this example for illustration, $I = 22$, the number of $\gamma_{ijt}$ for fixed $t$ is 528, so it is impossible to examine all the results in this paper. For this reason, we pick up a few affinity effects that may interest readers in terms of interpretation. For affinity $\gamma_{ijt}$ with retrospective posterior c.d.f $\Phi_{ijt}(\gamma)$, we use the Bayesian credible value $p_{ijt} = \min\{\Phi_{ijt}(1), 1 - \Phi_{ijt}(1)\}$ as a simple numerical measure of deviation from the “neutral” value of 1. This highlights the practical relevance of the affinity effect and its changes over time.

First, we focus on the flows from Homepage (domain 1). Those flows are crucial in understanding the user’s preference from the aggregated data since Homepage is usually the landing page for visitors. Flows from it must have information on which domain the user wants to access first, which would be an important finding in advertisement and marketing. Fig. 12 shows that the affinity trajectory and credible values for flows from Homepage (domain 1) to Science (domain 7) are entirely less than 1. This implies that visitors to the FoxNews landing page tend not to access articles in the Science domain (during this time period). In contrast, Fig. 13 displays the affinity trajectory and credible values for the flow from Homepage to Opinion (domain 4). The affinity effect is significantly positive for much of the time interval, implying that at the beginning of the hour, people on the FoxNews landing page tend to check the websites within the Opinion domain. But, as time passes, the trajectory gradually decreases to 1 and becomes insignificant. This change of significance is clear in the Bayesian credible values, which are almost zero at first but suddenly begin to increase at around 9:40-9:45am and quickly exceed the reference line 0.05.

The change in significance of the Homepage $\rightarrow$ Opinion affinities shows that the DGM has the flexibility to detect time-varying sparsity in parameters. This phenomenon is peculiar to time series analysis and modeling, and the example demonstrates that such sparsity actually exists in the dataset. It also has important implications for on-line advertising, as it shows that users have different interests at different times of the day.

### 6.2 Comparison Across Days

The FoxNews dataset covers both the morning (09:00-10:00am) and the afternoon (01:00-02:00pm) period on each of the 6 days. We have already discussed a range of comparisons, and differences, across days for the morning periods in Section 4. Moving to the DGM, we now explore additional features concerning time-of-day effects as well as day-to-day variation. This is based on running the coupled BDFM-DGM analysis separately on each time period/day.

Figure 14 shows the DGM trajectories for the retrospective baseline parameter process $\mu_{1,T}$ for each of the 12 fifty-minute intervals. Trajectories are similar across days but for notable differences on February 24 and March 3. On February 24, the afternoon flow is significantly lower than the
Figure 12: Upper: DGM-based smoothed trajectories of transition affinities $\gamma_{17,1:T}$, Homepage $\rightarrow$ Science. Lower: Corresponding Bayesian credible values.

Figure 13: Upper: DGM-based smoothed trajectories of transition affinities $\gamma_{14,1:T}$, Homepage $\rightarrow$ Opinion. Lower: Bayesian credible values corresponding to the affinity trajectories.
morning flow, while the morning flow that day is much larger than across other days. One plausible reason is increased morning traffic in response to discussions following the Academy Awards ceremony, with a resulting lull in the afternoon traffic. The reverse happens on March 3 where, although the morning traffic seems typical, the afternoon traffic is unusually high. This was the day on which FoxNews posted an article concerning Hillary Clinton’s use of her personal email account for all correspondence during her tenure as Secretary of State. It is plausible that this led to larger than usual afternoon traffic flows as the controversy unfolded.

![Figure 14: DGM-based inference on baseline flow level trajectories for all six days, with 95% credible intervals. The red trajectories correspond to the 09:05-09:55am time window, and the blue trajectories correspond to the 01:05-01:55 p.m. time window.](image)

One of the advantages of the DGM representation is that it allows an easy path to check these speculative explanations. For example, examination of the destination flow effects $\beta_{5.1:T}$ show that the Entertainment node was unusually popular on the morning of February 24, and that the Politics and Opinion nodes were unusually popular on the afternoon of March 3, compared to similar flows on other days.

7. Closing Comments

The BDFM framework is adaptive to time-varying rates of flows within dynamic networks and able to coherently quantify non-stationary changes in within- and into-/out of- network flow rate processes. The sequential analysis of this Bayesian dynamic flow model is fast and efficient; computational demands scale linearly in time and quadratically in node number. Importantly, this almost semi-parametric approach generates a parallelizable analysis yielding full posterior distributions for underlying rate parameter process parameters across nodes and pairs of nodes in a scalable
manner. While the model inherently reflects the complexity of interactions among the full set of nodes, and their changes over time, the BDFM approach “decouples” analysis to individual nodes and pairs of nodes, and “recouples” them (via the map from decoupled gamma to recoupled Dirichlet posteriors over time) for formal inferences. Our analysis of the FoxNews network time series data sets shows the utility of the BDFM in generating initial inferences on flow rate processes, in highlighting differences across days and in generating potential practical “leads”. On the latter, for example, it is immediately clear from the BDFM results that most visitors go to just one domain, rather than traversing to multiple domains. This has potential decision implications for computational advertising, and also likely highlights a difference between on-line news consumers and traditional newspaper readers.

