Bayesian Dynamic Modeling for Streaming Network Data

by

Xi Chen

Department of Statistical Science
Duke University

Date: __________________________

Approved:

______________________________
Mike West, Supervisor

______________________________
David Banks

______________________________
Alexander Volfovsky

______________________________
Andrew Cron

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistical Science in the Graduate School of Duke University 2017
ABSTRACT
Bayesian Dynamic Modeling for Streaming Network Data
by
Xi Chen

Department of Statistical Science
Duke University

Date: ______________________________
Approved:

________________________________
Mike West, Supervisor

________________________________
David Banks

________________________________
Alexander Volfovsky

________________________________
Andrew Cron

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Statistical Science
in the Graduate School of Duke University
2017
Abstract

Streaming network data of various forms arises in many applications, raising interest in research to model and quantify the nature of stochasticity and structure in dynamics underlying such data. One example context is that of traffic flow count data in networks, such as in automobile or aviation transportation, certain directed social network contexts, and Internet studies. Using an example of Internet browser traffic flows through site-segments of an international news website, I present Bayesian analyses of two new, linked classes of models which, in tandem, allow fast, scalable and interpretable Bayesian inference on dynamic patterns over time underlying flows. I develop two kinds of flexible state-space models for streaming count data, able to adaptively characterize and quantify network dynamics efficiently in real-time. These models are then used as emulators of more structured, time-varying gravity models that allow formal dissection of network dynamics. This yields interpretable inferences on traffic flow characteristics, and on dynamics in interactions among network nodes. Bayesian monitoring theory defines a strategy for sequential model assessment and adaptation in cases when network flow data deviates from model-based predictions. Exploratory and sequential monitoring analyses of evolving traffic on a network of web site-segments in e-commerce demonstrate the utility of this coupled Bayesian emulation approach to analysis of streaming network count data.

A second, different dynamic network context is that involving relational data. Examples include contexts of binary network data indicating communications or
relationships between pairs of network nodes over time. Some popular examples include friendships over social networks and communications between different functional zones in brain. Using an example of co-movements of company stock indices, I develop and compare two different approaches. One involves latent threshold models mapping latent processes to binary entries via a probabilistic link function, a second involves dynamic generalized linear models for binary outcomes. Analyses implemented using Markov chain Monte Carlo methods are available for these models, but naturally computationally demanding and not scalable to relevant network dimensions for many contexts. In contrast, dynamic generalized linear models can be implemented using fast, effective approximate Bayesian computations for both sequential and retrospective analyses to enable linear-time computations. I also demonstrate the use of a model decoupling/recoupling strategy to enable scaling in network size.
To my family.
Contents

Abstract iv
List of Tables xi
List of Figures xii
List of Abbreviations and Symbols xx
Acknowledgements xxii

1 Introduction 1
2 Web Site-Segment Network Flows 8
   2.1 Context and Data . . . . . . . . . . . . . . . . . . . . . . . . . . . 8
   2.2 Hourly Fox News Data . . . . . . . . . . . . . . . . . . . . . . . . . 9
   2.3 Daily Fox News Data with Text Mining Categorization . . . . . . . 11
   2.4 Network Structure and Notation . . . . . . . . . . . . . . . . . . . . 14
   2.5 Summary Comments . . . . . . . . . . . . . . . . . . . . . . . . . . . 15
3 Bayesian Dynamic Flow Models 16
   3.1 Background: Discrete-Time Volatility Models for Poisson Rates . . 16
      3.1.1 Forward filtering (FF) . . . . . . . . . . . . . . . . . . . . . 17
      3.1.2 Backward sampling (BS) . . . . . . . . . . . . . . . . . . . . 19
      3.1.3 Model marginal likelihood (MML) and discount factor . . . . 19
      3.1.4 Comments . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 20
   3.2 Network Inflows: Poisson Dynamic Models . . . . . . . . . . . . . 21
5.5 Model Mapping for DGMs by LLGMs .............................. 70
  5.5.1 Overall mean ........................................... 71
  5.5.2 Origin and destination effects .............................. 72
  5.5.3 Affinity effect ......................................... 73
  5.5.4 Singular value decomposition for DGM results .............. 99
  5.5.5 Findings in daily Fox News data ........................... 103
5.6 Summary Comments ............................................. 105

6 Bayesian Sequential Monitoring ................................. 107
  6.1 General Comments .......................................... 107
  6.2 Bayesian Model Monitoring and Automated Intervention ....... 107
  6.3 Monitoring for Sparse Flows ................................. 110
    6.3.1 Discounted process with time varying discount factor .... 113
    6.3.2 Influence on Bayes factor ............................... 115
  6.4 Hourly Fox News Network Example with BDFM .................. 116
  6.5 Comparison with Piecewise Median ............................ 117
  6.6 Summary Comments .......................................... 120

7 Dynamic Relational Data Analysis .............................. 122
  7.1 Introduction .............................................. 122
  7.2 Latent Threshold Model for Relational Data ................... 124
  7.3 Dynamic Generalized Linear Model for Relational Data ....... 125
  7.4 Example of S&P 500 Index Co-movement ......................... 126
    7.4.1 S&P 500 index co-movement data ........................ 126
    7.4.2 Analysis by LTM with MCMC ............................. 127
    7.4.3 Analysis by DGLM with FFBS ............................ 128
  7.5 Summary Comments .......................................... 131
List of Tables

7.1 Five chosen S&P 500 sectors for analysis. . . . . . . . . . . . . . . . . 127
List of Figures

2.1 A snapshot of counts and flows on the Fox News network at time $t = 1$ (09:05.30 on February 23rd 2015). The circular numbered nodes represent the site-segments, with diameters and colors proportional to occupancy $n_{it}$ for node $i$ at this time point $t$. Each arrow $i \rightarrow j$ has width proportional to flow $x_{ijt}$. ................................................. 12

2.2 Tree structure of 25 sub-categories under main category “News”. Each node represents a category with its name as a combination of all nodes on the path originated from the root. For example, following the path $1 \rightarrow 11 \rightarrow 17 \rightarrow 23$, the classification for Node 23 is “/News/Business News/Company News/Company Earnings”. ................................. 13

2.3 Network schematic and notation for flows at time $t$. ....................... 14

3.1 Marginal log-posteriors for baseline discount factors $d_i$ for the inflows to Fox News nodes $i = 1:22$ for the February 23rd 2015 am period. ............... 27

3.2 BDFM-based inference on inflows to node $j = 10$, the Leisure site-segment. $Upper$: data $x_{0,10,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. $Center$: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{0,10,t}|x_{0,10,0:t})$. $Lower$: revised trajectory under full retrospective posteriors $p(\phi_{0,10,t}|x_{0,10,0:T})$. ................................. 29

3.3 BDFM-based inference on transitions from $i = 1$ (Homepage) to $j = 2$ (Politics). $Upper$: data $x_{1,2,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. $Center$: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{1,2,t}|x_{1,2,0:t})$. $Lower$: revised trajectory under full retrospective posteriors $p(\phi_{1,2,t}|x_{1,2,0:T})$. .................. 29

3.4 BDFM-based inference on transitions from $i = 1$ (Homepage) to $j = 7$ (Science). $Upper$: data $x_{1,7,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. $Center$: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{1,7,t}|x_{1,7,0:t})$. $Lower$: revised trajectory under full retrospective posteriors $p(\phi_{1,7,t}|x_{1,7,0:T})$. ................................. 30
3.5 BDFM-based inference on transitions from $i = 1$ (Homepage) to $j = 11$ (World). Upper: data $x_{1,11,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{1,11,t}|x_{1,11,0:t})$. Lower: revised trajectory under full retrospective posteriors $p(\phi_{1,11,t}|x_{1,11,0:T})$.

3.6 Retrospective posterior means and 95% pointwise intervals of the trajectories of transition probability $\theta_{1,0,t}$ (Homepage $\rightarrow$ External) from analysis on each of the six mornings.

3.7 Retrospective posterior means and 95% pointwise intervals of the trajectories of transition probabilities $\theta_{1,5,t}$ (Homepage $\rightarrow$ Entertainment) for each of the six mornings.

4.1 Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on flows staying at $i =$ “Arts & Entertainment” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{3,3,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{3,3,t}$.

4.2 Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on flows staying at $i =$ “Arts & Entertainment” on September 17, 2015. Notations are the same as above but now for retrospective results.

4.3 Forward filtering analysis by LLGM (discount factor $\delta = 0.80$) on flows staying at $i =$ “Arts & Entertainment” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{3,3,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{3,3,t}$.

4.4 Retrospective analysis by LLGM (discount factor $\delta = 0.80$) on flows staying at $i =$ “Arts & Entertainment” on September 17, 2015. Notations are the same as above but now for retrospective results.

4.5 Forward filtering analysis by LLGM (discount factor $\delta = 0.95$) on flows staying at $i =$ “Arts & Entertainment” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{3,3,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{3,3,t}$.

4.6 Retrospective analysis by LLGM (discount factor $\delta = 0.95$) on flows staying at $i =$ “Arts & Entertainment” on September 17, 2015. Notations are the same as above but now for retrospective results.
4.7 Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on transitions from $i =$ “Games/Computer & Video Games” to $j =$ “Games/Online Games” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{41,105,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{41,105,t}$. .......................... 48

4.8 Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on transitions from $i =$ “Sports” to $j =$ “News” on September 17, 2015. Notations are the same as above but now for retrospective results. .......................... 48

4.9 Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on transitions entering from $i =$ “Autos & Vehicles” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{0,47,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{0,47,t}$. .......................... 49

4.10 Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on transitions entering from $i =$ “Autos & Vehicles” on September 17, 2015. Notations are the same as above but now for retrospective results. .......................... 49

4.11 Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on transitions exiting via $i =$ “News” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{16,0,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{16,0,t}$. .......................... 50

4.12 Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on transitions exiting via $i =$ “News” on September 17, 2015. Notations are the same as above but now for retrospective results. .......................... 50

4.13 Summaries from analysis of flow staying at “Arts & Entertainment” by LLGM and BDFM, both with discount factor 0.90. Data is from September 17, 2015. Data (+) with one-step forecast means and 95% intervals from the LLGM analysis in red/light red compared to the analysis standard BDFM in black/gray. .......................... 51

5.1 Bayesian credible value is defined to be posterior $p$–value as shaded area (left: larger than 1; right: less than 1). .......................... 60

5.2 Smoothed trajectories of selected node-specific outflows $\alpha_{i,1:T}$. The + symbols indicate empirical values computed from the raw data (with cases of 0 occupancy leading to missing values). .......................... 61

5.3 Smoothed trajectories of selected node-specific inflows $\beta_{j,1:T}$. .......................... 61
5.4 Posterior summaries for DGM parameters for transitions from node $i=$Homepage $\rightarrow j=$Science. As in Figures 5.2 and 5.3, the + symbols indicate empirical values (with cases of 0 occupancy leading to missing values). Upper left: Posterior trajectory for the latent flow level process $\phi_{ijt}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{ijt}$ with raw frequencies (crosses). Center left: Posterior trajectory for the Homepage origin (outflow) effect process $\alpha_{it}$. Center right: Posterior trajectory for the Science destination (inflow) effect process $\beta_{jt}$. Lower left: Posterior trajectory for the Homepage:Science affinity/interaction process $\gamma_{ijt}$. Lower right: Corresponding trajectories of Bayesian credible values.

5.5 Posterior summaries for DGM parameters for transitions from node $i=$Homepage $\rightarrow j=$Entertainment, with details as in Figure 5.4.

5.6 Posterior summaries for DGM parameters for transitions staying at node $i=$Entertainment, with details as in Figure 5.4.

5.7 Posterior summaries for DGM parameters for transitions from node $i=$Entertainment $\rightarrow j=$Leisure, with details as in Figure 5.4.

5.8 Posterior summaries for DGM parameters for transitions from node $i=$World $\rightarrow j=$US, with details as in Figure 5.4.

5.9 Heat-map images of standardized posterior means of DGM parameters for all nodes across time in the February 23rd am period. Means are max-min standardized separately for each of the $\alpha_{it}$, $\beta_{jt}$ and $\gamma_{1jt}$ (1=Homepage). Shading runs from 0 (dark blue) to 1 (yellow) on these standardized scales. The nodes are ordered based on correlation over time of the main outflow effects $\alpha_{it}$ with that of Homepage; this is an arbitrary ordering chosen simply for visual presentation.

5.10 DGM-based inference on trajectories of baseline flow levels $\mu_t$ across days with 95% intervals. Red are for morning periods, blue lines are for afternoons.

5.11 DGM-based inference on smoothed trajectories of baseline activity level of Fox News data on September 17th, 2015 with 95% credible interval. Blue + symbols indicate empirical values computed from the raw data.

5.12 Smoothed trajectories of selected node-specific origin effects $\alpha_{i,1:T}$ with 95% credible interval. The + symbols indicate empirical values computed from the raw data.
5.13 Smoothed trajectories of selected node-specific destination effects $\beta_{j,1:T}$ with 95% credible interval. The + symbols indicate empirical values computed from the raw data. ............................. 75

5.14 Posterior summaries for DGM parameters for transitions staying at $i =$ Arts & Entertainment. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{it}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{it}$ with raw frequencies (crosses). Center left: Posterior trajectory for the Arts & Entertainment origin (outflow) effect process $\alpha_{it}$. Center right: Posterior trajectory for the Arts & Entertainment destination (inflow) effect process $\beta_{it}$. Lower left: Posterior trajectory for the Arts & Entertainment self-affinity/interaction process $\gamma_{it}$. Lower right: Corresponding trajectories of Bayesian credible values. ............................. 77

5.15 Posterior summaries for DGM parameters for transitions staying at $i =$ Finance/Investing with details as in Figure 5.14. ............................. 78

5.16 Posterior summaries for DGM parameters for transitions staying at $i =$ Computers & Electronics/Software with details as in Figure 5.14. 79

5.17 Posterior summaries for DGM parameters for transitions staying at $i =$ Arts & Entertainment/TV & Video with details as in Figure 5.14. 80

5.18 Posterior summaries for DGM parameters for transitions entering $i =$ News/Weather. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{0it}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{0it}$ with raw frequencies (crosses). Center left: Posterior trajectory for the external origin (outflow) effect process $\alpha_{0it}$. Center right: Posterior trajectory for the News/Weather destination (inflow) effect process $\beta_{0it}$. Lower left: Posterior trajectory for the News/Weather entering affinity/interaction process $\gamma_{0it}$. Lower right: Corresponding trajectories of Bayesian credible values. ............................. 82

5.19 Posterior summaries for DGM parameters for transitions entering $i =$ News/Politics with details as in Figure 5.18. ............................. 83

5.20 Posterior summaries for DGM parameters for transitions entering $i =$ Health/Pharmacy/Drugs & Medications with details as in Figure 5.18. An interesting bump is noted in the afternoon (about 12:00 - 16:00). 84
5.21 Posterior summaries for DGM parameters for transitions entering $i =$ Finance/Investing with details as in Figure 5.18. Note a maintained high level during the day (8:00 - 16:00). ............................... 85

5.22 Posterior summaries for DGM parameters for transitions entering $i =$ Shopping with details as in Figure 5.18. ............................... 86

5.23 Posterior summaries for DGM parameters for transitions entering $i =$ Beauty & Fitness/Fashion & Style with details as in Figure 5.18. 87

5.24 Posterior summaries for DGM parameters for transitions exiting from $i =$ Arts & Entertainment. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{i0t}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{i0t}$ with raw frequencies (crosses). Center left: Posterior trajectory for the Arts & Entertainment origin (outflow) effect process $\alpha_{it}$. Center right: Posterior trajectory for the external destination (inflow) effect process $\beta_{0t}$. Lower left: Posterior trajectory for the Arts & Entertainment leaving affinity/interaction process $\gamma_{ijt}$. Lower right: Corresponding trajectories of Bayesian credible values. ................................................ 89

5.25 Posterior summaries for DGM parameters for transitions exiting from $i =$ Food & Drink/Cooking & Recipes with details as in Figure 5.24. 90

5.26 Posterior summaries for DGM parameters for transitions exiting from $i =$ News/Weather with details as in Figure 5.24. ............................... 91

5.27 Posterior summaries for DGM parameters for transitions exiting from $i =$ Arts & Entertainment/Music & Audio with details as in Figure 5.24. A huge peak is noted at night (around 20:00). ............................... 92

5.28 Posterior summaries for DGM parameters for transitions from node $i =$ Games/Online Games $\rightarrow j =$ Games/Computer & Video Games. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{ijt}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{ijt}$ with raw frequencies (crosses). Center left: Posterior trajectory for the Games/Online Games origin (outflow) effect process $\alpha_{it}$. Center right: Posterior trajectory for the News/Weather destination (inflow) effect process $\beta_{jt}$. Lower left: Posterior trajectory for the Games/Online Games : Games/Computer & Video Games affinity process $\gamma_{ijt}$. Lower right: Corresponding trajectories of Bayesian credible values. ................................................ 95
5.29 Posterior summaries for DGM parameters for transitions from node 
$i = \text{News} \rightarrow j = \text{News/Local News}$ with details as in Figure 5.28. ... 96

5.30 Posterior summaries for DGM parameters for transitions from node 
$i = \text{Home & Garden} \rightarrow j = \text{Reference/General Reference/How to DIY & Expert Content}$ with details as in Figure 5.28. ... 97

5.31 Posterior summaries for DGM parameters for transitions from node 
$i = \text{News/Technology News} \rightarrow j = \text{Shopping}$ with details as in Figure 5.28. ... 98

5.32 Dominant factors in SVD results for posterior mean trajectories of origin effects with data of FoxNews visiting flows on September 17, 2015 (left: cumulative variation explained by factors; right: time varying patterns of five dominant factors). ... 101

5.33 SVD results for posterior mean trajectories of origin effects: time varying first component and its loadings (data: FoxNews visiting flows on September 17, 2015). ... 102

5.34 Dominant factors in SVD results for posterior mean trajectories of destination effects with data of FoxNews visiting flows on September 17, 2015 (left: cumulative variation explained by factors; right: time varying patterns of five dominant factors). ... 103

5.35 SVD results for posterior mean trajectories of destination effects: time varying first component and its loadings (data: FoxNews visiting flows on September 17, 2015). ... 104

6.1 Schematic of monitoring, routine outlier/change assessment and automatic intervention to allow for more adaptability in times of change, as discussed and detailed in Chapter 6. This relies on a Bayes factor threshold $\tau$ (e.g. $\tau = 0.1$) and run-length threshold $r$ (e.g. $r = 4$). This schematic is a modified version of Figure 11.9 in West and Harrison (1997). ... 111
6.2 Summaries from analysis of transitions from node $i=$Homepage $\rightarrow j=$World with Bayesian model monitoring and discount-based intervention. Data is from the February 23rd am period. *Upper:* Symbol + indicates observations judged consistent with the standard model; x indicates cases identified as potential outliers by low $H_t$; * indicates those flagged as potential change points via low $L_t$; o indicates cases with $l_t > 1$. The vertical arrows indicate times of automatic intervention. The full line and shaded region represent one-step forecast means and 95% intervals. *Center:* Tracks of $\log(H_t)$ (above center) and $\log(L_t)$ (below center) over time. *Lower:* Data (+) with one-step forecast means and 95% intervals from the standard BDFM analysis in red compared to the analysis with monitoring and intervention in gray.

6.3 Summaries from analysis of transitions from node $i=$Homepage $\rightarrow j=$Science under monitoring and intervention, with details as in Figure 6.2.

6.4 A comparison of anomaly detection results by piecewise median monitoring (green asterisks) and sequential Bayesian monitoring (red circles) for the flow from node $i=$Homepage $\rightarrow j=$World on the morning of Feb. 23.

7.1 Frequencies of S&P 500 Index co-movement between five sectors in 01/02/2008 - 12/31/2008. Due to the symmetry property of relational data, I am only modeling the upper triangle of adjacency matrix, and thus set the diagonal and lower triangle entries as 0.

7.2 Posterior mean by MCMC for LTM of latent processes $y_{1:P,1:T}$ in 01/02/2008 - 12/31/2008 with fluctuations in February and July-August.

7.3 Posterior samples by MCMC for LTM of auto-regressive coefficients $\phi_{1:P}$ with dashed vertical red line as posterior mean.

7.4 Posterior mean by MCMC for LTM of correlations between sector pairs.

7.5 Sequential analysis by DGLM of of binomial probability $\mu_{1:T}$ with 95% credible interval in 01/02/2008 - 12/31/2008. The ‘*‘s are real data.

7.6 Retrospective analysis by DGLM of binomial probability $\mu_{1:T}$ with 95% credible interval in 01/02/2008 - 12/31/2008. Notations are the same as in Figure 7.5 but now for retrospective results.
List of Abbreviations and Symbols

Symbols

\( Po(\lambda) \) Poisson distribution with rate parameter \( \lambda \).

\( Ga(a, b) \) Gamma distribution with shape parameter \( a \) and rate parameter \( b \).

\( NB(k, p) \) Negative binomial distribution with number of failures \( k \) and success probability in each experiment \( p \).

\( Mn(n, \theta) \) Multinomial distribution with population \( n \) and probability vector \( \theta \).

\( Be(a, b) \) Beta distribution with shape parameters \( a \) and \( b \).

\( IW(s, S) \) Inverse-Wishart distribution with degree of freedom \( s \) and scale matrix \( S \).

\([m, V]\) Distribution with mean \( m \) and covariance matrix \( V \).

\( \mathcal{D}_t \) All information available at time \( t \).

\( s : t \) Indices \( s, s + 1, \ldots, t - 1, t \).

Abbreviations

AR(p) Autoregressive process with order \( p \).

BDFM Bayesian Dynamic Flow Model.

DGLM Dynamic Generalized Linear Model.

DGM Dynamic Gravity Model.

DLM Dynamic Linear Model.

FFBS Forward Filtering Backward Smoothing/Sampling.
LBE  Linear Bayesian Estimation.
LLGM  Local Linear Growth Model.
LTM  Latent Threshold Model.
MCMC  Markov Chain Monte Carlo.
SGDLM  Simultaneous Graphical Dynamic Linear Model.
SVD  Singular Value Decomposition.
Acknowledgements

First, I would like to express my deepest gratitude to my advisor, Professor Mike West, for his invaluable mentoring and support. During the four years of my PhD studies, he always provided me with insightful guidance on research, which inspired me intellectually and led to the accomplishment of my thesis. I would also like to thank him for introducing me to various academic events and internship opportunities, and his kind advice on communication skills. These are beneficial not only for my study and career development, but also for my whole life.

I am grateful to all the faculty members in the Department of Statistical Science. Their instruction during classes not only provided me with statistical knowledge, but has also encouraged me to analyze real-world problems from different perspectives. In particular, I would like to thank Professor David Banks, Professor Alexander Volfovsky and Professor Andrew Cron for serving as my preliminary exam and/or dissertation committee members. Special thanks to Professor David Banks for being my first-year advisor who kindly introduced me to the Duke statistics community and helped coordinate my research schedules and collaborative projects.

I would also like to thank MaxPoint for supporting my PhD studies. Discussions with my MaxPoint collaborators, Mark Lowe, Rob Haslinger and Jewell Thomas, provided useful insights into online commerce industry, and inspired my research. Further thanks to Dr. Rob Haslinger for his time and effort as a preliminary exam committee member.
I would like to acknowledge my colleagues in Professor Mike West’s group for sharing their understanding of research with me. Special thanks to Kaoru Irie who helped a lot when I was stuck in research. Finally, I would like to thank my parents, Renwen and Ruxiao, for their love and support.
Increasing access to streaming data on dynamic networks drives interest in models to quantify stochasticity and structure of latent processes underlying observable data streams. Key challenges are to develop real-time/sequential analysis that is computationally efficient and scalable with network size and sampling rates. This thesis concerns appropriate and efficient Bayesian dynamic modeling, and focuses on two types of problems. Chapters 2 to 6 is the first part, developing models, analysis and monitoring for dynamic network flow count data. The second part, Chapter 7, looks at dynamic relational data.

The first part of this thesis is about modeling and methodological development for dynamic network flow count data, and the research innovations are prompted by an application to Internet traffic in e-commerce. Online advertisers are interested in many statistical issues related to traffic flow and site-segment 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 site-segments have not received much
attention. In particular, there is commercial value in identifying the change of a site’s popularity on short time scales, and the interactions between sites with respect to traffic. In addition, sequential monitoring and appropriate interventions are also important to signal dynamics that may reflect interesting departures from the norm. Proposed models also elucidate the complexities and dynamics in network structure in both single sample inference and multi-sample comparisons across contexts. The main contributions of the work are as follows.

- A flexible and customized statistical modeling framework for: (i) characterizing patterns of temporal variation in network flows at the levels of nodes and pairs of nodes; (ii) model-based exploratory data analyses of network flows within and across contexts; and (iii) the ability to scale to large networks.

- Use of these flexible, efficient models as Bayesian emulators of more structured network flow models. This yields computationally efficient insight into the dynamics of node-specific and dependence/interaction effects between nodes in a structured model where analysis is otherwise computationally challenging in large networks.

- Formal Bayesian model assessment methodology for sequential monitoring of flow patterns with the ability to signal departures from predictions in real-time and allow informed interventions as a response, again in a scalable framework.

- Development and validation of the above in exploratory and monitoring analyses of data from the motivating application. The observations are streaming counts of visitors in a set of defined web site-segments (collections of webpages) in a well-defined but dynamic/evolving website. This includes evaluation of node-specific and node-pair interactions in the flow dynamics within the network over a given time period, comparisons across time periods and across days, and analyses that use Bayesian monitoring and adaptation to respond to
departures from predicted flow patterns.

Following a discussion of the motivating application in Chapter 2, details of Bayesian dynamic flow models (BDFMs) are in Chapter 3. BDFMs are flexible univariate dynamic models for series of counts representing flows into/out of the network and between within-network node pairs. These non-stationary and non-normal state-space models for streaming count data rely on discrete-time gamma processes historically used in volatility modeling (West and Harrison, 1989a, 1997, Section 10.8), and that have recently been used for flexible smoothing (Ameen and Harrison, 1985; Harrison and West, 1987; Quintana and West, 1987) with “steady” evolutions in non-Gaussian dynamic models (Smith, 1979). These gamma-beta discount models yield closed form Bayesian analyses. Using these models for within-network flows is novel and involves methodological extension to adapt and customize them to provide suitable univariate emulators of the underlying dynamic multinomial structures governing flows at each time point. This use of sets of decoupled univariate models that are then recoupled to define the actual multinomial probability processes (a) is explicitly designed to be computationally efficient in on-line data analysis, scaling quadratically in the number of network nodes and enabling distributed implementation for streaming data on large networks; (b) allows for diverse patterns in the dynamics of flow rates that a time-varying Dirichlet-multinomial model simply cannot; and (c) relates to the recent development of conceptually similar approaches that have advanced multivariate dynamic modeling in conditionally normal contexts (Gruber and West, 2016, 2017; Zhao et al., 2016). Section 3.4 summarizes some aspects of an exploratory analysis of the network flow data on the Fox News website. This highlights the use of customized BDFMs, with a focus on exploring flow dynamics on the network across the same time periods on different days.

Chapter 4 introduces a more flexible class of models for the latent flow rate pro-
cess: the local linear growth models (LLGMs). Chapter 3 indicates a strong need for model improvement to track local linear increases or decreases in some network flows. Chapter 4 is inspired by previous research in dynamic generalized linear models (DGLMs) (West and Harrison, 1997; Migon, 1984; Migon and Harrison, 1985; West et al., 1985, Chapter 14). Related research can be traced back to McCullough and Nelder (1989) on generalized linear models (GLMs), and Bayesian approaches as in West (1985). As a special kind of DGLMs, LLGMs introduce a local linear term able to capture local trends, which can help the model to adapt more quickly and appropriately to trend changes in data. The analysis developed adopts sequential Bayesian computation that combines variational Bayes approximation with linear Bayes updates, extending earlier DGLM methods (West and Harrison, 1997; Hartigan, 1969; Goldstein, 1976). A detailed example of Fox News flows modeled by LLGMs and comparison with BDFMs is given in Section 4.3.

Chapter 5 introduces a class of more highly structured dynamic gravity models (DGMs) for network flows. These are non-normal, log-linear random-effects models with time-varying parameters for flow rate contributions of origin nodes, destination nodes and origin-destination interaction effects. The proposed DGMs extend prior work with static gravity models (e.g. West, 1994; Tebaldi and West, 1998; Sen and Smith, 1995; Congdon, 2000) to the time-varying parameter context, defining a class of models able to represent complicated patterns of dependency structure, and their temporal variation, across nodes. In a 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. Similar work of using Bayesian approach to analyze and predict traffic flows also includes Anacleto et al. (2013a,b). Congdon (2000) explored a similar approach in studies of patient flows to hospitals, and Jandarov et al. (2014) studied the spatio-temporal dynamics of infectious diseases by looking
at Bayesian inferences from a gravity model. However, MCMC is hugely computationally demanding, and importantly, in this thesis, it is shown that the flexible and computationally simple BDFM/LLGM framework can be mapped one-to-one to that of DGMs. This underlies another novel contribution of this work: the use of the fast, efficient BDFMs/LLGMs as emulators of DGMs. This is key from the viewpoint of scalability; fitting gravity models, even without time-varying parameters, is a challenging issue in all but small networks, and infeasible in any dynamic extension appropriate for scalable, on-line analyses of streaming network flow data. Further, the proposed models avoid the challenge of defining and parameterizing time-evolution models for DGMs directly, adopting the implicit structures induced in the mapping from BDFMs/LLGMs, where model specification and fitting is relatively easy. Example results and highlights from the Bayesian emulation analysis of DGMs for the MaxPoint Fox News study appear in Sections 5.4 and 5.5.

Chapter 6 develops methods of formal, sequential Bayesian model monitoring and adaptation (automatic intervention) for BDFMs. The aim here is to build into the fast, decoupled analysis an ability to efficiently evaluate incoming flows against model predictions so as to signal data—at the level of individual nodes and node-pairs—that appear discrepant, and that may indicate outliers or changes in flow trends/rates. In addition to signaling such events and thus providing opportunity for direct intervention, monitoring is coupled with the use of automatic intervention to allow the model to appropriately adapt to data at the next few time points. This Bayesian testing/adaptation strategy builds on core theory underlying its use in time series forecasting contexts with dynamic linear models (West, 1986; West and Harrison, 1986; Harrison and West, 1987; West and Harrison, 1989b; see also Chapter 11 of West and Harrison, 1997). The novelty here lies in the use of these ideas in non-linear, non-normal dynamic models for count data—BDFMs. Importantly, monitoring is applied in parallel across nodes and node-pairs, so is also scalable with
network size. Some departures from normal variation in patterns of flow may be related across nodes, and the approach has an ability to explore and evaluate this within the BDFM context and then map the relation to more structured DGMs that reflect interaction effects. Application to the Fox News network data illustrates this method for selected network nodes.

In the second part of this thesis, Chapter 7 discusses modeling for dynamic relational data, which is a kind of binary network data indicating communication or relationships between pairs of nodes in the network over time. Some common examples include friendship in social network, and communication between different functional zones in the brain. Dynamic relational data has become a popular research topic due to its wide applications, especially in social network analysis. Various statistical models have been developed (Fu et al., 2009; Berger-Wolf and Saia, 2006; Durante and Dunson, 2014). In this thesis, modeling is considered from two perspectives: different modeling strategies for streaming binary data, and the utilization of network structure. Developments involve binary probit models with dynamic latent thresholding ideas inspired by previous work (Nakajima and West, 2013b,a; Zhou et al., 2014; Nakajima and West, 2015, 2017), and dynamic logistic regressions (West and Harrison, 1997, Chapter 14). For the network structure, models based on some form of latent variable structure can be used to induce dependence between nodes, while node-specific covariates can be incorporated. There is also an emerging literature in simultaneous graphical dynamic linear model with the decoupling/recoupling idea which encourages using contemporary network information to improve performance (Gruber and West, 2016, 2017; Zhao et al., 2016). An application to stock index co-movement data is developed in Sections 7.4.2 and 7.4.3. As in prior work in related studies (Tsay, 2005), developing models with a focus on the binary summaries of market up/down movements can provide useful insights into the market relationship and dynamics, while somewhat side-stepping the challenges of
model building and misspecification for the original multivariate time series of stock
prices or returns.

The thesis concludes in Chapter 8 with some summary comments and points to
open questions and potential future research.
2

Web Site-Segment Network Flows

2.1 Context and Data

The data are traffic flows among site-segments (defined sets of pages) of the Fox News website. Site-segments include the “Homepage” and several categories of news and consumer content—such as “Politics”, “Entertainment”, “Travel”, “Science”, etc.—as defined by Fox News. While the site-segment structure is persistent, the nature of webpage definition and content within a site-segment is dynamic; content changes on a daily basis (updated at least at midnight, and often more frequently). MaxPoint\(^1\) places ads on pages in these Fox News site-segments, and thus can track flows of anonymized users as they move through its pages. While some users can be tracked individually, this is not the norm, and I focus in this study on aggregated flow counts, not the trajectories of individuals.

On-line advertisers are interested in many statistical issues related to traffic flow

\(^1\) MaxPoint Interactive, Inc. (https://maxpoint.com/us/) is a marketing technology company providing brands and retailers insights into customers’ shopping behaviors both physically and online. Its business includes placing ads on website pages to targeted users, and Fox News is one of the websites it provides service to. The two Fox News data sets discussed in this thesis are both provided by MaxPoint.
and site-segment 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 site-segments have not received much attention. In particular, there is commercial value in identifying how the popularity of a site changes on short time scales, and how sites interact with respect to traffic. For example, the data showed a morning-after spike in traffic to the “Entertainment” site-segment following the Grammy awards, which was an opportunity to advertise concert tickets. Similarly, unusual interactions in flows between “Science” and “Health” may reflect new medical findings that incline people to purchase gym memberships, and increased flows from “Homepage” to “Technology” may signal marketing opportunities related to a new iPhone release.

As pages within a site-segment are updated, patterns in browsing traffic patterns may change. To discover changes, it is necessary to understand stochastic variation in recent traffic so that comparisons can be made between current traffic and short-term predictions to identify significant discrepancies. Companies that have flow models which enable them to predict how traffic will change as content changes, that are able to sensitively characterize and monitor patterns of change in both overall rates and interactions, and that can signal anomalous changes that drive business decisions, will be advantaged. They can recognize opportunities more quickly, and, for example, adapt bidding strategies for relevant keywords to dynamically reflect expected revenue.

2.2 Hourly Fox News Data

The first data set contains Fox News website visit data during 09:00–10:00 a.m. and 01:00–02:00 p.m. EST on each of six days, February 23-24, March 2-3 and 9-10, 2015.
These days are Mondays and Tuesdays.

Since the Fox News website structure changes often, with new pages being added and old pages being archived, the analysis aggregates webpages into groups specified by the host site www.foxnews.com, and the set of first url paths after the host site, such as www.foxnews.com/politics/* and www.foxnews.com/US/*. These classify all pages into 22 site-segments: 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. These categories are the most refined network structure in the Fox News website that is stable. Other applications that have similarly stable network structure include air transportation and (approximately) Wikipedia flows.

The data set includes anonymized visitors 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 I 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.

Aggregate data give time series of counts in half-minute intervals, i.e., 120 time points of site-segment occupancy, flows from each site-segment, and flows into each site-segment. In each half minute interval, if the record shows the same user in two or more site-segments, then each of her/his moves is counted in the flow data into each of these site-segments. 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 site-segment; this can be done as the web browsing tool performs automatic refresh. Importantly, if a user stays in the same site-segment for more than five minutes, s/he is declared to be inactive and is counted as leaving the Fox News site. If such a user later appears in one site-segment, s/he counts as inflow from outside the Fox
News site. Finally, from this data set alone, it is impossible to track user information either before or after the one-hour observation window; thus attention is restricted to the period 09:05–09:55 a.m. and 01:05–01:55 p.m., 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. Figure 2.1 provides a visualization of data at the first time interval.

Aggregation at half-minute intervals balances interest in fine-time scale modeling against the information content of node pairs with low flow rates. If flow rates are too low, it causes excessive volatility in the BDFM and noisy parameter estimates in the DGM. No single window is good for all node pairs at all times, but preliminary exploration found half-minutes to be a good compromise.

The decision that a user is inactive after five minutes is based on previous research on how users access on-line articles. Few people read more than the first paragraph of a news story. Investigations of on-line session length have focused on dynamics of search engine use (e.g. Silverstein et al., 1999; Qiu et al., 2005, and references therein) in settings where the full breadth of user browsing behavior is visible. These studies indicate that for more than 90% of the websites that people visit, the time spent is less than five minutes (Jansen et al., 2007).

2.3 Daily Fox News Data with Text Mining Categorization

Another data set is the visiting flow data on the Fox News website for the entire day on September 17th, 2015. The website is segmented by Adex Category, a categorization method based on text mining results of webpage contents. Nowadays, the Adex Category is widely used by online advertising companies for webpage content analysis. This categorization method has 2,208 pre-defined categories, including 26 main categories and different levels of sub-categories under them. The 26 main categories are Arts & Entertainment, Computers & Electronics, Finance, Games, Home
Figure 2.1: A snapshot of counts and flows on the Fox News network at time $t = 1$ (09:05.30 on February 23rd 2015). The circular numbered nodes represent the site-segments, with diameters and colors proportional to occupancy $n_{it}$ for node $i$ at this time point $t$. Each arrow $i \rightarrow j$ has width proportional to flow $x_{ijt}$.

& Garden, Business & Industrial, Internet & Telecom, People & Society, News, Shopping, Law & Government, Sports, Books & Literature, Real Estate, Beauty & Fitness, Health, Autos & Vehicles, Hobbies & Leisure, Pets & Animals, Travel, Food & Drink, Science, Online Communities, Reference, Jobs & Education and World Localities, which covers almost all the topics online advertising strategists care about. There is a tree structure with each main category as the root. For instance, there are 25 sub-categories under Category “News”, as in Figure 2.2. Though there is overlap between the meanings of different categories, they are treated as independent by the text mining tool and form a partition of the website. Exploratory study also shows no clustering pattern among categories sharing the same main category, and thus we assume independence for all categories in the analysis.

Data are aggregated with five-minute intervals. As suggested by stability test
results, it is appropriate to use a time interval of five minutes for data aggregation, which leads to a time series with 288 time points. That is to say, at each time point, over the directed network with nodes classified by Adex Category, the data include counts of flows between each pair, incoming flows from the external to the Fox News website for each node, as well as the total number of people visiting each node.

As discussed in Section 2.2, previous studies show most people spend no more than five minutes on a single webpage (Jansen et al., 2007); therefore, visitors spending more than five minutes on the same webpage are set as inactive and declared as having left the Fox News website in data aggregation. Moreover, the users' information is unavailable both before and after the chosen day, so by the five-minute leaving rule, the first and last five minutes should be eliminated from analysis and
the total length of this time series to be analyzed is $T = 284$.

Finally, since Fox News is not a comprehensive website, there are some categories with very low or even no traffic during the entire chosen day, so only those categories with sufficient data are considered in the analysis. By applying a threshold of 1 for the total traffic across all $T = 284$ time periods, $I = 237$ out of 2,208 categories are left for analysis.

2.4 Network Structure and Notation

Referring to all sites external to the Fox News website as node 0, we have 23 network nodes for the first data set and 238 for the second. “External” and the $I = 22$ or $I = 237$ actual site-segments are indexed as $i = 0:I$. At each time $t = 1:T$, define $x_{ijt}$ as the flow count from node $i$ to $j$, including the inflows $x_{0it}$ and outflows $x_{i0t}$ relative to External. Also, denote the number of occupants of node $i$ at the end of the $t$ period by $n_{it}$—a random quantity at the start of the period, but then known and given by the sum of inflows minus outflows at the end of the period. The schematic in Figure 2.3 reflects the notation.

![Figure 2.3: Network schematic and notation for flows at time $t$.](image)

For later development, notation for various distributions arising in the network flow models is defined and summarized here. If $\phi > 0$ has a gamma distribution with
shape $r$ and scale $1/c$, we write $\phi \sim Ga(r, c)$, noting that $E(\phi) = r/c$ and $V(\phi) = r/c^2$.

If $\eta \in (0, 1)$ has a beta distribution with p.d.f. proportional to $\eta^{a-1}(1 - \eta)^{b-1}$ for $a > 0, b > 0$, we write $\eta \sim Be(a, b)$. If the $n-$vector of counts $\mathbf{z}$ has a multinomial distribution with total counts $n$ and probability vector $\theta$, we write $\mathbf{z} \sim Mn(n, \theta)$.

For any series of random quantities $x_t$ over $t = 0, 1, \ldots$, we use the succinct notation $x_{h:k} = \{x_h, x_{h+1}, \ldots, x_k\}$ for any indices $h \leq k$.