The Bayesian emulation “map” from the BDFM to the dynamic gravity model represents a modelling/computational strategy of increasing interest in many areas, and especially in emerging Big Data applications. That is, we fit a flexible, adaptive model in a set of (conditionally) decoupled analyses, and then directly map posterior samples to the more substantively interesting and interpretable parameter processes in a model, the DGM, that is otherwise challenging to fit. Applied to the FoxNews flow data, we see that this indicates “time-varying sparsity” in node-node interaction effects over time, nicely captured by the DGM. Our use of Bayesian credible values over time is one nice way of focusing attention on this, allowing us to highlight the “significance” of inferred DGM interactions across time. Interestingly, many of the interaction effects (affinities) appear significant at some points in time but not in others. A number of the specific node-node inferences mentioned in the application section highlight additional results of substantive interest, some of which are initially unexpected. These include, for example, the sustained positive affinity of Opinion for Homepage, but a similarly sustained but negative affinity of Science for Homepage. Additionally, comparisons across different times of the day identified and quantified patterns related to anomalous flows corresponding to identifiable news events that appear to have driven traffic to specific nodes on the FoxNews site.

From this point, we see opportunities to now develop these models as a basis to characterize the stochastic dynamics of website flows, and hence feed into modeling and decision analysis that addresses the needs to respond to changing patterns in computational advertising. An ability to rapidly signal potential anomalies in a small subset of domains in real-time will be of huge interest in this field. We also note, as remarked in the introduction to the paper, the potential connections with problems of physical traffic flows, origin-destination problems, and other kinds of dynamic network studies including social networks, capital flows between financial institutions, electrical power grids, and others.

More immediately, some of the evident questions arising from the current study concern the overlay of the “unbiased” inferences about changes and structure in network flows with substantive covariate information. In many applications, including computational advertising but also capital and transportation flows, there are useful covariates that could inform the analysis. Our perspective here has been more exploratory, aiming to define a formal basis for effectively characterizing non-stationary stochastic dynamics in flow data. A next step is to overlay any particular application with covariate information as descriptive/explanatory as we exemplified with some vignettes from the FoxNews study. At a more predictive level, the DGM is naturally extensible to incorporate covariates— in main effects and/or interaction terms— so that some consideration of how to extend the flexible, computational efficient and scalable BDFM-DGM in that direction is warranted.
Appendix

This Appendix summarizes the context and details of the gamma-beta “steady” dynamic model for time-varying Poisson rates, which is the basic univariate model that underlies the BDFM. This adapts the venerable stochastic volatility model for time-varying normal variances to the new context of Poisson rates. Background on the volatility literature, as well as theory and references, can be accessed in West and Harrison (1997, Section 10.8) as well as in Prado and West (2010, Section 4.3.6 and Section 4.6, problem 4).

Using generic notation, a series of non-negative counts $x_t$ over $t = 1:T$ is modeled via $x_t | \phi_t \sim \text{Poi}(\phi_t)$ conditionally independently over time, where the underlying/latent Poisson rate process $\phi_t$ follows a gamma-beta stochastic model. This is effectively a non-stationary, non-Gaussian random walk model, so it has enormous flexibility in adapting to changes over time. The extent of anticipated stochastic change over time is defined by a single discount factor parameter $\delta \in (0, 1)$.

We detail the model concept and structure, and the implied machinery for Bayesian learning and forecasting that includes the forward filtering, backward sampling (FFBS) algorithm for conditionally Poisson time series coupled with the gamma-beta steady process model.

A1. Forward Filtering (FF)

At time $t = 0$, introduce an hypothetical latent state $\phi_0$ and use $x_0$ as a synthetic notation for all available initial information. Specify an initial gamma prior, so $\phi_0 \sim \text{Ga}(r_0, c_0)$ with log p.d.f. given by a constant plus $(r_0 - 1) \log(\phi_0) - c_0 \phi_0$ on $\phi_0 > 0$, where $r_0 > 0, c_0 > 0$ are known.

For each $t = 1:T$, the model and forward/sequential analysis are then as follows.

Posterior at time $t - 1$: Standing at time $t - 1$, the posterior for the current Poisson rate given the initial information and all data observed over past times $1:t - 1$ is gamma,

$$\phi_{t-1} | x_{0:t-1} \sim \text{Ga}(r_{t-1}, c_{t-1}) \quad (5)$$

where the defining parameters are known, evaluated from past information $x_{0:t-1}$.