2.5 Summary Comments

Traffic flow count data in networks arise in many applications. With a specific interest in online advertisement placement, I study statistical issues related to traffic flows and site-segment contents. Two data sets of Internet browsing traffic flow through site-segments of an international news website, Fox News, are introduced. One is an hourly data set with 22 site-segments and 110 time points using 5 minutes as aggregation interval for 12 chosen hours in late February and early March in 2015, either in the morning or in the afternoon on 6 chosen days. The other is for the entire day on September 17th, 2015 with 237 site-segments indexed with Adex categories and 284 time points using half hour as the aggregation interval.

With the two data sets, there are five goals that I want to achieve by studying site-segment networks: i) to efficiently analyze a single flow’s change over time and how it is driven by the assumed latent flow-rate process; ii) to give precise short-term predictions of entries in the network; iii) to implement scalable algorithms for model fitting in order to study the node-node interactions; iv) to compare network flow behavior across different days or in the context of different news events; and, v) to identify latent network structure and quickly flag anomalies.
3 Bayesian Dynamic Flow Models

3.1 Background: Discrete-Time Volatility Models for Poisson Rates

BDFMs are based on gamma-beta random walks that have been key to stochastic volatility modeling historically (West and Harrison, 1989a, 1997, Section 10.8). Based initially on Bayesian discount concepts related to exponential smoothing for volatilities (Ameen and Harrison, 1985; Harrison and West, 1987; Quintana and West, 1987) and with “steady” evolutions in non-Gaussian dynamic models (Smith, 1979), these gamma-beta discount models yield closed-form Bayesian analyses. They have seen some—though limited—use as models for rates underlying time series of conditionally Poisson counts (Harvey and Fernandes, 1989; Brandt et al., 2000), which is a starting point here. The basic model is extended in a number of ways, with novel model forms customized to the network flow context and that define flexible models for conditional multinomial data with time-varying probabilities that go beyond prior use. Full Bayesian simulation of posteriors for latent rate processes is also heavily utilized, which extends the use from the normal/linear dynamic volatility modeling context (Prado and West, 2010, Chapter 4).
Using generic notation, a series of non-negative counts $x_t$ over $t = 1:T$ is modeled via $x_t|\phi_t \sim Poi(m_t \phi_t)$ conditionally independently over time, where the underlying/latent Poisson rate process $\phi_t$ follows a gamma-beta stochastic model and each $m_t$ is a scaling constant known at time $t$. 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 discount factor parameter $\delta_t \in (0, 1)$, potentially different at each time $t$. The following sections summarize model concepts and structure, and the implied machinery for Bayesian learning and forecasting that includes forward filtering, backward sampling (FFBS) for conditionally Poisson time series coupled with the gamma-beta steady process model.

3.1.1 Forward filtering (FF)

At time $t = 0$, introduce an hypothetical latent state $\phi_0$ and use $x_0$ as notation for all available initial information. Specify an initial gamma prior, so $\phi_0 \sim Ga(r_0, c_0)$ 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 $0:t - 1$ is gamma,

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

(3.1)

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

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

(3.2)
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 $\delta_t$ 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.

Note that the beta innovations distribution for $\eta_t$ at time $t$ depends in the accumulated information content about the time $t - 1$ level through the shape parameter $r_{t-1}$. The discount factor $\delta_t$ acts to decrease the information content between times $t - 1$ and $t$ in a natural way. That is, information loss rates are constant over time, rather than parameters of the innovation distribution. The specific choice of beta distribution ensures that the implied time $t$ prior has a conjugate gamma form.

**Prior for time $t$:** The time $t - 1$ gamma posterior of eqn. (3.1) couples with the beta innovation to give the time $t - 1$ prior for the next state as

$$\phi_t|x_{0:t-1} \sim Ga(\delta_tr_{t-1}, \delta_tc_{t-1}).$$

(3.3)

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 generalized negative binomial with p.d.f.

$$p(x_t|x_{0:t-1}, \delta_{1:t}) = \frac{\Gamma(\delta_tr_{t-1} + x_t)}{\Gamma(\delta_tr_{t-1})\Gamma(x_t + 1)} \frac{m_t^x(\delta_tc_{t-1})^{\delta_tr_{t-1}}}{(\delta_tc_{t-1} + m_t)^{\delta_tr_{t-1} + x_t}} \quad (3.4)$$
on $x_t = 0, 1, \ldots$.
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 $r_t = \delta_t r_{t-1} + x_t$ and $c_t = \delta_t c_{t-1} + m_t$.

3.1.2 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 based on theory in West and Harrison (1989a, Section 10.8); see also Prado and West (2010, Section 4.3.7 and problem 4 of Section 4.6). Specifically:

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

Repeating the backward sampling generates a Monte Carlo sample of the trajectory $\phi_{1:T}$ from the full posterior $p(\phi_{1:T}|x_{0:T}, \delta_{1:T})$ for summary inferences.

3.1.3 Model marginal likelihood (MML) and discount factor

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

$$p(x_{1:t}|x_0, \delta_{1:t}) = \prod_{s=1}^{t} p(x_s|x_{0:s-1}, \delta_s).$$

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

$$p(x_{1:t}|x_0, \delta_{1:t}) = p(x_t|x_{0:t-1}, \delta_{1:t}) p(x_{1:t-1}|x_0, \delta_{1:t-1})$$  \hspace{1cm} (3.5)
where the contribution at time \( t \) derives from the one-step ahead predictive density of eqn. (3.4) evaluated at the datum \( x_t \). These are trivially computed.

One of the most useful roles of the marginal likelihood is in comparing models based on different (sets of) discount factor values. As one key special case, suppose \( \delta_t = \delta \) is fixed over the time period of interest. Then eqn. (3.5) gives the value of the marginal likelihood \( p(x_{1:t}|x_0, \delta) \) at any chosen value of \( \delta \). 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 \( p(\delta|x_{0:t})xp(\delta|x_0)p(x_{1:t}|x_0, \delta) \) 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 \( p(x_t|x_{0:t-1}, \delta) \)—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).

### 3.1.4 Comments

The model can be defined by any sequence of specified discount factors \( \delta_t \). A constant value over time defines a global smoothing rate; values closer to 1 constrain the stochastic innovation and hence the change from \( \phi_{t-1} \) to \( \phi_t \); smaller discount factor values lead to greater random changes in these Poisson levels. Intervention to specify smaller discount factors at some time points, to reflect or anticipate higher levels of dynamic variation at those times, are sometimes relevant. In the network flow models below, the specification of the discount factor sequence is customized to address issues that arise in cases of low flow levels. That extension of discount-based
modeling defines the $\delta_t$ as time-varying functions of an underlying base discount rate, and the latter are then evaluated using MML measures.

This model provides the basis for flows into network nodes, and it can be adapted and generalized to define components of flexible multinomial dynamic models for flows between nodes in a network.

### 3.2 Network Inflows: Poisson Dynamic Models

With notation for inflows as in Figure 2.3, the general model of Section 3.1 is adopted by adding suffixes $i$ for network nodes and setting the Poisson mean scaling factors to 1. This model is now customized via specification of discount factor sequences. At any node $i$, the time $t$ inflow to node $i$ is $x_{0it} \sim \text{Poi}(\phi_{it})$ independently across nodes $i = 1:I$, and the latent levels $\phi_{it}$ follow node-specific gamma-beta discount models with discount factor $\delta_{it}$ at time $t$. The time $t \rightarrow t + 1$ update/evolve steps are: (i) the time $t$ prior $\phi_{it}|x_{0i,0:t-1} \sim \text{Ga}(\delta_{it}r_{i,t-1}, \delta_{it}c_{i,t-1})$ updates to the posterior $\phi_{it}|x_{0i,0:t} \sim \text{Ga}(r_{it}, c_{it})$ with $r_{it} = \delta_{it}r_{i,t-1} + x_{0it}$ and $c_{it} = \delta_{it}c_{i,t-1} + 1$. This then evolves to the time $t + 1$ prior $\phi_{i,t+1}|x_{0i,0:t} \sim \text{Ga}(\delta_{i,t+1}r_{i,t}, \delta_{i,t+1}c_{it})$, and so on. Specifying discount factors $\delta_{it}$ relates to the information content of gamma distributions as measured by the shape parameters $r_{it}$; evolution at each time point reduces this by the discount factor, the latter representing a per-time-step decay of information induced by the stochastic evolution. The specification of discount rates is motivated by the following considerations. First, baseline levels of variation on $\phi_{it}$ are likely to be node specific, so that each node should have its own baseline discount rate that is assessed in data analysis. Second, in cases of zero flow rates for a period of time, $r_{it}$ is continually discounted and shrinks towards 0 while $c_{it}$ is incremented by 1 at each update step. That is, discounting is not balanced by the prior-posterior update and thus generates more and more diffuse posteriors favoring lower and lower $\phi_{it}$. Ideally, the posterior and prior should be very similar in cases of zero flows, which can be
addressed with the specification $\delta_{it} = d_i + (1 - d_i) \exp(-kr_{i,t-1})$ at each $i, t$, where $d_i$ is a constant baseline discount factor for node $i$ and $k > 0$ is a specified constant. The aim is that $\delta_{it}$ be close to the baseline unless information content is very low; thus the applied studies take $k = 1$ (so that $\delta_{it}$ be close to—within 10% of—the baseline when $r_{i,t-1} > 2$). In cases of high information content, the effective $\delta_{it}$ is close to $d_i$; otherwise, $\delta_{it}$ is close to 1, appropriately limiting the decay of information in such cases.

Node-specific MML measures feed into model assessment to aid in selection of the baseline discount factors $d_i$. These measures of short-term predictive fit of the models can also be monitored sequentially over time for on-line tracking of model performance. This ability to flag anomalous data at one node or any subset of nodes is key to commercial application of the analysis, since that corresponds to new opportunities or new threats (e.g., offer concert tickets after the Grammy Awards, but not on David Bowie’s obituary). This view on anomaly detection is extended below, in Chapter 6, using Bayesian model monitoring. One aspect of this is the need to temporarily reduce the value of the discount factor for a node at a time when predictive performance degrades, presumably due to changes in $\phi_{it}$ that are larger than the “standard” baseline discount factor value $d_i$ determines.

3.3 Transitions from Network Nodes: Multinomial Dynamic Models

Transitions from any node $i$ at time $t$ are inherently multinomial with time-varying transition probabilities. To build flexible and scalable models for dynamics and dependencies in transition probability vectors is a challenge, with computational issues for even simple models quickly dominating. Novel models here adapt and extend the univariate Poisson/gamma-beta random walk models to enable flexibility in modeling node-pair specific effects as they vary over time as well as scalability.
3.3.1 Model structure

Considering flows from node $i$ to node $j$ at time $t$, and using notation as in Figure 2.3, the core model is $x_{i,0;I,t} \sim Mn(n_{i,t-1}, \theta_{i,0;I,t})$ where the current node $i$ occupancy level is $n_{i,t-1}$, and $\theta_{i,0;I,t}$ is the $(I + 1)$-vector of transition probabilities $\theta_{ijt}$ (including the “external” node—i.e., leaving the Fox News network—at $j = 0$). The decoupled BDFMs are structured in terms of positive flow rates $\phi_{ijt}$ underlying each $x_{ijt}$. Specifically,

$$x_{ijt} \sim Poi(m_{it}\phi_{ijt}), \quad m_{it} = n_{i,t-1}/n_{i,t-2}, \quad (3.6)$$

independently, with independent gamma-beta evolutions for each latent level $\phi_{ijt}$. These BDFMs for each node pair can be customized with node-pair specific discount factors, allowing greater or lesser degrees of variation by node pair. The set of models for elements of $\phi_{i,0;I,t}$ implies a dynamic model for the vector of transition probabilities $\theta_{i,0;I,t}$ having elements $\theta_{ijt} = \phi_{ijt}/\sum_{j=0;I} \phi_{ijt}$. Independence across nodes enables scaling, as the analyses can then be decoupled and run in parallel for the $\phi_{ijt}$ and then recoupled to infer the $\theta_{ijt}$. Dependencies in patterns of changes in the $\phi_{ijt}$ are recovered in evaluating the posterior distributions and, as in Chapter 5, in using this set of models to emulate gravity models that explicitly characterize pairwise dependencies.

3.3.2 Discussions on occupancy ratio

A key and critical component of the model is the definition of the scaling factors $m_{it}$ in eqn. (3.6). In decoupling the multinomial flows from node $i$ into parallel Poisson models for nodes $i \to j$, the inherent dependency on total occupancy of node $i$ is lost. That dependence is restored by using this specific definition of scaling factors to explicitly correct for occupancy changes. This recognizes that the decoupled, scaled models are not predictive of overall occupancy—rather, they are decoupled,
tractable models that are relevant to tracking and short-term prediction of relative occupancy levels through the implied multinomial probabilities. The relevance of this scaling factor is most evident in cases of major changes in occupancy, when an abrupt increase in node \( i \) occupancy \( n_{i,t-1} \) at time \( t-1 \) relative to its prior value \( n_{i,t-2} \) will lead to increased flows to other nodes at time \( t \) even if the underlying transition probabilities \( \theta_{i,0:t} \) are essentially constant. In such a case, the scaling factor will encourage the appropriate view that the \( \phi_{ijt} \) are stable. In other words, variation in flows can be explained by change in either \( n_{i,t-1} \) or \( \theta_{ijt} \). Use of the modification term eliminates the influence of occupancy count, providing a smoother Poisson rate and better control when flow counts are small.

Before and after introducing the \( m \) (occupancy ratio) term, the models of the latent processes are

\[
\phi_{ijt} = \phi_{ij,t-1}\eta_{ijt}/\delta_{ijt}, \tag{3.7}
\]

\[
\phi^*_{ijt} = m_{it}\phi^*_{ij,t-1}\eta_{ijt}/\delta_{ijt}, \tag{3.8}
\]

with priors as \( \phi_{ij0} \sim \text{Ga}(r_{ij0}, c_{ij0}) \) and \( \phi^*_{ij0} \sim \text{Ga}(r^*_{ij0}, c^*_{ij0}) \), and with discount factors \( \{\delta_{ijt}\}_{t=1:T} \) and \( \{\delta^*_{ijt}\}_{t=1:T} \). By induction, it can be proved that

\[
r_{ijt} = \prod_{k=1:t} \delta_{ijk}r_{ij0} + \sum_{l=2:t} \prod_{k=1:t} \delta_{ijk} x_{ijk} + x_{ijt},
\]

\[
c_{ijt} = \prod_{k=1:t} \delta_{ijk}c_{ij0} + \sum_{l=2:t} \prod_{k=1:t} \delta_{ijk} + 1,
\]

\[
r^*_{ijt} = \prod_{k=1:t} \delta^*_{ijk}r^*_{ij0} + \sum_{l=2:t} \prod_{k=1:t} \delta^*_{ijk} x_{ijk} + x_{ijt},
\]

\[
c^*_{ijt} = \prod_{k=1:t} \delta^*_{ijk}c^*_{ij0} + \sum_{l=2:t} \prod_{k=1:t} \delta^*_{ijk} + 1.
\]

Under the condition that all the outgoing flows from node \( i \) share the same set of discount factors (\( \delta^*_{ijt} = \delta_{ijt} \) for all \( j = 0:I \) and \( t = 1:T \)), and the two processes share the same prior information (\( r^*_{ij0} = r_{ij0} \) and \( c^*_{ij0} = c_{ij0} \) for all \( j = 0:I \)), which is met by
the analysis in Section 3.4 and latter studies, it can be proved that for all $j = 0 : I$,

$$\phi_{ijt}^* = m_{it}\phi_{ijt}. \quad (3.9)$$

Then, inferences on the $\phi_{ijt}$ directly yield inferences on the transition probabilities of interest: the conditional multinomial probabilities are simply not impacted by the scaling factors, i.e.,

$$\theta_{ijt} = m_{it}\phi_{ijt}/\sum_{j=0:I} m_{it}\phi_{ijt} = \phi_{ijt}/\sum_{j=0:I} \phi_{ijt}. \quad (3.10)$$

In addition, for sparse flows, a slight change in the occupancy count might have a large influence on both inference and predictions if using the scaling factor defined in eqn. (3.6), which causes instability in the model. Therefore, in practice, the following strategy is applied:

- Use $m_{it} = 0$ without any thresholding when $n_{i,t-1}$ is zero;
- Use $m_{it} = 1$ if either occupancy count is less than or equal to 10 and $n_{i,t-1}$ is not zero;
- Use $m_{it} = n_{i,t-1}/n_{i,t-2}$ if both occupancy terms exceed 10.

### 3.3.3 More model details

The theory and analysis details of Section 3.1 now apply with data and latent flow levels indexed by origin node $i$ and destination node $j$. As with inflow models, it is flexible to choose discount factors specific to the context. Following the discussion of Section 3.2, the discount factor is specified as $\delta_{ijt} = d_{ij} + (1 - d_{ij}) \exp(-kr_{ij,t-1})$ at each $i, j, t$, where $d_{ij}$ is a constant baseline discount factor for node pair $i, j$, and $k > 0$ a specified constant. Again, this is later overlaid with intervention to adjust discount factor values as needed, based on sequential monitoring of flow patterns and
using the MML measures—now of course for each node pair—as one formal guide to
model adequacy.

In sequential analysis of transitions, the node-pair specific models generate full
joint predictions one-step ahead (or more, if desired) for the theoretically exact set of
multivariate flow vectors \( x_{i,0:t,t} \) across all nodes. The one-step forecast distribution
does not have an analytic closed form, but is trivially simulated to define forecasts.
That is: (i) simulate directly from each of the gamma-beta evolutions for the \( \phi_{ij,t} \);
(ii) transform sampled values to the conditional multinomial probabilities \( \phi_{ij,t} \); then
(iii) sample the multinomial \( x_{i,0:t,t} \sim Mn(n_{i,t-1}, \theta_{i,0:t,t}) \) at these parameter values.
Similarly, for both on-line and retrospective inference about transition probabilities,
samples from posteriors for the \( \phi_{ij,t} \) again simply transform to the required probability
scale.

3.4 Aspects of BDFM Analysis of Fox News Data

The analysis was applied separately to data from each of the six days. I focus
on the morning of February 23rd 2015 for initial summaries, and then make some
comparisons across days. Priors for the inflow rates \( \phi_{i0} | x_{00} \sim Ga(r_{i0}, c_{i0}) \) with
\( c_{i0} = 1 \) and \( r_{i0} = c_{i0} z_i \) where \( z_i \) is from inflows in the five minutes prior to the
start of model analysis at \( t = 1 \). The priors for the node-node flows are, similarly,
\( \phi_{ij0} | x_{0ij0} \sim Ga(r_{ij0}, c_{ij0}) \) with \( c_{ij0} = 1 \) and with \( r_{ij0} = c_{ij0} z_{ij} \) where each \( z_{ij} \) is a
point estimate based on the initial five minutes. With \( c_s = 1 \) the priors are relatively
diffuse, and for most nodes one or a few initial observations “wash out” the effect
of the prior. While some node-node pairs have very low counts, they all see traffic
that quickly updates the shape parameters—some network links have counts in the
thousands, while the average is around 40–50 in the fifty-minute window.

The prior for each baseline discount factor is a \( Be(19, 1) \) truncated to \((0.9, 0.999)\);
this encourages smoothness. Reanalysis using uniform priors on this range found
Figure 3.1: Marginal log-posteriors for baseline discount factors $d_i$ for the inflows to Fox News nodes $i = 1:22$ for the February 23rd 2015 am period.

little difference, as the marginal likelihoods at $t = T$ dominate. Running models in parallel across discrete grids for each discount factor and evaluating MML measures at $t = T$ gives marginal likelihoods that are mapped to posteriors. Figure 3.1 plots posteriors for the $d_i$ in the inflow models. Some nodes exhibit higher volatility in flows than others, consistent with smaller discount factors; these are particularly associated with site-segments with high flow counts (e.g., inflows to Homepage). Constraining the range to higher values reflects the expectation of generally “smooth” trajectories for the $\phi_{st}$ processes, which is the case with the majority of flows; allowing smaller values has little impact on much of the reported analysis. However, for a few nodes whose flows have high levels of change, lower discount factors would lead to posteriors that suggest more volatile trajectories. But these flows are better addressed in a model that uses a higher discount factor as standard, with interventions to allow
increased uncertainty and adaptation in response to discrepant observations (whether singletons or batches). The work in Chapter 6 is heavily motivated by this. Of course, users may use whatever priors they regard as relevant in this model.

Summary inferences on selected model components are reported from models with discount factors set at posterior modes. Figure 3.2 gives one example of learning about inflow to the “Leisure” site-segment. This 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 3.3 is an example of flow between two network nodes: from “Homepage” to “Politics”. As with Figure 3.2, it should be noted that the concordance of incoming data with the one-step predictive intervals as they are successively revised in the forward analysis, and the enhanced smoothing of trajectories in the retrospective analysis.

With careful choice of temporal discount factors, BDFM is prevented from overacting to outliers, and thus becomes more stable. However, smoothing can sometimes make the model slow to adapt to change points and rate trends. For most of the cases, when the rate change is not very significant, BDFM is still able to make accurate predictions. As shown by the inference on transitions from “Homepage” to “Science” in Figure 3.4, BDFM captures the steady increase in trend from 09:05 to 09:35. However, the model sometimes make less accurate one-step ahead predictions when there is a sharp change in the trend or level of the rate process. For example, on the transitions from “Homepage” to “World” in Figure 3.5, BDFM acts slowly to the change point at 09:23, and does not capture the trend change until 09:30. It does not perform well during 09:45-09:50 either.

On transition probabilities, it is natural to look at examples involving the “Homepage”, the most popular single site-segment on Fox News. For example, Figure 3.6 shows that the probability of people leaving the Fox News website from “Homepage”
Figure 3.2: BDFM-based inference on inflows to node \( j = 10 \), the Leisure site-segment. Upper: data \( x_{0,10,t} \) (plus signs) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors \( p(\phi_{0,10,t}|x_{0,10,0:t}) \). Lower: revised trajectory under full retrospective posteriors \( p(\phi_{0,10,t}|x_{0,10,0:T}) \).

Figure 3.3: BDFM-based inference on transitions from \( i = 1 \) (Homepage) to \( j = 2 \) (Politics). Upper: data \( x_{1,2,t} \) (plus signs) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors \( p(\phi_{1,2,t}|x_{1,2,0:t}) \). Lower: revised trajectory under full retrospective posteriors \( p(\phi_{1,2,t}|x_{1,2,0:T}) \).
Figure 3.4: BDFM-based inference on transitions from $i = 1$ (Homepage) to $j = 7$ (Science). Upper: data $x_{1,7,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{1,7,t}|x_{1,7,0:t})$. Lower: revised trajectory under full retrospective posteriors $p(\phi_{1,7,t}|x_{1,7,0:T})$.

Figure 3.5: BDFM-based inference on transitions from $i = 1$ (Homepage) to $j = 11$ (World). Upper: data $x_{1,11,t}$ (plus signs) with one-step ahead forecast means and 95% intervals. Center: trajectory of mean and 95% intervals from on-line posteriors $p(\phi_{1,11,t}|x_{1,11,0:t})$. Lower: revised trajectory under full retrospective posteriors $p(\phi_{1,11,t}|x_{1,11,0:T})$. 
Figure 3.6: Retrospective posterior means and 95\% pointwise intervals of the trajectories of transition probability $\theta_{1,0,t}$ (Homepage $\rightarrow$ External) from analysis on each of the six mornings.

Figure 3.7: Retrospective posterior means and 95\% pointwise intervals of the trajectories of transition probabilities $\theta_{1,5,t}$ (Homepage $\rightarrow$ Entertainment) for each of the six mornings.
increases in this 50 minute window for each of the six mornings. Note that there are significant day effects; e.g., visitors were more likely to leave Fox News on the morning of March 9th compared to the other mornings. More detailed insights based on the gravity model analysis are noted in Section 5.4. Similar figures (not shown) highlight patterns and day-to-day differences for other transitions. For example, most “Homepage” visitors stay on “Homepage” for a while and have a high probability of exiting the Fox News site entirely from that page. Across all six days, the probability of staying on the “Homepage” each time interval generally decreases over the course of the 50-minute morning period.

As an illustration of a more detailed analysis of a very specific flow, consider Figure 3.7. Among the visitors who leave the “Homepage” for other Fox News site-segments, “Entertainment” is generally the most popular destination. Across all mornings, there are major differences in trajectories of transition probabilities; in particular, February 23rd and 24th have higher rates than the other days. It is noteworthy that the Academy Awards ceremony was held on the night of February 22nd, which may have driven this uptick. Additional summaries of inference on trajectories of selected $\phi_{ijt}$ and $\theta_{ijt}$ processes appear in the mapping to dynamic gravity models; see the upper row of frames in Figures 5.4 and 5.5, for examples.

3.5 Summary Comments

This chapter introduces BDFMs for fast, efficient sequential tracking of dynamic network flows. The model first decouples the network into individual flows as univariate streaming count data. For each individual flow, the model is basically a state-space model with Poisson/Gamma conjugate structure using multiplicative gamma random walks as latent processes. The evolution of the latent gamma processes is defined using time varying discount factors, which enhances the smoothing of retrospective trajectories of the rate processes. As a modification term, the occupancy factor en-
ables scale-free Poisson/Gamma analysis, which can not only provide more stable posterior inferences on the latent gamma processes, but also adjusts for low-rate flows—an important issue in applications with sparse network structures.

Thanks to the highly independent structure after decoupling, the analysis is decoupled and parallelizable across node pairs. This enables use of FFBS for all the flows, which allows fast, scalable and interpretable Bayesian inference. Then, full network-wide inferences are enabled by recoupling across all nodes.

With the hourly Fox News streaming flow data, I show some examples of the dynamics for individual flows as well as a comparison of transition probabilities between six chosen hours. BDFM is able to capture data dynamics and make accurate predictions most of the time; however, it fails to adapt promptly to big level or trend change in data. Therefore, two approaches are suggested to address these issues: i) local trend dynamic models to better capture the trend change; ii) sequential monitoring and intervention to modify model performance. These are developed in Chapters 4 and 6.
4.1 Basic Ideas of Flexible Modeling Structure

As in Chapter 3, the network has $I$ nodes and data arises at $T$ time points. An origin node 0 is introduced to denote nodes external to the network: that is, there are $I + 1$ network nodes with $I$ actual nodes and one “External”, indexed as $i = 0:I$. At each time point $t = 1:T$, $n_{it}$ is the number of occupants of node $i$, and $x_{ijt}$ is the flow count from node $i$ to $j$, including the in-flows $x_{0it}$ and out-flows $x_{i0t}$ relative to the “External” node; refer back to Figure 2.3.

The basic modeling idea is the latent space model assuming a Markovian structure for latent parameters. Conditionally independent Poisson ($Poi$) models are assumed for in-flows to the network and conditionally independent multinomials ($Mn$) for flows from each node, all with time-varying parameters: for all time $t = 1:T$, and all network nodes $i = 1:I$,

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

and the conditional independence is across nodes and over time. As for the conditioning, there are two kinds of (latent) process parameters: $\phi_{0jt}$ is the time-varying rate
The process governing flows into node $j$ from the external, and $\boldsymbol{\theta}_i = (\theta_{i0t}, \theta_{i1t}, \cdots, \theta_{iIt})'$ is the vector of time-varying transition probabilities from node $i$ to all the other nodes and $j = 0$ representing leaving the network.

Flexibility and efficiency is a key problem in dynamic network analysis especially for large scale networks. One possible application of the model is to optimize online advertisement placement strategies by analyzing website browsing traffic, which requires fast and large-scale analyses. To address this, a decoupling and recoupling idea is introduced to enable parallel computation. In the decoupling step, the network flows are considered as a collection of conditionally independent Poissons

$$x_{ijt} | \phi_{ijt} \sim \text{Poi}(m_{it} \phi_{ijt}),$$

where the occupancy ratio $m_{it}$ is defined to correct for changes as

$$m_{it} = n_{i,t-1}/n_{i,t-2}.$$  \hfill (4.2)

For each individual flow rate process, a conjugate gamma process is assumed as

$$\phi_{ijt} | \mathcal{D}_t \sim \text{Ga}(r_{ijt}, c_{ijt}).$$  \hfill (4.3)

With this model structure, analysis for individual flows can be updated independently and separately, which provides a suitable structure for parallel implementation. Then in the recoupling step, after having the inferences from all the flows, information about the whole network structure can be obtained. For instance, key concerns are those of inferences on transition probabilities

$$\theta_{ijt} = m_{it} \phi_{ijt} / \left( \sum_{k=0}^{I} m_{it} \phi_{ikt} = \phi_{ijt} / \sum_{k=0}^{I} \phi_{ikt}. \right.$$  \hfill (4.4)

Another way to analyze the network structure is to map the results to dynamic gravity models by decomposing the Poisson rate $\phi_{ijt}$ into four components at each
time point: i) baseline process $\mu_t$ representing the overall activity level of the network, ii) origin effect $\alpha_{it}$ as the intensity of the outflow process for node $i$, iii) destination effect $\beta_{jt}$ as the attractiveness of the inflow process for node $j$, and iv) affinity effect $\gamma_{ijt}$ which shows the time-varying directional affinity for node $i$ to node $j$, as in eqn. (4.5).

$$
\phi_{ijt} = \mu_t \alpha_{it} \beta_{jt} \gamma_{ijt}.
$$

(4.5)

I develop this Bayesian model emulation idea in Chapter 5 below.

4.2 Local Linear Growth Model

In LLGMs, the evolution of each latent rate $\phi_{ijt}$ is characterized by a local linear growth model. Previous work on dynamic network flow analysis (Chen et al., 2017) indicates a strong need for model improvement of keeping track of local linear increase/decrease, for which dynamic generalized linear models (DGLMs) can provide entree (West and Harrison, 1997, Chapter 14). This work is inspired by previous research in dynamic exponential family models (Migon, 1984; Migon and Harrison, 1985; West et al., 1985). As a special kind of DGLM, LLGM introduces a local linear term into the model which is able to capture local changes in underlying Poisson rates via models of “local linear” change.

4.2.1 Model structure

For a generalized linear model, the natural parameter of a Poisson rate is its logarithm

$$
\lambda_t = \log(\phi_t),
$$

(4.6)

which can be modeled with a linear model with latent Markovian structure

$$
\begin{align*}
\lambda_t &= F_t' \theta_t, \\
\theta_t &= G_t \theta_{t-1} + w_t, w_t \sim [0, W_t],
\end{align*}
$$

(4.7)
where the latent state vector $\theta_t$ includes two parameters to be modeled dynamically, the level evolution, $\lambda_t$, and the local linear growth term, $\beta_t$; i.e. the state vector

$$\theta_t = \begin{pmatrix} \lambda_t \\ \beta_t \end{pmatrix}. \quad (4.8)$$

Here

$$F_t = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \quad (4.9)$$

This model meets the need for tracking local linear trends, and provides a flexible structure for incorporating more covariates, such as text mining information from website content, in future studies.

The state evolution noise $w_t$ is only specified by its mean $0$ and variance matrix $W_t$ here defined with a discount factor $\delta$ as

$$W_t = \frac{1 - \delta}{\delta} P_t, \quad (4.10)$$

where $P_t = G_t C_{t-1} G_t$ (West and Harrison, 1997, Chapter 6).

4.2.2 Sequential, retrospective updating and forecasting

Forward filtering analysis follows (West et al., 1985; West and Harrison, 1997, Chapter 14) in using a variational Bayes strategy to condition one-step ahead priors to conjugate gamma forms. This is coupled with the use of linear Bayes updates to define filtering (Hartigan, 1969; Goldstein, 1976; West and Harrison, 1997, Chapter 4). Linear Bayesian estimation (LBE) provides a Bayesian estimation decision-theoretic approach to estimating non-linear functions by linear regression (West and Harrison, 1997; Hartigan, 1969; Goldstein, 1976). With model specification in eqn. (4.7), a gamma-distributed variational Bayes approximation is applied as a bridge to connect the prior information and data to carry out the sequential updating.
Sequential updating

Standing at time $t$, with prior information $\mathcal{D}_{t-1}$ defining summaries of the posterior distribution $\theta_{t-1}|\mathcal{D}_{t-1}$, we now address forecasting and prior-posterior updating on the current data $x_t$. The time $t-1$ state posterior is characterized via its mean and variance matrix as

$$\theta_{t-1}|\mathcal{D}_{t-1} \sim [m_{t-1}, C_{t-1}],$$

(4.11)

where $m_{t-1}$ and $C_{t-1}$ are the posterior mean and variance of the state vector $\theta_{t-1}$.

By the state evolution equation, the prior at time $t$ is

$$\theta_t|\mathcal{D}_{t-1} \sim [a_t, R_t],$$

(4.12)

where $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t^t + W_t$. By expanding the matrix form and extracting the information on $\lambda_t$, the prior of $\lambda_t$ at time $t$ is now

$$\lambda_t|\mathcal{D}_{t-1} \sim [f_t, q_t],$$

(4.13)

where the mean and variance are

$$f_t = F_t^t a_t \text{ and } q_t = F_t^t R_t F_t.$$ (4.14)

I now apply the variational Bayes strategy to map to an assumed conjugate gamma prior. By the property of gamma random variables, for $\phi_t|\mathcal{D}_{t-1} \sim Ga(r_t, c_t)$, the mean and variance of its logarithm $\lambda_t$ are

$$\begin{cases} f_t := \mathbb{E}[\log \phi_t|\mathcal{D}_{t-1}] = \gamma(r_t) - \log c_t, \\ q_t := \mathbb{V}[\log \phi_t|\mathcal{D}_{t-1}] = \gamma'(r_t), \end{cases}$$

(4.15)

where $\gamma(\cdot)$ is the digamma function and $\gamma'(\cdot)$ is the trigamma function. The Newton-Raphson method can be then applied to solve eqn. (4.15) for $r_t$ by iterating

$$r_{t,n+1} = r_{t,n} - \frac{\hat{\gamma}(r_{t,n}) - q_t}{\hat{\gamma}'(r_{t,n})}$$

(4.16)
until convergence, and \( c_t \) can be then updated as

\[
c_t = \exp\{\gamma(r_t) - f_t\}. \tag{4.17}
\]

Given the conjugate Poisson/gamma structure, the posterior of \( \phi_t \) at time \( t \) can be easily updated as

\[
\phi_t | D_t \sim Ga(r_t + x_t, c_t + m_t). \tag{4.18}
\]

The posterior mean and variance of \( \lambda_t \) are then exactly computed as

\[
\begin{aligned}
    f_t^* &:= \mathbb{E}[\lambda_t | D_t] = \gamma(r_t + x_t) - \log(c_t + m_t), \\
    q_t^* &:= \mathbb{V}[\lambda_t | D_t] = \dot{\gamma}(r_t + x_t).
\end{aligned} \tag{4.19}
\]

Linear Bayes updating (Hartigan, 1969; Goldstein, 1976; West and Harrison, 1997, Chapter 14) then yields the feedback of information to define the posterior mean and variance of \( \theta_t \) as

\[
\begin{aligned}
    m_t &= a_t + R_t F_t (f_t^* - f_t) / q_t, \\
    C_t &= R_t - R_t F_t F_t' R_t (1 - q_t^* / q_t) / q_t.
\end{aligned} \tag{4.20}
\]

**k-step ahead forecasting**

Over future times \( k = 1, 2, 3, \ldots \), the evolution of the latent process can be characterized as

\[
\theta_{t+k} | D_t \sim [a_t(k), R_t(k)], \tag{4.21}
\]

where

\[
\begin{aligned}
    a_t(k) &= G(k)m_t, \\
    R_t(k) &= G(k)C_t G'(k) + \sum_{h=1:k} G(k-h)W_{t+h} G'(k-h). \tag{4.22}
\end{aligned}
\]

with

\[
G(k) := \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}. \tag{4.23}
\]
Marginally, the prediction for the logarithm of the latent rate at time \(t\) is

\[
\lambda_{t+k}|D_t \sim [F'_{t+k}a_t(k), F'_{t+k}R_t(k)F_{t+k}].
\] (4.24)

With \(r_t(k)\) and \(c_t(k)\) solved by Newton-Raphson as in eqn. (4.16) and eqn. (4.17), it follows that the gamma approximation to \(k\)-step ahead predictions is

\[
\phi_{t+k}|D_t \sim Ga(r_t(k), c_t(k)).
\] (4.25)

When \(k = 1\), since the occupancy ratio at time \(t + 1\) can be calculated using known information as

\[
m_{i,t+1} = n_{it}/n_{i,t-1},
\] (4.26)
the one-step ahead predictive distribution can be obtained by integrating out \(\phi_t\) to give

\[
p(y_{t+1}|D_t) \sim NB(r_{t+1}, \frac{c_{t+1}}{c_{t+1} + m_{i,t+1}}),
\] (4.27)

with number of failures \(r_{t+1}\) and success probability \(p_{t+1} = c_{t+1}/(c_{t+1} + m_{i,t+1})\).

However, when \(k\) is larger than 1, doing prediction is not as simple as for \(k = 1\). That is because, as in eqn. (4.2), the occupancy rescaling ratio \(m_{it}(k)\) is unknown for \(k > 1\), and needs to be predicted to define forecasts. In order to get predicted samples of \(m_{it}(k)|D_t\), it is necessary to do sampling for \(m_{it}(l)\) sequentially over \(l = 2 : k\) by following the steps:

- With the information from \((l-2)\)-step, sample the \((l-1)\)-step ahead predicted flows \(\{x_{ijt}(l-1)\}_{i=0:I, j=0:I}\).

- Calculate the occupancy data \((l-1)\)-step ahead as

\[
n_{it}(l-1) = n_{it}(l-2) - \sum_{j=0:I} x_{ijt}(l-1) + \sum_{j=0:I} x_{jlt}(l-1).
\] (4.28)
• Calculate the occupancy ratio as $m_{it}(l) = n_{it}(l-1)/n_{it}(l-2)$ for the next step’s flow prediction.

With samples of the occupancy ratio $m_{it}(k)$, the predictive distribution can be obtained by integrating out $\phi_{t+k}$ and $m_{it}(k)$ numerically.

Retrospective smoothing

The retrospective sampling is done in the same way as for normal DLMs. At time $t$, the smoothed posterior can be updated as

$$\lambda_t | \mathcal{D}_T, \phi_{t+1} \sim [a_t^*, R_t^*], \quad (4.29)$$

where

$$\begin{align*}
  a_t^* &= (1 - \delta) m_t + \delta G_t^{-1} a_{t+1}^*, \\
  R_t^* &= (1 - \delta) C_t + \delta^2 G_t^{-1} R_{t+1}^* G_t^{-1}.
\end{align*} \quad (4.30)$$

The variational Bayes gamma approximation is then applied in the same way as in the forward sequential updating by Newton-Raphson method; the difference here is that the variational approximation is applied to retrospectively updated posteriors.

4.2.3 Model advantages

By introducing the local linear growth term into the latent process, the LLGM is able to capture the local trend, which helps the model to adapt more quickly to trend changes in data, and thus may outperform the original model with multiplicative random walks. The occupancy factor enables scale-free Poisson/gamma analysis, which can not only provide smooth posterior inferences, but also adjusts for low-rate flows—an important issue in applications with sparse network structures.
4.3 Example with Daily Fox News Data

4.3.1 Forward filtering and backward smoothing

The analysis was applied to the Fox News visiting flow data with Adex Category segments on September 17th, 2015. I focus on the analysis on individual flows, and also evaluate its performance by comparing the one-step ahead predictions with predictions by BDFMs. In the LLGM analysis, for each individual flow, the latent state vector $\theta_{ij0}$ has two components: the rate parameter on the log scale, $\lambda_{ij0}$, and the local linear growth term, $\beta_{ij0}$. The prior mean for $\lambda_{ij0}$ is chosen as a point estimate based on the data in the five minutes prior to the beginning of the model, and the prior mean for $\beta_{ij0}$ is chosen to be 0. The prior covariance matrix of $\theta_{ij0}$ is chosen to be a diagonal matrix, which assumes independence between $\lambda_{ij0}$ and $\beta_{ij0}$ initially and would allow the model to discover the covariance structure from the data.

Thanks to the highly independent model structure, LLGMs can be applied simultaneously to all the individual flows. The LLGM analysis of flow staying at category “Arts & Entertainment” is chosen to illustrate the model performance and efficacy. Figure 4.1 shows the forward filtering results for both the rate parameter and the local linear growth term, while Figure 4.2 shows a more smoothed analysis by backward sampling. The discount factor is chosen as a constant $\delta = 0.9$, which encourages smoothness but also provides flexibility for the model to adapt to changes. In general, both the sequential and retrospective analyses capture the data well, especially the retrospection. There are periods when the data changes too rapidly for the model to adapt to data. For example, in the morning (8:00 - 10:00) and at late night (20:00 - 24:00), the model tends to underestimate the rate parameter $\phi_t$. The sequential analysis also tends to be easily influenced by outliers in the data. For instance, at around 16:00, there are two data points with very low counts compared to the data.
both before and after them, and are thus very likely to be outliers. These two data points drive the inferred rate down in forward filtering. Fortunately, the retrospective analysis is able to resolve these issues by using information from the latter data, and provides an improved and more smoothed analysis for the rate parameter.