Evolution to time $t$: The Poisson rate evolves to time $t$ via the gamma-beta evolution

$$\phi_t = \phi_{t-1} \eta_t / \delta, \quad \eta_t \sim \text{Be}(\delta r_{t-1}, (1 - \delta) r_{t-1}), \quad (6)$$

where the random “shock”, or innovation, $\eta_t$ is independent of $\phi_{t-1}$. This is a multiplicative random walk model in that $E(\phi_t | \phi_{t-1}) = \phi_{t-1}$, hence the use of the “steady model” terminology. A lower value of $\delta$ leads to a more diffuse beta innovation distribution and the ability to adapt to changing rates over time, while a value closer to one indicates a steady, stable evolution. The random walk nature of the model allows for changes, but does not anticipate specific directional changes. The model results in a fully Bayesian solution to rather simple, flexible smoothing of discrete time series in the context of variation in the underlying latent process.

Prior for time $t$: The time $t - 1$ gamma posterior of eqn. (5) couples with the beta innovation to give the time $t - 1$ prior for the next state as

$$\phi_t | x_{1:t-1} \sim \text{Ga}(\delta r_{t-1}, \delta c_{t-1}) \quad (7)$$
Here we see the discounting effect of the random walk model: the prior for the evolved rate is more diffuse than the time $t-1$ posterior, reflecting increased uncertainty due to evolution.

**One-step ahead predictions:** Predicting the data at time $t$, the one-step ahead forecast distribution is negative binomial with p.d.f.

$$p(x_t|x_{1:t-1}, \delta) = \left( \frac{\delta r_{t-1} + x_t - 1}{x_t} \right)^{\delta c_{t-1}} (\delta r_{t-1})(\delta c_{t-1})^{-\delta r_{t-1} + x_t}, \quad \text{on } x_t = 0, 1, \ldots. \quad (8)$$

**Posterior at time $t$:** Observing $x_t$, the resulting posterior is $\phi_t|x_{0:t} \sim Ga(r_t, c_t)$, which has the same form as that at time $t-1$ but with updated parameters $r_t = \delta r_{t-1} + x_t$ and $c_t = \delta c_{t-1} + 1$.

**A2. Model Marginal Likelihood (MML)**

A key ingredient of formal model assessment is the model marginal likelihood that, in this first-order Markov model, is computed as the product of one-step forecast p.d.f.s evaluated at the realized data. At time $t$, this product is

$$p(x_{1:t}|x_0, \delta) = \prod_{s=1:t} p(x_s|x_{1:s-1}, \delta).$$

The product is most usefully written in its one-step updated form

$$p(x_{1:t}|x_0, \delta) = p(x_t|x_{1:t-1}, \delta)p(x_{1:t-1}|x_0, \delta) \quad (9)$$

where the contribution at time $t$ derives from the one-step ahead predictive density of eqn. (8) evaluated at the datum $x_t$. These are trivially computed.

In parallel analyses using a discrete set of $\delta$ values, the log of the marginal likelihood is linearly accumulated as data are sequentially processed. At any time $t$ this can be mapped to a posterior $p(\delta|x_{0:t}) \propto p(\delta|x_0)p(x_{1:t}|x_0, \delta)$ and then normalized over the grid of values for inference on $\delta$ at any time of interest. This can be used to identify/choose a modal value for inference on the $\phi_t$ conditional on a chosen $\delta$, or for model averaging.

The sequentially computed contributions to the marginal likelihood—the realized p.d.f. ordinates $p(x_t|x_{1:t-1}, \delta)$—can be monitored sequentially over time to provide an on-line tracking of model performance, with potential uses in flagging anomalous data at one node or any subset of nodes, using standard Bayesian model monitoring concepts; see West (1986), West and Harrison (1986, 1989 and Chapter 11 of 1997), and Prado and West (2010, Section 4.3.8).

**A3. Backward Sampling (BS)**

Reaching the end time $T$, we look back over time and revise the summary posterior distributions for the full trajectory of the latent gamma process $\phi_{1:T}$ based on all the observed data. This uses backward sampling as follows.

- Sample the final rate from the time $T$ posterior $\phi_T|x_{1:T}, \delta \sim Ga(r_T, c_T)$.
- Recurse back over time $t = T-1, T-2, \ldots, 1$, at each stage sampling $\phi_t$ from $p(\phi_t|\phi_{t+1}, x_{1:t}, \delta)$ via $\phi_t = \delta \phi_{t+1} + \epsilon_t$ with a “backward innovation” $\epsilon_t$ drawn from $\epsilon_t \sim Ga((1-\delta)r_t, c_t)$, independently of $\phi_{t+1}$.
Repeating the backward sampling generates a Monte Carlo sample of the trajectory $\phi_{1:T}$ from the full posterior $p(\phi_{1:T} | x_{0:T}, \delta)$ for summary inferences.

**References**