In addition, both forward filtering and backward smoothing for the local linear growth term enable us to have a better knowledge of how trends vary with time during the day. The Poisson rate for people staying at category “Arts & Entertainment” firstly decreases at 0:00 - 4:00, but its decreasing speed also goes down, and then it reaches the minimum at around 4:00 am. Afterwards, the rate goes up rapidly and reaches the first peak during the day at around 8:00 am, and then keeps a high and stable activity level during the day until around 19:00; thereafter, \( \phi_{ijt} \) then goes up slightly reaching its second peak at around 20:00, and after that, the activity level keeps going down more and more rapidly until midnight.

When applying LLGMs, the level of smoothing should be carefully decided, since different choices of discount factor values lead to different model performances. With discount factor \( \delta = 0.80, 0.90, 0.95 \), the corresponding posterior inferences of transitions staying at “Arts & Entertainment” are summarized here. In addition to the discussion above for a discount factor to be 0.90, when using a smaller discount, the model becomes more adaptive to changes in data, but also less stable; while a higher discount factor leads to higher model stability, it sometimes over-smooths.

Figures 4.3 and 4.4 show the sequential and retrospective trajectories of both latent rate process and the local linear term when \( \delta = 0.80 \). The model now gives more precise one-step ahead predictions; however, the local linear term drops significantly lower than 0 even in retrospective analysis when several data points are obvious outliers at 16:00. That is to say, the model with \( \delta = 0.80 \) overacts to outliers in data. Figures 4.5 and 4.6 summarize the corresponding results of \( \delta = 0.95 \). As we can see from the one-step ahead predictions during 18:00-24:00, with too high a discount
factor, the model over-smooths the latent rate and local linear processes, and thus results in inaccurate forecasting.

Examples are also shown for transitions between two nodes, entering or leaving the Fox News website. Figures 4.7 and 4.8 summarize the FFBS results of transitions from “Games/Computer & Video Games” to “Games/Online Games”, Figures 4.9 and 4.10 are for flow entering from “Autos & Vehicles”, and Figures 4.11 and 4.12 are for transitions exiting Fox News via “News”. In all the three examples, the LLGM has satisfactory performance if assessed by the accuracy of one-step ahead predictions. It should also be noted that in the LLGM results for flow from “Games/Computer & Video Games” to “Games/Online Games”, the LLGM gives wide credible intervals when there is less information and data spreads out during 16:00-23:00, and tighter intervals when data concentrates.

4.3.2 Comparison with BDFM

In order to compare the performance of LLGMs and BDFMs, both classes of models are applied to the flow staying at category “Arts & Entertainment” with the same discount factor $\delta = 0.9$. Both models perform well when the data is relatively stable, as during 8:00-19:00; the one-step ahead forecasts by both models match the data well. However, the BDFM tends to underestimate when the data is going up and overestimate when the data is going down, as during 4:00-8:00 and 20:00-24:00 respectively, while LLGM still provides precise point-wise predictions. Generally speaking, LLGMs outperform BDFMs when the performance is measured by one-step ahead forecast accuracy, as in Figure 4.13.

4.4 Summary Comments

By introducing the local linear growth term into the latent process underlying network link flows, LLGMs are able to capture local trends, and thus help the model
Figure 4.1: Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on flows staying at $i = \text{“Arts & Entertainment”}$ on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{3,3,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{3,3,t}$.

Figure 4.2: Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on flows staying at $i = \text{“Arts & Entertainment”}$ on September 17, 2015. Notations are the same as above but now for retrospective results.
Figure 4.3: Forward filtering analysis by LLGM (discount factor \( \delta = 0.80 \)) on flows staying at \( i = "Arts & Entertainment" \) on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate \( \phi_{3,3,t} \); lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term \( \beta_{3,3,t} \).

Figure 4.4: Retrospective analysis by LLGM (discount factor \( \delta = 0.80 \)) on flows staying at \( i = "Arts & Entertainment" \) on September 17, 2015. Notations are the same as above but now for retrospective results.
Figure 4.5: Forward filtering analysis by LLGM (discount factor $\delta = 0.95$) on flows staying at $i = \text{"Arts & Entertainment"}$ on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{3,3,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{3,3,t}$.

Figure 4.6: Retrospective analysis by LLGM (discount factor $\delta = 0.95$) on flows staying at $i = \text{"Arts & Entertainment"}$ on September 17, 2015. Notations are the same as above but now for retrospective results.
Figure 4.7: Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on transitions from $i =$ “Games/Computer & Video Games” to $j =$ “Games/Online Games” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{41,105,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{41,105,t}$.

Figure 4.8: Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on transitions from $i =$ “Sports” to $j =$ “News” on September 17, 2015. Notations are the same as above but now for retrospective results.
Figure 4.9: Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on transitions entering from $i = "Autos & Vehicles"$ on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{0,47,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{0,47,t}$.

Figure 4.10: Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on transitions entering from $i = "Autos & Vehicles"$ on September 17, 2015. Notations are the same as above but now for retrospective results.
Figure 4.11: Forward filtering analysis by LLGM (discount factor $\delta = 0.90$) on transitions exiting via $i =$ “News” on September 17, 2015. Upper: data (crossings), forward filtering mean (solid line) and 95% CI (shaded area) of trajectories of transition rate $\phi_{16,0,t}$; lower: forward filtering mean (solid line) and 95% CI (shaded area) of local linear growth term $\beta_{16,0,t}$.

Figure 4.12: Retrospective analysis by LLGM (discount factor $\delta = 0.90$) on transitions exiting via $i =$ “News” on September 17, 2015. Notations are the same as above but now for retrospective results.
to adapt more quickly to trend change in data, and thus outperforms BDFMs. The model also uses the decoupling/recoupling strategy as in BDFMs. That is to say, it first decouples the network into individual flows to do LLGM analysis separately, and then recouples inferences to generate an overview of the network structure by looking at transition probabilities. The retrospective results in LLGMs help us understand the evolution of the individual flow’s latent structure without influence from noise. The decoupling/recoupling strategy also provides an ideal structure to perform parallel computation. Together with the efficient FFBS for each individual flow, the analysis is scalable and efficient for increasingly large dynamic networks. For each flow modeled with LLGM, the algorithm only needs $O(1)$ time for updating with each new observation, and thus the computation complexity for analyzing the whole time series is $O(T)$. Performed on a computer with $K$ cores over a network with $I$ nodes, the computational time of the proposed method scales as $O(T^2T/K)$.

The evolution of the latent process is defined using discount factors which determine the smoothing of LLGMs. With the daily Fox News data, I show that different
choices of discount factor values lead to different model performances. An optimal
discount factor can be chosen by maximizing the marginal likelihood, which is a
routine for choosing this temporal smoothing factor value in other DLM applica-
tions. Alternative ways include cross validation, etc. In order to improve LLGM
performance, an additional step of choosing discount factors is necessary in future
studies.

Moreover, as a special kind of dynamic generalized linear model, LLGMs also
provide flexible structure for incorporating time-varying covariate information, which
can potentially improve model performance. Unfortunately, there is no covariate
information in the application with the Fox News data, but further studies on data
sets with covariates would also be a potential next step. For instance, dynamic
traffic flow with geographical information and brain network data with cell connection
information would be interesting applications.
5

Emulation-Based Mapping to Dynamic Gravity Models

5.1 Dynamic Gravity Models

A more intricate, multivariate dynamic model involves node-specific main effects and node-node interaction terms, representing dependencies in patterns of flows linked to inflow/outflow and node-node relationships. For flows between all node pairs $i = 0: I$ and $j = 0: I$ excluding the unobservable flow $0 \rightarrow 0$, the model is

$$\phi_{ijt} = \mu_t \alpha_{it} \beta_{jt} \gamma_{ijt} \tag{5.1}$$

with: (i) a baseline process $\mu_t$; (ii) a node $i$ main effect process $\alpha_{it}$, adjusting the baseline intensity of flows—the origin or outflow parameter process for node $i$; (iii) a node $j$ main effect process $\beta_{jt}$, representing the additional “attractiveness” of node $j$—the destination or inflow parameter process for node $j$; and (iv) an interaction term $\gamma_{ijt}$, representing the directional “affinity” of node $i$ for $j$ over time relative to the combined contributions of baseline and main effects.

Models of this and more elaborate forms have seen some use in transportation studies (e.g. West, 1994; Tebaldi and West, 1998; Tebaldi et al., 2002; Sen and Smith, 53
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 $j$. In this thesis, the $\gamma_{ij}$ interactions are referred to as “affinities”, a
term that nicely describes the role of the terms linked to node-node intensities. In
dissecting the network flow activity, I am 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 a 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 net-
work. Later Bayesian studies that analyze and predict traffic flows include Anacleto
et al. (2013a,b). Congdon (2000) explored a similar approach in studies of patient
flows to hospitals. More recently, Jandarov et al. (2014) studied such models for
spread of infectious diseases, and used Gaussian process approximations for approx-
imate inference rather than full MCMC or other computational methods. Analysis
via MCMC is computationally very demanding. The burden increases quadratically
in $I$, and analysis is inherently non-sequential. To be more specific, for a network
with $I$ nodes and $T$ time steps, there are $I^2$ flows and thus the computational load
of analysis of the BDFM is $O(T)$ for each decoupled time series, or $O(TI^2)$ for the
entire network in the forward filtering step. And the time complexity is thus $O(TI^2)$
for both forward filtering and backward sampling (FFBS). After finding the pos-
teriors, complexity is $O(TI^2)$ to map the results to the DGM, and thus the total
computation is $O(TI^2)$. In contrast, with MCMC having $M$ sample draws for each
flow, the total complexity is $O(MTI^2)$, where $M$ is typically of the order $10^3 - 10^4$
to ensure convergence. So FFBS is faster, and furthermore, can be implemented in parallel.

The innovation here is to use the efficiently implemented BDFM as a path to fitting the gravity model, the latter now involving time-varying effect parameter processes. However, the affinity parameters $\gamma_{ijt}$ are not constrained as a function of covariates of any kind, simply treated as dynamic, random effects. This leads to a direct parameter mapping between the BDFM and 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 for Bayesian Emulation of DGMs by BDFMs

Given a set of flow rates $\phi_{ijt}$ for all $i=1:I, j=0:I$, excluding $i=j=0$ at each time $t=1:T$, the mapping to DGM parameters in eqn. (5.1) requires aliasing constraints to match dimensions. The common zero-sum constraint is adopted on logged values. Define $h_t = \log(\mu_t)$, $a_{it} = \log(\alpha_{it})$, $b_{jt} = \log(\beta_{jt})$ and $g_{ijt} = \log(\gamma_{ijt})$. Using the + notation to denote summation over the range of identified indices, constrain via $a_{+t} = b_{+t} = 0$, $g_{+jt} = g_{i+t} = 0$ for all $i, j, t$. We then have a bijective map between BDFM and DGM parameters; given the $\phi_{ijt}$ we can directly compute implied, identified DGM parameters. The DGM is saturated– there are exactly as many parameters in the DGM as there are observations in the data set. However, the emulating BDFM enforces smoothness over time in parameter process trajectories, and this acts to substantially reduce the effective model dimension– one key attribute of the emulation approach. Note that this overall strategy necessarily adopts the view that temporal structure for DGM parameter processes are those induced by the mapping from BDFMs. In current form, the evolution of latent rate processes in the latter are random walks with levels of variation defined by rate-specific discount factor.
sequences, so the evolutions for the induced DGM parameters will be more elaborate but still basically of random walk form.

5.2.1 Computation

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

- the baseline level $\mu_t = \exp(h_t)$ where

$$h_t = f_{+++}/((I + 1)^2 - 1);$$

- for each $i = 1:I$, the origin node main effect $\alpha_{it} = \exp(a_{it})$ where

$$a_{it} = f_{i+t}/(I + 1) - h_t;$$

and the origin node main effect of the external $\alpha_{0t} = \exp(a_{0t})$ where

$$a_{0t} = f_{0+t}/I - h_t;$$

- for each $j = 1:I$, the destination node main effect $\beta_{jt} = \exp(b_{jt})$ where

$$b_{jt} = f_{+=j}/(I + 1) - h_t;$$

and the destination node main effect of the external is $\beta_{0t} = \exp(b_{0t})$ where

$$b_{0t} = f_{+0t}/I - h_t;$$

- for each $i = 0:I$ and $j = 0:I$ excluding $i = j = 0$, the affinity $\gamma_{ijt} = \exp(g_{ijt})$ where

$$g_{it} = f_{ijt} - h_t - a_{it} - b_{jt}.$$ 

In the data analysis below, I 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; this can unduly impact 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 (e.g. Bishop et al., 1975, Chapter 5). This is implemented by simply restricting the summations in identifiability constraints to node pairs for which $x_{ijt} > d$, for some small $d$, and adjusting divisors to count the numbers of terms in each summation. For this study, I use $d = 3$. With this adjustment, very small $\phi_{ijt}$ appropriately leads to very small affinities $\gamma_{ijt}$, i.e., small rates underlying very sparse flows.

5.2.2 Occupancy ratio in DGM

For the emulator of each individual flow, the occupancy ratio is introduced as a modification to the rate of the latent gamma process. This ratio is defined as that of the occupancy count at time $t - 1$ to time $t - 2$. In the original multinomial assumption, variation in flows can be explained by change in either occupancy $n_{i,t-1}$ or transition probability $\theta_{ijt}$. By introducing the modification term, the influence of occupancy count is eliminated, and thus it is possible to have a smoother Poisson rate process.

Though inference on the transition probabilities remains the same as without the modification term, it does influence the results mapped to the gravity model. It is shown in Section 3.3.2 that, if using the same prior values and the same discounted process, the relationship between the latent processes with and without the modification term is

$$\phi_{ijt}^* = m_{it} \phi_{ijt}. \quad (5.2)$$
The relationship between the DGM mapping results of both processes can be then derived as follows ($\phi_{00t}$ is set to be 0 here for simplicity of notation):

- Overall level $\mu^*_t$:
  \[
  \log \mu^*_t = \sum_{i=0:I} \sum_{j=0:I} \log \phi^*_{ijt}/(I + 1)^2 \\
  = \sum_{i=0:I} \sum_{j=0:I} (\log \phi_{ijt} + \log m_{it})/(I + 1)^2 \\
  = \sum_{i=0:I} \sum_{j=0:I} \log \phi_{ijt}/(I + 1)^2 + \sum_{i=0:I} \log m_{it}/(I + 1) \\
  = \log \mu_t + \sum_{i=0:I} \log m_{it}/(I + 1).
  \]

- Origin effect $\alpha^*_{it}$ for $i = 0:I$:
  \[
  \log \alpha^*_{it} = \sum_{j=0:I} \log \phi^*_{ijt}/(I + 1) - \log \mu^*_t \\
  = \sum_{j=0:I} (\log \phi_{ijt} + \log m_{it})/(I + 1) - (\log \mu_t + \sum_{i=0:I} \log m_{it}/(I + 1)) \\
  = \left[ \sum_{j=0:I} \log \phi_{ijt}/(I + 1) - \log \mu_t \right] + \log m_{it} - \sum_{i=0:I} \log m_{it}/(I + 1) \\
  = \log \alpha_{it} + \sum_{j \neq i} [\log m_{it} - \log m_{jt}] / I.
  \]

- Destination effect $\beta^*_{jt}$ for $j = 0:I$:
  \[
  \log \beta^*_{jt} = \sum_{i=0:I} \log \phi^*_{ijt}/(I + 1) - \log \mu^*_t \\
  = \sum_{i=0:I} (\log \phi_{ijt} + \log m_{it})/(I + 1) - (\log \mu_t + \sum_{i=0:I} \log m_{it}/(I + 1)) \\
  = \sum_{i=0:I} \log \phi_{ijt}/(I + 1) - \log \mu_t \\
  = \log \beta_{jt}.
  \]
- Affinity effect $\gamma_{ijt}$ for $i, j = 0: I$ excluding $i=j=0$:  
\[
\log \gamma_{ijt}^* = \log \phi_{ijt}^* - \log \alpha_{it}^* - \log \beta_{jt}^* - \log \mu_t^*
\]
\[
= (\log \phi_{ijt} + \log m_{it}) - \left[ \log \alpha_{it} + \log m_{it} - \sum_{i=0: I} \log m_{it}/(I + 1) \right] - \log \beta_{jt} \\
- \left[ \log \mu_t + \sum_{i=0: I} \log m_{it}/(I + 1) \right]
\]
\[
= \log \phi_{ijt} - \log \alpha_{it} - \log \beta_{jt} - \log \mu_t
\]
\[
= \log \gamma_{ijt}.
\]

To conclude, the two processes share the same destination effect process and affinity effect process, but the overall level and origin effect processes are different. The overall level of the modified process can be regarded as the original overall level modified by the geometric mean of the occupancy ratios of all nodes as origins, while the origin effect process of node $i$ is modified by the geometric mean of the difference between the occupancy ratios of node $i$ and the remaining nodes.

### 5.3 Bayesian Credible Value

In the DGM, in order to informally summarize the significance of the origin, destination and affinity effects, a Bayesian credible value is defined to characterize the time varying significance. At each time point, an influential origin/destination/affinity effect is indicated by an effect value significantly either larger or smaller than 1, while a value close to 1 means a negligible effect on the flow. For affinity $\gamma_{ijt}$ with retrospective posterior c.d.f $\Phi_{ijt}(\gamma)$, the Bayesian credible value is calculated as

\[
p_{ijt} = \min\{\Phi_{ijt}(1), 1 - \Phi_{ijt}(1)\},
\]

as shown in the shaded areas in Figure 5.3. A sense of significance can be obtained by looking at the time series or some summary of the Bayesian credible values over a certain time period.
5.4 Sequential Monitoring of Fox News Data

5.4.1 February 23rd 2015, 09:00–10:00am.

The gravity model decomposition can be directly computed for the 9:00-10:00 a.m. data on February 23rd. Following the BDFM analysis as in Section 3.4, posterior simulations (5,000 Monte Carlo samples) of flow rates are mapped to posterior samples from the corresponding DGM.

Inference on outflow (origin) parameters $\alpha_{it}$ and inflow (destination) parameters $\beta_{jt}$ for six chosen nodes are shown in Figures 5.2 and 5.3. The posteriors for origin effects show that large-scale site-segments, such as “Homepage”, have higher values of $\alpha_{it}$, while site-segments with low or zero flows, such as “Science”, naturally have lower values. Across all site-segments, subsets show similar patterns but there are also major differences apparent. In particular, the posterior analysis shows that several site-segments, such as “Homepage” and “Entertainment”, are substantially above average as both origin and destination nodes. Several nodes, such as “Opinion”, have above (or below) average destination effects but their origin effects are about the norm. These distinctions between the two effects show the roles of $\alpha_{it}$ and $\beta_{jt}$.
Figure 5.2: Smoothed trajectories of selected node-specific outflows $\alpha_{i,1:T}$. The + symbols indicate empirical values computed from the raw data (with cases of 0 occupancy leading to missing values).

Figure 5.3: Smoothed trajectories of selected node-specific inflows $\beta_{j,1:T}$. 
\( \beta_{jt} \) as representing common factors across the origin and destination of the flows node-by-node. They are also naturally related over time in most site-segments; this captures the effect of the overall scale, or popularity, of some site-segments such as “Homepage” and “Entertainment” here, while also showing up in clearly similar patterns over time in less active site-segments, such as “World”. Further, while some trajectories are relatively stable over time, others show marked changes in the node-specific effects over the morning period. “Opinion”, for example, has a roughly constant and above-average inflow effect for much of the morning, but it decays toward the end of the morning period; “World” starts off at a level slightly above the norm in both inflow and outflow effects, which both then increase substantially as the morning progresses; “Science”, in contrast, has roughly constant effects across the full time period.

For the affinity effects \( \gamma_{ijt} \), there are \((I + 1)^2 - 1\) 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 \(I\). Here \(I = 22\), so the number of \( \gamma_{ijt} \) for fixed \(t\) is 528, and it is impossible to examine all those results in this application. Instead, a few affinity effects with interesting behavior are picked here.

Traffic from “Homepage” to other site-segments is central to understanding normal patterns of variation as “Homepage” is usually the landing page for visitors. Where users tend to go next, and how the flow patterns begin to evolve from “Homepage” generally, is of key interest in computational advertising. Figure 5.4 displays some relevant summaries for flows from “Homepage” to “Science”. First, overall counts and also relative frequencies of transitions tend to increase over this morning period. The BDFM appropriately tracks these slowly evolving trends (while not, of course, predicting them). Second, origin/outflow and destination/inflow parameter processes are relatively constant over time, although the former exhibits a slight increase in the later morning period. Of more interest—and highlighting the flexibility
**Figure 5.4**: Posterior summaries for DGM parameters for transitions from node $i=$Homepage $\rightarrow j=$Science. As in Figures 5.2 and 5.3, the + symbols indicate empirical values (with cases of 0 occupancy leading to missing values). **Upper left**: Posterior trajectory for the latent flow level process $\phi_{ijt}$ with raw counts (crosses). **Upper right**: Posterior trajectory for the transition probability process $\theta_{ijt}$ with raw frequencies (crosses). **Center left**: Posterior trajectory for the Homepage origin (outflow) effect process $\alpha_{it}$. **Center right**: Posterior trajectory for the Science destination (inflow) effect process $\beta_{jt}$. **Lower left**: Posterior trajectory for the Homepage:Science affinity/interaction process $\gamma_{ijt}$. **Lower right**: Corresponding trajectories of Bayesian credible values.
and incisiveness of the BDFM→DGM emulation map—the affinity process is clearly time-varying. Fox News visitors tend to be much less likely to go to “Science” from “Homepage” during the first half of the morning period, but this trends upwards after about 40 minutes. Note that, while the overall outflow and inflow processes are roughly constant over time for this pair of nodes, the raw data indicate continued growth in traffic towards the end of the time period, and thus the model responds by inferring the upward drift in the interaction/affinity process.

The pattern over time of the affinity effect also relates to “dynamic sparsity.” While it is impossible to have models that are explicitly exploring sparsity in main or affinity effects, the emulation approach has enabled the identification of an affinity process that is relevant for some periods of time but irrelevant for others. In contrast to this easy and scalable methodology, other more formal Bayesian approaches to dynamic sparsity modeling (e.g. Nakajima and West, 2013a,b, 2015; Zhou et al., 2014) are difficult or impossible to reliably implement in a sequential context, being reliant on intense MCMC methods for batch data processing. It should be noted, however, that this does not formally test consistency of posteriors for affinities against the value of 1, but simply explores the trajectories to generate insights. More formal assessment is possible by considering differences in affinities over time from the full posterior sample.

A second example, chosen to represent node pairs with high inflow and outflow levels, concerns transitions from “Homepage” to “Entertainment”; see Figure 5.5. Here again the trivially implemented emulation approach is able to identify a high level of stability over time in the main effects, while indicating reduced affinity in the latter part of the morning period. Considering the downward trending patterns in raw data/relative frequencies of flows from “Homepage” to “Entertainment” in the latter period, it would not otherwise be easy to isolate these patterns as idiosyncratic to the node pair. Inferences reflected here on the trajectory of the affinity process
clearly show significantly reduced levels later on, with $\gamma_{ijt}$ falling from around 6.5 to around 5.5; relative to the network-wide structure, high affinity is maintained throughout at a practical level, but at a reduced level later on for this node pair.

A node’s affinity effect of attracting its own visitors—referred to self-affinity—is also important; a high self-affinity reflects strong attraction of the site’s content. The node “Entertainment” is a good example, as in Figure 5.6. Its origin and destination effects are relatively low, both are less than 3; however, the self-affinity is quite high, around 1,000, which suggests strong interest in entertainment news, and thus it should be a good target for entertainment related advertisement such as entertainment magazines or movie trailers.

There are also pairs of nodes with interesting interactions. Two examples of transitions from “Entertainment” to “Leisure” and from “World” to “US” are chosen here, as shown in Figures 5.7 and 5.8 correspondingly. The affinity effect from “Entertainment” to “Leisure” is very weak, as suggested by the fact that its temporal Bayesian credible value is above 0.05 most of the time. In the posterior inferences on transitions from “World” to “US”, the directed interaction plays a more important role, and the increasing affinity effect partially explains the increase in the latent rate trend.

Heat-maps in Figure 5.9 show aspects of relationships in some estimated DGM parameters across nodes and across time. These show patterns in the values of the posterior means of $\alpha_{it}, \beta_{jt}, \gamma_{ijt}$ over time; this includes all main effects and the directional affinities/interactions of all network nodes for flows from site-segment 1, the “Homepage”. For visual display, in all three images the nodes are ordered in terms of correlation over time with the estimated “Homepage” outflow effect $\alpha_{it}$. The values shown are standardized within each image so that the min/max across time are 0/1. Note common patterns that reflect interdependencies in dynamics across subsets of network nodes. The $\alpha_{it}$ image reflects natural evolution in the morning period

65
Figure 5.5: Posterior summaries for DGM parameters for transitions from node $i=$Homepage $\to j=$Entertainment, with details as in Figure 5.4.

Figure 5.6: Posterior summaries for DGM parameters for transitions staying at node $i=$Entertainment, with details as in Figure 5.4.
Figure 5.7: Posterior summaries for DGM parameters for transitions from node $i=$ Entertainment $\rightarrow j=$ Leisure, with details as in Figure 5.4.

Figure 5.8: Posterior summaries for DGM parameters for transitions from node $i=$ World $\rightarrow j=$ US, with details as in Figure 5.4.
of traffic from network nodes, showing the increasing rates of transitions from some of the more popular, core site-segments (Homepage, Politics, World, Entertainment and others) in later morning. The $\beta_{jt}$ image shows consonant patterns in a subset of these core site-segments—in that their attractiveness increases in later morning—but with some clear exceptions. For the “Homepage” affinity processes $\gamma_{1jt}$, there are quite a few site-segments that see increased incremental traffic rates in the first half of the morning, which then drop later on.

5.4.2 Comparison across days

The study covers morning (09:00-10:00 a.m.) and afternoon (01:00-02:00 p.m.) periods on each of 6 days, as already discussed and explored in Section 3.4. Moving to the DGM, I 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 5.10 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 24th and March 3rd. On February 24th, the afternoon flow is significantly lower than the morning flow, while the morning flow that day is much larger than across other days. One plausible reason is increased morning traffic following the Academy Awards ceremony, with a resulting lull in the afternoon. 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 Fox News 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.

An advantage of the DGM representation is that it allows investigation of such speculative explanations. For example, examination of destination effects (not shown)
Figure 5.9: Heat-map images of standardized posterior means of DGM parameters for all nodes across time in the February 23rd am period. Means are max-min standardized separately for each of the $\alpha_{it}$, $\beta_{jt}$ and $\gamma_{1jt}$ (1=Homepage). Shading runs from 0 (dark blue) to 1 (yellow) on these standardized scales. The nodes are ordered based on correlation over time of the main outflow effects $\alpha_{it}$ with that of Homepage; this is an arbitrary ordering chosen simply for visual presentation.
Figure 5.10: DGM-based inference on trajectories of baseline flow levels $\mu_t$ across days with 95% intervals. Red are for morning periods, blue lines are for afternoons.

$\beta_{5,1:T}$ confirm that the “Entertainment” node was unusually popular on the morning of February 24, and that the “Politics” and “Opinion” nodes were unusually popular on March 3 in the afternoon, compared to similar flows on other days.

5.5 Model Mapping for DGMs by LLGMs

The data analyzed in this section is the Fox News data with Adex Category as network segments on September 17th, 2015. This streaming network data spans the entire chosen day with $T = 284$ time points. At each time point $t$, the data is stored as a square matrix with entries $x_{ijt}$ representing the number of visitors traveling from node $i$ to $j$ during the time period $(t - 1, t]$. With the retrospective trajectories of log rate parameters generated from the LLGM, the mapping strategies in Section 5.2 are directly extensible with the same identification constraints and low rate flow controls. The results of DGM mapping provide a detailed understanding
of the network structure in four aspects: i) the baseline process $\mu_t$ characterizing the general activity intensity, ii) the origin effect $\alpha_{it}$ as node’s influence on outgoing flows from node $i$, iii) the destination effect $\beta_{jt}$ as attractiveness for incoming flows to node $j$, and iv) directed pairwise affinity $\gamma_{ijt}$ as interactions between nodes $i \rightarrow j$.

In this section, I first analyze the four components of DGM results separately to explore how users’ visiting behavior and the network structure vary across the day. Related results are summarized in Sections 5.5.1, 5.5.2 and 5.5.3. This is followed by exploratory posterior analysis to investigate clusterings of categories over time, developed in Section 5.5.4.

5.5.1 Overall mean

The baseline activity level reflects the daily trend of users’ visiting intensity. Generally, this maintains a high level during the day and early evening, and drops to a relatively low level at night, as in Figure 5.11. A day starts at midnight when the baseline activity intensity is at around 1.025 with a tight credible interval. Since more and more users go offline, this overall activity level keeps going down until reaching the lowest level at around 4:00, and then users begin to wake up and start visiting the website, so the overall intensity increases sharply until around 8:00 when the trend becomes flat. During the working hours (8:00-16:00), the website maintains a relatively high level of activity with three small bumps at around 10:00, 12:30 and 15:00; these correspond to regular times for taking breaks from work in the morning, at noon and in the afternoon. Then there is a slight decreasing trend from 16:00 to 18:00 when people usually commute from work and spend time at dinner. After dinner, people usually have more free time that they could spend on the Internet, and the trend increases to a peak at around 20:00. Afterwards, the majority of users go offline and thus the activity level goes down rapidly from about 21:00 to the end of the data period.
Figure 5.11: DGM-based inference on smoothed trajectories of baseline activity level of Fox News data on September 17th, 2015 with 95% credible interval. Blue + symbols indicate empirical values computed from the raw data.

5.5.2 Origin and destination effects

For most of the categories, there is not much difference between the trend of origin and destination effects of the same category, so I only analyze origin effects in detail here. For origin effects $\alpha_{it}$, there are mainly three different kinds of trends, and three examples of each kind are chosen for illustration in Figures 5.12 and 5.13. The first kind has trend similar to that of the overall activity level $\mu_t$. The origin effect of categories in this kind increases to a high level during the early morning (from about 4:00 to 8:00), keeps a high activity level during the day and early night (from about 8:00 to 20:00), and then goes down sharply after 20:00 until 4:00 the next day. Such categories include “Arts & Entertainment”, “News”, and “Shopping”. Those categories are also among the most popular categories in the network and evidently have content most interesting to users all day long. The second kind only keeps a stable high activity level during the day time (from about 8:00 to 16:00), but drops to a low level at other times; see, for example, “Health”, “Beauty & Fitness”, and
“News/Weather”. For categories of this kind, it can be interpreted that people are more interested in related content during the day time. For example, as shown in the trend of category “News/Weather”, most users are interested in the weather when they leave home for school or work, i.e. around 8:00 in the morning, to decide what kind of clothes they should wear and whether they should bring an umbrella with them. They still pay close attention to the weather during the day, but they lose their interest in weather in the evening since most people would stay home and do not go outside then. The third kind of trend keeps increasing from early morning (about 4:00) to night (about 20:00), reaching its peak at around 20:00 and likely to have other interesting peaks. Categories “Games”, “Online Communities/Social Networks”, and “Food & Drink/Food/Baked Goods” are some examples of this kind. Such categories attract more and more people during the day, and there are times when people have particular interest in them. For instance, Category “Food & Drink/Food/Baked Goods” has two peaks during the day: one in the early morning (around 8:00) for cooking breakfast, and the other in the afternoon (around 17:00) for cooking dinner. The day I choose for analysis, September 17th of 2015, is a workday (Thursday), and most people only cook two meals—breakfast and dinner—since most people eat lunch at their school or workplace. The two peaks in people visiting this category can reflect this.

5.5.3 Affinity effect

Affinity effects can reflect the communication between different categories. Pairs of categories with high affinity can possibly reflect that people have more common interest in accessing information or purchasing goods in both categories, thus affinity effects are the most valuable for online advertisement placement. There are four different kinds of affinity effects: affinity effects of i) staying at a certain category
Figure 5.12: Smoothed trajectories of selected node-specific origin effects $\alpha_{i,1:T}$ with 95% credible interval. The + symbols indicate empirical values computed from the raw data.
Figure 5.13: Smoothed trajectories of selected node-specific destination effects $\beta_{j,t}$ with 95% credible interval. The + symbols indicate empirical values computed from the raw data.
γ_{iit}, ii) entering the network γ_{0jt}, iii) leaving the network γ_{i0t}, and iv) between two different categories γ_{ijt}, and I analyze each of them separately.

1. First of all, the affinity effects of staying at a single category. A category with a high self-affinity reflects its ability to provide users interesting content and encourage visitors to spend more time on it; this is thus a good target for online ad placement. By inspecting the trend of self-affinity processes, I can focus on daily periods with high affinity effects as ad placement target. For those categories with high self-affinity, there are mainly three different kinds of trends, and I analyze each of them with examples in detail.

The first kind resembles the overall activity intensity trend \( \mu_t \), and the self-affinity at \( i = \text{"Arts & Entertainment"} \) is a good example, as in Figure 5.14. The self-affinity maintains a high level during the day and at evening (8:00-20:00) with a huge bump at night (around 20:00).

The second kind has a high activity level during the day (8:00-16:00), and then drops to a relatively low level at night. Such categories include “Finance/Investing” and “Computers & Electronics/Software” as in Figures 5.15 and 5.16. It should be noted that the self-affinity of “Finance/Investing” has three peaks: the first one is around 10:00 in the morning, the second one is around 15:00 in the afternoon, and the last one is around 20:30 at night, which indicates three time periods when people are most interested in financial and investing content. Category “Computers & Electronics/Software” is of special interest, since for most categories, the self-affinity drops a little bit after 8:00 and is not very high at noon, while that of “Computers & Electronics/Software” keeps increasing in the entire morning and reaches its highest at noon and then declines, which indicates that it is better to put more ads related to computer software at noon than other content.
Figure 5.14: Posterior summaries for DGM parameters for transitions staying at $i = \text{Arts & Entertainment}$. The + symbols indicate empirical values. **Upper left:** Posterior trajectory for the latent flow level process $\phi_{it}$ with raw counts (crosses). **Upper right:** Posterior trajectory for the transition probability process $\theta_{it}$ with raw frequencies (crosses). **Center left:** Posterior trajectory for the Arts & Entertainment origin (outflow) effect process $\alpha_{it}$. **Center right:** Posterior trajectory for the Arts & Entertainment destination (inflow) effect process $\beta_{it}$. **Lower left:** Posterior trajectory for the Arts & Entertainment self-affinity/interaction process $\gamma_{it}$. **Lower right:** Corresponding trajectories of Bayesian credible values.
Figure 5.15: Posterior summaries for DGM parameters for transitions staying at $i = \text{Finance/Investing}$ with details as in Figure 5.14.
Figure 5.16: Posterior summaries for DGM parameters for transitions staying at $i = \text{Computers \& Electronics/Software}$ with details as in Figure 5.14.
Figure 5.17: Posterior summaries for DGM parameters for transitions staying at $i = \text{Arts \& Entertainment/TV \& Video}$ with details as in Figure 5.14.
The third kind keeps increasing from early morning until reaching its peak at night. It is a very unusual trend, and “Arts & Entertainment/TV & Video” is one of the few examples, as in Figure 5.17. For categories of this kind, visitors’ interest keeps the same level during the day, and there are no specific periods when related ads are of special interest.

2. Second, the affinity effects of entering the network. A category with high entering affinity means it is more attractive to users outside the website, and the ads shown on it should be given more attention. There are mainly four kinds of trends, and I analyze each of them in detail separately. It should be noted that the “transition probability” \( \theta_{0i} \) shown here is actually the proportion of incoming users into category \( i \) versus the total number of incoming users into Fox News website, which is not exactly the transition probability, and serves as an approximate measure of a category’s popularity to the external audience.

The first kind of trajectory only has one huge peak in the early morning; examples include the incoming flows to “News/Weather” and “News/Politics”. The peak for both of the two categories is at about 6:00 in the morning, as in Figures 5.18 and 5.19. For the category “News/Weather”, people are most interested in weather before they leave for school or work in the morning, which leads to this peak in the early morning. Therefore, ads related with weather, such as ads for umbrellas, are suggested to be placed in this category during the morning. As for the category “News/Politics”, the chosen date is the day after a Republican debate, so people are more interested in politics in early morning this day to check about details of the debate and media response; this peak is highly likely to be driven by this event. In order to modify ad placement strategy due to this trend feature, it is suggested to intervene and improve the probability of showing political ads if there is an influential political event.
Figure 5.18: Posterior summaries for DGM parameters for transitions entering $i =$ News/Weather. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{0it}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{0it}$ with raw frequencies (crosses). Center left: Posterior trajectory for the external origin (outflow) effect process $\alpha_{0it}$. Center right: Posterior trajectory for the News/Weather destination (inflow) effect process $\beta_{it}$. Lower left: Posterior trajectory for the News/Weather entering affinity/interaction process $\gamma_{0it}$. Lower right: Corresponding trajectories of Bayesian credible values.
Figure 5.19: Posterior summaries for DGM parameters for transitions entering $i =$ News/Politics with details as in Figure 5.18.
Figure 5.20: Posterior summaries for DGM parameters for transitions entering $i = \text{Health/Pharmacy/Drugs \& Medications}$ with details as in Figure 5.18. An interesting bump is noted in the afternoon (about 12:00 - 16:00).
Figure 5.21: Posterior summaries for DGM parameters for transitions entering $i =$ Finance/Investing with details as in Figure 5.18. Note a maintained high level during the day (8:00 - 16:00).
Figure 5.22: Posterior summaries for DGM parameters for transitions entering $i =$ Shopping with details as in Figure 5.18.
Figure 5.23: Posterior summaries for DGM parameters for transitions entering $i =$ Beauty & Fitness/Fashion & Style with details as in Figure 5.18.
The second kind maintains a high activity level with bumps during the day (8:00-16:00), and stays relatively low level except these hours. As in Figures 5.20 and 5.21, categories such as “Health/Pharmacy/Drugs & Medications” and “Finance/Investing” have this kind of entering affinity, which means the content in those categories is more interesting to users during the working time. Therefore, medical and financial ads, especially those targeting employees in those industries, should be placed in the corresponding category during the day. Moreover, the huge bump in the afternoon of the entering affinity of “Health/Pharmacy/Drugs & Medications” should be given special attention.

The third kind of trajectory increases to a high intensity in the morning and then stays stable during the entire day, which means website users keep a high level of interest all the day long. The entering affinity of “Shopping” is a good example as in Figure 5.22, so relevant ads should be placed constantly onto the website though maybe not with a high rate.

The fourth kind of trajectory is almost stable during the entire day but with some occasional bumps which are signals for adding more related ads. For example, there are three interesting peaks in the entering affinity of “Beauty & Fitness/Fashion & Style” as in Figure 5.23: the first one is about 7:00-8:00 in the morning, the second one is about 12:00-16:00 in the afternoon, and the third one is about 20:00-24:00 at late night. The chosen date is during the New York Fashion Week, which presumably explains why people pay attention to this category more often than usual and thus indicates the requirement for more fashion ads.

3. Third, the affinity effects of flows exiting from category $i$. Ad placement strategy should ensure that users have access to the ads the online advertising companies want to show them right before they leave the Fox News website.
Figure 5.24: Posterior summaries for DGM parameters for transitions exiting from $i =$ Arts & Entertainment. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{it}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{it}$ with raw frequencies (crosses). Center left: Posterior trajectory for the Arts & Entertainment origin (outflow) effect process $\alpha_{it}$. Center right: Posterior trajectory for the external destination (inflow) effect process $\beta_{it}$. Lower left: Posterior trajectory for the Arts & Entertainment leaving affinity/interaction process $\gamma_{ijt}$. Lower right: Corresponding trajectories of Bayesian credible values.
Figure 5.25: Posterior summaries for DGM parameters for transitions exiting from $i =$ Food & Drink/Cooking & Recipes with details as in Figure 5.24.
Figure 5.26: Posterior summaries for DGM parameters for transitions exiting from $i =$ News/Weather with details as in Figure 5.24.
Figure 5.27: Posterior summaries for DGM parameters for transitions exiting from \( i = \) Arts & Entertainment/Music & Audio with details as in Figure 5.24. A huge peak is noted at night (around 20:00).
A detailed analysis of the exiting affinity is thus very important. There are mainly three different kinds of trends in all exiting affinity processes and I analyze each of them in detail with examples.

The first kind of trajectory keeps increasing in the morning until reaching a high level at about 8:00, and then stays at a stable and high intensity almost the entire day. Most of the categories with this kind of exiting affinity attract their certain group of users, and those users usually do not visit other categories after they find the targeted content; examples include “Arts & Entertainment” and “Food & Drink/Cooking & Recipes” as in Figures 5.24 and 5.25. Though the affinity intensity is quite stable during the day, there might be bumps occasionally which informs the requirement for putting in more related ads. For instance, there is a small bump during 16:00-18:00 in the exiting affinity of “Food & Drink/Cooking & Recipes” which could be caused by increased interest in cooking instructions while people are preparing food for dinner. The implication is to add more ads about cooking recipes or ingredients into this category, otherwise we will lose those users since they are leaving Fox News.

The second kind of trajectory has a huge peak while staying low for the rest of time. For example, there is a peak which is about four times higher than the average level in the exiting affinity process of the category “News/Weather”, as in Figure 5.26. Considering the fact that its entering affinity process also has the same pattern, it can be interpreted as most visitors come to visit this category just for weather information and would leave the Fox News website right after they check the weather, so weather related ads should be shown directly in this category to ensure the targeted users are able to see them.

The third kind of trajectory keeps increasing from 4:00 in the morning until
the end of day, which indicates the webpages of these categories are losing more and more of their users. Such categories include “Arts & Entertainment/Music & Audio”, as in Figure 5.27. Therefore, there is a need for adjusting related ad placements according to this trend.

4. Fourth, transition affinity from a category to another. A high affinity value between categories indicates strong interactions between them, and thus become a sign for putting more ads related to the latter category also in the webpages belonging to the former category. Therefore, the bumps and peaks in the process are of special interest. Here four examples are chosen for detailed analysis.

The first example is from “Online Games” to “Video Games”. As in Figure 5.28, it has a huge bump in the evening and at night (16:00-24:00) during which the average is six times higher than the usual intensity level. This is a strong indication that users who have read about online games also have high interest in computer and video games during this period.

The second example is from “News” to “News/Local News”. As in Figure 5.29, this affinity process has three interesting peaks in a day: the first one is around 8:00 in the morning, the second one is at noon, and the last one is around 20:00 at night. Those times are when people typically have some time off work, and those peaks indicate that people who are interested in news also check the local news quite often when they have time.

The third example is from “Home & Garden” to “Reference/General Reference/How to DIY & Expert Content”. There is an interesting bump in the morning, which is about 8:00-12:00, as in Figure 5.30. Moreover, both the origin effect of “Home & Garden” and the destination effect of “Reference/General Reference/How to DIY & Expert Content” are low, while the affinity effect be-
Figure 5.28: Posterior summaries for DGM parameters for transitions from node $i =$ Games/Online Games $\rightarrow j =$ Games/Computer & Video Games. The + symbols indicate empirical values. Upper left: Posterior trajectory for the latent flow level process $\phi_{ijt}$ with raw counts (crosses). Upper right: Posterior trajectory for the transition probability process $\theta_{ijt}$ with raw frequencies (crosses). Center left: Posterior trajectory for the Games/Online Games origin (outflow) effect process $\alpha_{it}$. Center right: Posterior trajectory for the News/Weather destination (inflow) effect process $\beta_{jt}$. Lower left: Posterior trajectory for the Games/Online Games : Games/Computer & Video Games affinity process $\gamma_{ijt}$. Lower right: Corresponding trajectories of Bayesian credible values.
Figure 5.29: Posterior summaries for DGM parameters for transitions from node $i = \text{News}$ to $j = \text{News/Local News}$ with details as in Figure 5.28.
Figure 5.30: Posterior summaries for DGM parameters for transitions from node $i = \text{Home & Garden}$ to $j = \text{Reference/General Reference/How to DIY & Expert Content}$ with details as in Figure 5.28.
Figure 5.31: Posterior summaries for DGM parameters for transitions from node $i = \text{News/Technology News}$ to $j = \text{Shopping}$ with details as in Figure 5.28.
tween them is huge, which is an indication for their strong interaction. That is to say, people who are interested in doing home decorating and gardening refer to the content about doing DIY a lot, especially in the morning.

The last example is from “News/Technology News” to “Shopping”. As in Figure 5.31, this affinity process has a huge peak at night (around 20:00), which means the users who have read technology news tend to explore shopping at night. This can serve as a sign that on the webpages about technology news, online advertisement strategists might increase shopping related items such as links to online department store websites.

5.5.4 Singular value decomposition for DGM results

In order to have a better understanding of the network structure dynamics, especially how the origin and destination effects vary with time, some further exploration of posterior trajectory summaries is of interest. Decomposing estimates of effects over time using singular value decomposition (SVD) analysis is explored here. The posterior means of origin/destination effects are rearranged in a matrix $X$ of dimension $I \times T$ with row $i = 1, \cdots, I$ indicating the origin/destination effect of category $i$ over all time, and column $t = 1, \cdots, T$ indicating the origin/destination effects of all categories at a time $t$.

The matrix $X$ can be decomposed into an $I \times I$ unitary matrix $A$, an $I \times I$ diagonal matrix $D$ with diagonal singular values in decreasing order and an $I \times T$ matrix $F$ as

$$X = ADF', \quad (5.4)$$

with $f_i$, column $i$ in $F$, indicating the loadings of effects on empirical values in matrix $A$.

The columns in $A$ indicate the entry-to-pattern similarity, which is useful in exploring the clustering of entries. The columns in $F$ show the time-to-pattern
similarity, relevant to exploring the time varying structure of different clusters.

In the analyzed data set, there are 237 categories and 284 time points in both of the origin and destination effect matrices. The decomposition results, especially the time varying structure of some factors as well as their loadings $f_i$’s, show some interesting findings. For both origin and destination effects, the first factor, the one with the largest singular value, accounts for more than 80% of the total variation (sum of squared singular values), as in Figures 5.32 and 5.34. Moreover, for both origin and destination effects, the trend of the first factor looks almost the same as the overall activity level as in Figure 5.11, and the categories with high factor 1 loadings show a clustering pattern that they are more likely to fall into entertainment and news categories. More details of the SVD analysis about important factors and category clusterings are now discussed.

**SVD for origin effects**

Among all the 237 factors, the 5 dominant factors account for more than 90% of the variation in estimated origin effects (Figure 5.32). The time varying pattern of the first factor roughly summarizes the daily trend of overall activity level, and that of other factors also capture various patterns across different time periods. For example, the time varying pattern of the second factor demonstrates a high activity level during the day time (8:00-16:00), and that of the third loading factor shows two peaks in the morning (8:00) and at night (20:00).

The decompositions can also show clustering behaviors of categories. For the 1st factor, the categories with top 10 loadings are “Arts & Entertainment/Celebrities & Entertainment News”, “News”, “Arts & Entertainment/TV & Video/Online Video”, “External, /News/Sports News”, “Arts & Entertainment/Movies”, “Arts & Entertainment”, “Travel”, “Sports”, “Food & Drink/Cooking & Recipes”. Most of these are about entertainment or news, and are popular in the network, as in Figure 5.33.
Some other interesting factors include the 8th, 9th, 11th and 12th. Categories ranking highest in the 8th and 9th factors are mainly about games or leisure, and those with high loadings in the 12th and 13th factors are mainly about sports or education.

**SVD for destination effects**

As in Figures 5.34 and Figure 5.35, the results of SVD for posterior mean trajectories of destination effects are quite similar to those of origin effects: the first 5 factors are the most important and explain more than 90% of the information in all destination effects, and the time varying pattern of the first factor resembles the trend of general activity level of the entire network.

However, there is also difference especially in the category clustering behaviors. Only the loadings of the 1st factor shows similar clustering results as for the origin effects, and the top 10 categories are “Arts & Entertainment/Celebrities & Entertainment News”, “News”, “Arts & Entertainment/Movies”, “Arts & Entertainment/TV
Figure 5.33: SVD results for posterior mean trajectories of origin effects: time varying first component and its loadings (data: FoxNews visiting flows on September 17, 2015).

![Figure 5.34](image)

**Figure 5.34**: Dominant factors in SVD results for posterior mean trajectories of destination effects with data of FoxNews visiting flows on September 17, 2015 (left: cumulative variation explained by factors; right: time varying patterns of five dominant factors).

### 5.5.5 Findings in daily Fox News data

With the Fox News visiting flow data, I explore users’ daily visiting preferences over the defined categories of various topics using estimated affinities. Interesting findings include a strong interaction between two unpopular categories as either origin or destination, “Home & Garden” and “Reference/General Reference/How to DIY & Expert Content”. Moreover, time-varying Bayesian credible values of the affinity process of this pair of nodes show that this interaction is especially strong in the morning.

By applying SVD to the retrospective smoothing results of the origin and destination effect processes separately, I also detect clusterings among categories and
Figure 5.35: SVD results for posterior mean trajectories of destination effects: time varying first component and its loadings (data: FoxNews visiting flows on September 17, 2015).
discover the time varying structure of different clusters. For example, in the SVD results of the origin effects, the first component accounts for more than 80% of the total variation in the data. The time varying pattern approximates the overall activity level of the whole network, and those with high loadings for the first component are categories related with entertainment.

5.6 Summary Comments

DGM maps the retrospective trajectories of individual flows to four kinds of components characterizing the processes of overall activity level, each node’s outflow influence and inflow attractiveness as well as the affinity for each directed pair. In both the hourly and daily Fox News flow data sets, some examples show the dynamics of the four components as well as the time-varying sparsity in node-node interaction effects over time, the latter identified by the defined Bayesian credible values. High node-specific or interaction effects indicate visitors’ strong interest in a certain category or moving between two categories, and thus they have potential commercial value for online advertisement placement.

With the hourly Fox News visiting flow data, analysis generates insights into users’ daily visiting preferences over the defined categories of various topics. For example, “Homepage”, “Entertainment” and “World” are popular both as origin and destination of flows, and the category “Entertainment” has very strong ability to attract its visitors to spend more time on reading related content.

In the daily Fox News visiting flow data, interesting findings include a strong interaction between two unpopular categories as either origin or destination, “Home & Garden” and “Reference/General Reference/How to DIY & Expert Content”. Moreover, time-varying Bayesian credible values of the affinity process of this pair of nodes show that this interaction is especially significant in the morning. Moreover, by applying SVD to the retrospective smoothing results of origin and destination effect
processes separately, we can detect the clusterings among categories and discover the
time varying structure of different clusters as well. For example, in the SVD results
of the origin effect processes, the first component accounts for more than 80% of the
total variation in the data. Its time varying pattern approximates the overall activity
level of the whole network, and those with high loadings for the first components are
categories related with entertainment.
6

Bayesian Sequential Monitoring

6.1 General Comments

In using dynamic models to monitor flows, one key interest is to discover patterns that seem outside the norm and which may indicate events that warrant investigation or intervention. I address this here with methodology based on the theory for Bayesian monitoring in conditionally normal dynamic linear models. While this theory of sequential Bayesian model assessment is well-established, it does not seem to have been adapted to apply to dynamic models of counts; the contributions of my work include this extension and required customization of the approach.

6.2 Bayesian Model Monitoring and Automated Intervention

Now return to the standard Poisson-gamma model, as in Section 3.1, with observed count $x_t$ at time $t$ and underlying state $\phi_t$. This general model applies to all cases: inflows to any node $j$ when $x_t = x_{0jt}$, as well as transitions from a node $i$ to $j$ when $x_t = x_{ijt}$.

The sequential Bayesian testing approach in dynamic linear models (West and
Harrison, 1997, Chapter 11) is now extended to the dynamic Poisson-gamma model. For the standard model at time \( t \), relabel the one-step predictive density at each time \( t \) as \( p_0(x_t | \delta_t, \delta_{1:t-1}, x_{0:t-1}) \). The suffix 0 indicates the standard model, and we now explicitly recognize the dependence on the discount factors that define levels of stochastic change in the underlying state process \( \phi_t \). The two elements needed for monitoring and adaptation are as follows.

A. Alternative model predictions: A synthetic alternative model at time \( t \) that requires only the specification of the alternative predictive p.d.f. \( p_1(x_t | \delta'_t, \delta_{1:t-1}, x_{0:t-1}) \). This alternative differs from the standard model only in the current discount factor \( \delta'_t < \delta_t \). This implies that \( p_1(\cdot) \) is more diffuse than \( p_0(\cdot) \) but similarly located. In the Poisson-gamma case, the implied generalized negative binomial p.d.f.s \( p_0(\cdot) \) and \( p_1(\cdot) \) have precisely the same mean but the latter has a larger variance, allowing both smaller and larger values of \( x_t \).

B. Bayes factor comparisons: To compare the standard and alternative models, use Bayes factors defined by the predictive p.d.f.s. These marginal likelihood ratios are computed based on both the time \( t \) observation and recent consecutive observations to assess consistency of this local data with predictions from the standard model relative to the more diffuse synthetic alternative. Support for the standard model is regarded as a “business as usual” signal. Support for the alternative suggests the possibility that (i) the single observation to be discrepant, a possible outlier; or (ii) a relatively abrupt change in the \( \phi_t \) process at time \( t \), beyond that predicted by the model with current discount rate \( \delta_t \); or (iii) gradual, non-abrupt change in the \( \phi_t \) process at higher levels than the norm.

Define the following:

- The time \( t \) Bayes factor

\[
H_t = p_0(x_t | \delta_t, \delta_{1:t-1}, x_{0:t-1})/p_1(x_t | \delta'_t, \delta_{1:t-1}, x_{0:t-1}),
\]

(6.1)
assessing the current observation alone.

- The lag-*h* local Bayes factor

$$H_t(h) = \prod_{r=t-h+1:t} p_0(x_r|\delta_r, \delta_{1:r-1}, x_{0:r-1})/p_1(x_r|\delta'_r, \delta'_{1:r-1}, x_{0:t-1})$$  

(6.2)

based on the most recent $h \in \{1:t\}$ observations, including $x_t$.

- The local cumulative Bayes factor

$$L_t = \min_{h \in \{1:t\}} H_t(h)$$  

(6.3)

and corresponding run-length $l_t$ such that $L_t = H_t(l_t)$.

Bayesian testing theory (West, 1986 and Chapter 11 of West and Harrison, 1997) shows that the local test measures $L_t, l_t$ are trivially updated as time evolves. At time $t$, the updated pair is

$$[L_t, l_t] = \begin{cases} [H_t, 1], & \text{if } L_{t-1} \geq 1, \\ [H_tL_{t-1}, 1 + l_{t-1}], & \text{if } L_{t-1} < 1. \end{cases}$$  

(6.4)

Past consistency with the standard model ($L_{t-1} \geq 1$) means that the entire focus at time $t$ is on the single observation $x_t$. If, however, recent evidence weighs against the standard model ($L_{t-1} < 1$), then that evidence continues to accumulate based on the new observation. The pair $[L_t, l_t]$ define a tracking signal that can be used to formally intervene by rejecting potential outliers and adopting the smaller discount factor $\delta'_t$ at such times, as well as when $L_t$ and/or $l_t$ suggest cumulative drift in the $\phi_t$ process. This operates as follows.

Specify a Bayes factor threshold $\tau$ (e.g. $\tau = 0.1$) and run-length threshold $r$ (e.g. $r = 4$). When standing at time $t$, compute the single-period Bayes factor $H_t$. Then:

- If $H_t \leq \tau$, reject $x_t$ as potentially outlying.
  - Intervene to apply reduced discount factor $\delta'_t$ at time $t \to t + 1$ in case of changes.
• If $H_t > \tau$, then proceed to update $[L_t, l_t]$ to continue monitoring in case of potential changes.
  
  – If $L_t \leq \tau$ or $l_t \geq r$,
    
    ◦ Apply reduced discount factor $\delta'_t$ to allow for adaptation to potential changes;
    
    ◦ Update using time $t$ data as usual but with this increased prior uncertainty;
    
    ◦ Reset monitor to $L_t = 1$ and $l_t = 1$.
  
  – If $L_t > \tau$ and $l_t < r$,
    
    ◦ Proceed as usual with prior-posterior and monitor updates.

• Forecast ahead as desired, then proceed to time $t + 1$.

This process is displayed in schematic form in Figure 6.1, modified from West and Harrison, 1997, Chapter 11, which also discusses choices of thresholds $(\tau, r)$. I follow the recommendations there for these choices. As discussed in Sections 3.2 and 3.3, the discount factors in the standard models are based on $\delta_t = d + (1 - d) \exp(-kr_{t-1})$ where, in the generic notation here, $r_{t-1}$ is the shape parameter of the time $t - 1$ posterior gamma distribution for $\phi_{t-1}$ and $d$ is a baseline discount rate. Therefore, the alternative discount factor $\delta'_t$ for the intervention analysis is selected as $\delta'_t = d' + (1 - d') \exp(-kr_{t-1})$ where for some baseline $d' < d$. The following studies of the Fox News network data are based on $d = 0.1$, whereas the standard models are based on values of $d$ running between 0.9 and 0.99 across the sets of inflow and transition flow models.

6.3 Monitoring for Sparse Flows

Data sparsity is a common issue in many dynamic network applications, and thus it is important to verify that the proposed Bayesian sequential anomaly detection
Figure 6.1: Schematic of monitoring, routine outlier/change assessment and automatic intervention to allow for more adaptability in times of change, as discussed and detailed in Chapter 6. This relies on a Bayes factor threshold $\tau$ (e.g. $\tau = 0.1$) and run-length threshold $r$ (e.g. $r = 4$). This schematic is a modified version of Figure 11.9 in West and Harrison (1997).

The approach is also valid for monitoring sparse flows. For a very sparse flow, the flow count would be zero at most of the time, and thus the shape parameter of the latent gamma process $\phi_t \sim Ga(r_t, c_t)$ would be very small that would lead to not only detection issues but also computation issue in coding environment. The solution is to do thresholding on the shape parameter. When doing sequential updating, once the shape parameter drops below a pre-defined threshold value, $\tau_r$, it would be set as this threshold no matter what the real value is. In the study, a threshold of 1 is chosen, and in the remaining part of this section, I will discuss how this thresholding idea with $\tau_r = 1$ influences the performance of monitoring.

As discussed in Section 6.2, in Bayesian sequential monitoring, at each time
Bayes factor is calculated to evaluate the consistence of local data to model predictions under the two testing assumptions in eqn. (6.5).

\[ M_{0t} : x_t | \phi_t \sim \text{Poi}(m_{it}\phi_t), \quad \phi_t | D_{t-1} \sim \text{Ga}(\delta^t_{t-1}, \delta t_{c-1}/m_{it}); \]
\[ M_{1t} : x_t | \phi'_t \sim \text{Poi}(m_{it}\phi'_t), \quad \phi'_t | D_{t-1} \sim \text{Ga}(\delta'_{t-1}, \delta' t_{c-1}/m_{it}). \]  

(6.5)

where \( \delta'_t < \delta_t \) indicating that more flexibility to adapt to change in Poisson rates is allowed in the alternative model.

Under \( M_{0t} \), the predictive for \( x_t \) is a negative binomial distribution

\[ x_t | \delta_t, \delta_{1:t-1}, x_{0:t-1} \sim \text{NB}(\delta_t r_{t-1}, \delta_t/c_{t-1} + m_{it}). \]  

(6.6)

Similarly, under \( M_{1t} \), the predictive for \( x_t \) is also a negative binomial distribution but different parameters, namely

\[ x_t | \delta'_t, \delta'_{1:t-1}, x_{0:t-1} \sim \text{NB}(\delta'_t r'_{t-1}, \delta'_t/c'_{t-1} + m_{it}). \]  

(6.7)

The Bayes factor can be then calculated as

\[ H_t = \frac{p_0(x_t | \delta_t, \delta_{1:t-1}, x_{0:t-1})}{p_1(x_t | \delta'_t, \delta'_{1:t-1}, x_{0:t-1})} = \frac{\Gamma(\delta_t r_{t-1} + x_t)/\Gamma(\delta_t r_{t-1})}{\Gamma(\delta'_t r'_{t-1} + x_t)/\Gamma(\delta'_t r'_{t-1})} \]  

(6.8)

Since the monitoring procedure is a combination of three measures which are all based on the Bayes factor in eqn. (6.8), in order to prove its validity, it is only necessary to show that the Bayes factor behaves consistently before and after thresholding.
6.3.1 Discounted process with time varying discount factor

The thresholding idea influences the choice of dynamic discount factor. When there are many zeros in the flow data, there is little information, so it is inappropriate to discount too much. Therefore, instead of using a constant discount factor, it is appropriate to use a time varying discount factor incorporating information of the shape parameter. Moreover, due to the thresholding, a shape parameter close to \( \tau_r \) rather than 0 means a lack of information in data. Therefore, with a specified “nominal” discount factor \( d \), I use the discount factor as in eqn. (6.9), which is a slight modification from the process used in Section 6.2. That is

\[
\delta_t = d + (1 - d)e^{-k(r_{t-1} - \tau_r)}.
\] (6.9)

This time varying discount factor process ensures that:

- \( \delta_t \) is close to \( d \) when data is sufficient, i.e., the latent gamma process has a large shape parameter \( r_{t-1} \);

- \( \delta_t \) approaches 1 when there is limited information, i.e., \( r_{t-1} \) is close to \( \tau_r \).

The idea here is to limit the amount of information to be discarded when there is limited information, while targeting a nominal rate of change otherwise. The rest of this section discusses how the thresholding idea and the proposed discount factor process influence each other. With this time varying discount factor and specifying \( k \) and \( d \) in certain range, once there is a time \( t_0 \) when there is enough information so that \( r_{t_0} \geq \tau_r \), the gamma shape parameter under the original assumption \( M_{0t-}\delta_t r_{t-1} \) would be always larger than or equal to \( \tau_r \) after \( t_0 \), and thus \( r_t = \delta_t r_{t-1} + x_t \geq \tau_r \) for all \( t = t_0 + 1 : T \). In this case, there is no difference whether thresholding is applied or not, and thus it is necessary to check when \( r_t \) would become equal to or larger than \( \tau_r \). In the FFBS algorithm, the shape parameter is updated as in eqn. (6.10),
\[ r_t = \delta r_{t-1} = dr_{t-1} + (1 - d)r_{t-1}e^{-k(r_{t-1} - \tau_r)}. \quad (6.10) \]

Define a function \( f(\cdot) \) by

\[ f(r) = dr + (1 - d)re^{-k(r - \tau_r)}. \quad (6.11) \]

For the defined function, \( f(\tau_r) = \tau_r \), and the first derivative is

\[ f'(r) = d + (1 - d)(1 - kr)e^{-k(r - \tau_r)}. \quad (6.12) \]

In order to explore the conditions where \( f(r) \geq \tau_r \), \( \forall r \geq r_0 \), solve the equation \( f'(r) = 0 \) and find its nulls, i.e.,

\[ (kr - 1)e^{-kr} = \frac{d}{1 - d}e^{kr}. \quad (6.13) \]

Define another function \( g(\cdot) \) by

\[ g(u) = (u - 1)e^{-u}, \text{ where } u = kr, \quad (6.14) \]

and thus its first derivative is

\[ g'(u) = (2 - u)e^{-u}. \quad (6.15) \]

Since \( g'(2) = 0 \), \( \forall u > 2 \) \( g'(u) < 0 \), and \( \forall u > 2 \) \( g'(u) < 0 \), it follows that

\[ \max_{u \in \mathbb{R}} g(u) = g(2) = e^{-2}. \quad (6.16) \]

That is to say, if

\[ \frac{d}{1 - d} > e^{-2}, \text{ i.e. } d > \frac{1}{1 + e^{2-k\tau_r}}, \quad (6.17) \]

then \( g(u) < \frac{d}{1 - d} \forall u \in \mathbb{R} \), and thus \( f'(r) > 0 \forall r \in \mathbb{R} \). Since \( f(\tau_r) = \tau_r \), so it follows that \( \forall r \geq \tau_r, f(r) \geq \tau_r \). That is to say, once \( r_{t-1} \geq \tau_r \), we have \( r_t \geq \tau_r \).
In the analysis, I choose $\tau_r = 1$, $k = 1$ and $d = 0.99$ under $M_{0t}$, so the gamma shape parameter $r_t$ under $M_{0t}$ is always larger than or equal to 1 after the first non-zero observation or being thresholded for the first time. Therefore, at most one time of thresholding is needed under $M_{t0}$. While doing monitoring for sparse flows, a flow of all zeros is of no interest, and only large observations should be considered as reflecting possible anomalies. In this case, thresholding has very limited influence on the marginal probability under $M_{0t}$. For the alternative assumption, the proposal discount factor is chosen as $d' = 0.1$, where the condition in eqn. (6.17) is no longer satisfied. In this case, $g(\cdot)$ has two nulls $u_1^* \in (1, 2)$ and $u_2^* \in (2, \infty)$, so $f'(\cdot)$ would also have two nulls $r_1^* \in (1/k, 2/k)$ and $r_2^* \in (2/k, \infty)$, and $f'(r) > 0$ for $r \in (-\infty, r_1^*) \cup (r_2^*, \infty)$, $f'(r) < 0$ for $r \in (r_1^*, r_2^*)$. Therefore, it is possible that $\delta t r_{t-1}$ falls below 1 even if $r_{t-1} \geq 1$, and it is probably necessary to do thresholding for multiple times. In the next section, I discuss how Bayes factor changes with the thresholded shape parameter under the alternative assumption.

### 6.3.2 Influence on Bayes factor

The next step is to check the difference between the Bayes factors before and after thresholding. For a very sparse flow path, it is expected that most observations will be zero and prefer $M_{0t}$, and thus $r_t \approx 1$, while a non-zero flow count at time $t$ should be considered as an outlier and lead to the Bayes factor $H_t < 1$. Now analyze the two gamma processes under both assumptions.

For all $t > 1$, as discussed before, the thresholding can only be applied on the alternative assumption, and after thresholding, the marginal distribution under $M_{1t}$ is

$$p(y_t|\mathcal{D}_{t-1}, H_1) \sim NB(k_t^*, p_t^*), \quad (6.18)$$

where $k_t^* = \max\{1, x_t + \delta t r_{t-1}\} = 1$, and $p_t^* = \delta t c_{t-1}/(\delta t c_{t-1} + m_{it})$ which is the same
as before thresholding. The Bayes factor after thresholding becomes

\[ H_t^* = \frac{p_0(x_t \mid \delta_t, \delta_{t-1}, x_{0:t-1})}{p_1^*(x_t \mid \delta_t^*, \delta_{t-1}^*, x_{0:t-1})} = \frac{\Gamma(x_t + \delta_r r_{t-1}) \delta_t \delta_{t-1}^*}{\Gamma(x_t) \Gamma(\delta_r r_{t-1})} \frac{(m_{it} + 1)^{x_t r_{t-1}}}{(m_{it} + 1)^{x_t + \delta_r r_{t-1}}} \]  

(6.19)

The ratio of the two Bayes factors is then

\[ \frac{H_t}{H_t^*} = \left( \frac{\delta_r r_{t-1} + x_t - 1}{x_t} \right)^{-1} \left( \frac{\delta_r c_{t-1}/m_{it}}{\delta_r c_{t-1}/m_{it} + 1} \right)^{1 - \delta_r r_{t-1}}. \]  

(6.20)

When monitoring a flow of all zeros, threshold \( \tau_r \) is applied to both the null and alternative assumptions, that is to say, the shape parameters of both latent gamma processes are \( \tau_r \). Intuitively, since the gamma process under \( M_{0t} \) has a higher rate parameter and thus a lower mean, a new observation of 0 still prefers \( M_{0t} \). While monitoring a large observation \( x_t > \tau_r \) which is a potential anomaly, there is no need to threshold the shape parameter under both assumptions. That is to say, monitoring with thresholding performs in consistence with the version without thresholding, and gives credible monitoring results.

6.4 Hourly Fox News Network Example with BDFM

An extreme example from the Fox News study is summarized in Figure 6.2, regarding transition flows from “Homepage” to “World” on the morning of February 23. The series shows clear departure from the steady random-walk evolution of the BDFM, and this example highlights the efficacy of the on-line monitoring and automated intervention strategy.

There are several periods in which \( l_t > 1 \) but the evidence against the normal model is not so strong as to signal an exception and call for intervention. The period around 23–25 minutes saw a substantial upswing in flows that triggered interventions.
to adapt three times. Interventions at about 32.5 and 48 minutes were triggered
by a cumulated run-length $l_t$, suggesting gradual drift from the standard model.
This picture is very similar when shown in terms of the flow frequencies $x_{ijt}/n_{i,t-1}$
or the conditional transition probabilities $\theta_{ijt}$ rather than the raw counts $x_{ijt}$ and
rates $\phi_{ijt}$. This is an example where there were (at least) two periods of change in
transition characteristics beyond those defined by the BDFM, to which monitoring
and intervention is able to adapt on-line. In practice, more can and should be done
in terms of deciding upon responses that make business sense. The analysis here
simply shows the potential, and notes that it can be applied in parallel across all
inflow and node-node transition models in a wholly automated manner.

A second example in Figure 6.3 shows a more typically stable trajectory, with
only two interventions that appropriately adapt to modest and subtle level changes
in the latter part of the time period.

6.5 Comparison with Piecewise Median

I compare the Bayesian sequential monitoring tool to a piecewise median method
in Vallis et al. (2014). Both methods are used for anomaly detection, but the piece-
wise median method is post hoc, whereas the Bayesian sequential monitoring method
signals as data arrive. Also, the Bayesian analysis can intervene immediately to im-
prove forecasts by adapting to trends. Finally, the median method has a window size
of $n_w$, and makes $0.49 \times n_w$ tests for anomalies, imperiling the false positive rate, a
problem that the Bayes factor approach sidesteps.

Nonetheless, comparison is instructive. Figure 6.5 shows the times at which
anomalies are flagged in the flow from “Homepage” to “World” on the morning of
February 23. Red circles correspond to Bayesian monitoring, and green asterisks to
piecewise median monitoring. Both methods detect anomalies near 9:25, but the
median method does not flag change at 9:32.30 or 9:48. Also, median monitoring
Figure 6.2: Summaries from analysis of transitions from node $i=$Homepage $\rightarrow j=$World with Bayesian model monitoring and discount-based intervention. Data is from the February 23rd am period. Upper: Symbol + indicates observations judged consistent with the standard model; x indicates cases identified as potential outliers by low $H_t$; * indicates those flagged as potential change points via low $L_t$; o indicates cases with $l_t > 1$. The vertical arrows indicate times of automatic intervention. The full line and shaded region represent one-step forecast means and 95% intervals. Center: Tracks of $\log(H_t)$ (above center) and $\log(L_t)$ (below center) over time. Lower: Data (+) with one-step forecast means and 95% intervals from the standard BDFM analysis in red compared to the analysis with monitoring and intervention in gray.
Figure 6.3: Summaries from analysis of transitions from node $i=$Homepage $\rightarrow j=$Science under monitoring and intervention, with details as in Figure 6.2.
found many more anomalies than Bayesian sequential monitoring, raising concern about false positives.

![Graph showing anomaly detection results](image)

**Figure 6.4:** A comparison of anomaly detection results by piecewise median monitoring (green asterisks) and sequential Bayesian monitoring (red circles) for the flow from node $i=$ Homepage $\rightarrow j=$ World on the morning of Feb. 23.

### 6.6 Summary Comments

This chapter develops sequential Bayesian model monitoring and automatic intervention for BDFMs. The proposed approach is based on Bayes factor tests of incoming flows against short-term model predictions. By considering a combination of three detection measures (multiple-step Bayes factor, cumulative Bayes factor and history length), it can not only signal data that are outliers or change points, but also allow the model to appropriately adapt to data promptly. Since all the three detection measures can be computed sequentially, and monitoring can be applied in parallel across nodes and node-pairs, the efficiency and scalability of this monitoring and intervention method is ensured.

With the hourly Fox News visiting flow data as an example, the proposed Bayesian monitoring method is able to distinguish signal from noise, and appropriate interventions also largely improve the short-term predictions. When compared to another standard outlier detection method, Piecewise Median, the proposed approach outper-
forms not only because of its higher prediction precision, but also because it enables online monitoring while the other is only suitable for offline, post-hoc, analysis.
7

Dynamic Relational Data Analysis

7.1 Introduction

Relational data is often represented in the form of binary square matrices with entries denoting the links between nodes. For example, the links can be similarities between entries in multivariate analysis, friendship or messages between users in social networks, communications in the brain in biological studies, and co-movements in financial markets, etc. Dynamic relational data has become a popular research topic thanks to its wide applications especially in social network analysis, and many related statistical models have been developed (e.g. Fu et al., 2009; Berger-Wolf and Saia, 2006; Durante and Dunson, 2014). The goal of this study is to explore novel models for dynamic relational data with fixed nodes but time-varying links. Inferences on the interactions between nodes as well as the prediction precision of individual links are both of interest in various applications.

Dynamic relational data is basically time varying binary data with certain network structures, and thus, in this chapter, I develop two different models considering i) how to model streaming binary data, and ii) how to utilize the network structure.
The first model is a binary probit model with dynamic latent thresholding ideas inspired by prior work on latent threshold models (Nakajima and West, 2013a,b; Zhou et al., 2014; Nakajima and West, 2015, 2017). The second model is a dynamic logistic regression model as a special kind of dynamic generalized linear models (West and Harrison, 1997, Chapter 14). For the network structure, models based on some form of latent variable structure can be used to induce dependence between nodes, while node-specific covariates can be incorporated. There is also an emerging literature in simultaneous graphical dynamic linear models with the decoupling/recoupling idea which encourages using contemporary network information to improve performance (Gruber and West, 2016, 2017; Zhao et al., 2016). With this in mind, this chapter includes development and implementation of both latent threshold models analyzed with Monte Carlo methods and dynamic generalized linear models solved with forward filtering and retrospective smoothing.

The dynamic relational data of a total number of \( I \) nodes is saved as adjacency matrices \( A_{1:T} \). In this study, thanks to the symmetry property of relational data, it is enough to analyze just the upper triangle of the matrix entries which are denoted as \( z_{1:T} \) after vectorization, and thus the length of this vector is \( P = I(I - 1)/2 \).

As for data applications, stock index co-movement data is of interest. Co-movements are typically studied by analyzing time-varying covariance structure of series of prices or returns. For the actual indices data pair \( x_{it} \) and \( x_{jt} \), if they either both increase or decrease, then the relational data entry can be denoted as \( z_{ijt} = 1 \), and 0 otherwise. By modeling the co-movements instead of the actual time series, it is possible to develop a more targeted modeling while limiting the concerns about model misspecification (Tsay, 2005; Durante and Dunson, 2014).

Section 7.2 and Section 7.3 introduce the latent threshold models and dynamic generalized linear models, respectively. The example of S&P 500 co-movement data is developed using both models in Section 7.4. Section 7.5 concludes the study and
discusses potential next steps.

7.2 Latent Threshold Model for Relational Data

When modeling dynamic binary data, a natural idea is to assume a continuous latent process, and then apply some form of thresholding. In one example, this applies a probit link between the observation and the latent process. That is to say, the data is modeled as whether the latent process is larger than a certain pre-chosen threshold $\tau$. For example, in this application, if two indices tend to co-move a lot, and thus there are a lot of 1’s in the data, then the latent process tends to have a higher mean, larger than the threshold, and also very likely a small variance. It should be noted that if the threshold is deducted by a certain value and all the means of latent processes are also deducted the same value, the same results arise. Therefore, in order to avoid identification issues, the threshold is set as a constant 0 in the model without loss of generality.

Here the underlying process is chosen to be a linear, Gaussian $AR(1)$ process. The model is

\[
\text{Observation : } z_{ijt} = 1_{\{y_{ijt} > 0\}} \text{ for } i = 1 : I, j = (i + 1) : I, \quad (7.1)
\]
\[
\text{System evolution : } y_t = \mu + \Phi(y_{t-1} - \mu) + \epsilon_t, \epsilon_t \sim N(0, \Psi). \quad (7.2)
\]

For small-scale data, a straightforward approach is to use Markov chain Monte Carlo (MCMC). In order to carry out an efficient fully Bayesian analysis, I put a conjugate multivariate normal prior $MVN(m_0\mu, M_0\mu)$ on $\mu$, and an inverse-Wishart prior $IW(s_0, S_0)$ on $\Psi$. As for the $AR(1)$ coefficient matrix $\Phi$, it is assumed to be diagonal, $\Phi = \text{diag}(\phi)$. Then for $p = 1, \ldots, P$, $\phi^*_p = 2c^* - 1$ is proposed and then evaluated by Metropolis-Hastings with $c^*$ as a beta-distributed random variable. Actually, this proposal can be any distribution on $(0, 1)$ of which $Beta(a,b)$ is an example, as long as this prior ensures stationarity, i.e., $|\phi_p| < 1$ (Nakajima and
West, 2013a,b). By choosing those priors, it is possible to perform M-H within Gibbs and details about the full conditionals in this algorithm can be found in appendix material in Section 7.6.

However, computational efficiency is a big issue. Every time \( t \) with a new observation, the whole MCMC has to be run again to update inference on the dynamic network. For dynamic relational data of length \( T \) over a network of \( I \) nodes and with \( M \) MCMC samples, the computational complexity of this algorithm is \( O(TI^2M) \), which is too slow for sequential analysis. Thus, the approach using dynamic generalized linear models is of interest to enable efficient sequential analysis.

7.3 Dynamic Generalized Linear Model for Relational Data

The dynamic generalized linear model (DGLM) is a classical and efficient approach for dynamic exponential family time series (West and Harrison, 1997, Chapter 14). The dynamic relational data is modeled univariately using DGLM with network structure from previous data as covariates.

For a generalized linear model, the natural parameter of a binomial probability \( \mu_t \) is its logit as

\[
\lambda_{ijt} = \logit(\mu_{ijt}) = \log \frac{\mu_{ijt}}{1 - \mu_{ijt}},
\]

which can be modeled as a linear model with latent Markovian structure. For simplicity, a random walk is assumed for the latent state \( \theta_t \) with evolution noise \( \mathbf{w}_t \), that is

\[
\lambda_{ijt} = \mathbf{F}_{ijt}^\prime \mathbf{\theta}_{ijt},
\]

\[
\mathbf{\theta}_{ijt} = \mathbf{\theta}_{ij,t-1} + \mathbf{w}_{ijt}, \quad \mathbf{w}_{ijt} \sim [\mathbf{0}, \mathbf{W}_{ijt}],
\]

where \( \mathbf{F}_{ijt} \) is a vector of known constants or covariates, \( \mathbf{W}_{ijt} \) is the evolution variance matrix, and \( \mathbf{w}_{ijt} \) has zero mean, variance \( \mathbf{W}_{ijt} \) but its distribution is otherwise unspecified.
At time $t$, the chosen covariates include all the entries sharing either the same origin $i$ or destination $j$ in the network at time $t - 1$. That is to say,

$$F_{ijt} = (1, a_{i,i+1,t-1}, \ldots, a_{i,j-1,t-1}, a_{i,j+1,t-1}, \ldots, a_{i,t-1},$$

$$a_{1,j,t-1}, \ldots, a_{i-1,j,t-1}, a_{i+1,j,t-1}, \ldots, a_{j-1,j,t-1}, a_{j,t-1})' .$$  \hspace{1cm} (7.6)

The state vector $\theta_t$ includes the logit of binomial probability and corresponding regression coefficients as

$$\theta_{ijt} = (\lambda_{ijt}, \beta_{i,i+1,t}, \ldots, \beta_{i,j-1,t}, \beta_{i,j+1,t}, \ldots, \beta_{i,t,t},$$

$$\beta_{1,j,t}, \ldots, \beta_{i-1,j,t}, \beta_{i+1,j,t}, \ldots, \beta_{j-1,j,t}, \beta_{j,t,t})' .$$  \hspace{1cm} (7.7)

It should be noted that there is no specific distributional form for the state evolution noise $w_{ijt}$; this specified by its mean $0$ and covariance matrix $W_{ijt}$. The covariance matrix is defined by a discount factor $\delta$ as

$$W_{ijt} = \frac{1 - \delta}{\delta} C_{ijt} ,$$  \hspace{1cm} (7.8)

as in West and Harrison (1997, Chapter 6).

In this specification, the binomial probability $\mu_{ijt} = \exp(\lambda_{ijt})/(1 + \exp(\lambda_{ijt}))$ characterizes how likely two nodes are linked in the network, and is thus the most important parameter. The $\beta$’s are just for tracking the influence of network structure and historical data on this probability. With this model, it is possible to perform efficient forward filtering and backward smoothing algorithm as in West and Harrison (1997, Chapter 14).

### 7.4 Example of S&P 500 Index Co-movement

#### 7.4.1 S&P 500 index co-movement data

The Standard & Poor’s 500 Index (S&P 500) is a weighted index of 500 stocks, seen as a leading indicator of U.S. equities and one of the common benchmarks for the
<table>
<thead>
<tr>
<th>Abbr.</th>
<th>Sector</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDU</td>
<td>Industrials</td>
</tr>
<tr>
<td>CONS</td>
<td>Consumer staples</td>
</tr>
<tr>
<td>FINL</td>
<td>Financials</td>
</tr>
<tr>
<td>ENRS</td>
<td>Energy</td>
</tr>
<tr>
<td>INFT</td>
<td>Information Technology</td>
</tr>
</tbody>
</table>

Table 7.1: Five chosen S&P 500 sectors for analysis.

U.S. stock market. I choose five of its sectors with names and abbreviations as listed in Table 7.1, and analyze weekly co-movement between each pair of the five sectors, especially in the year of 2008 when the financial crisis happened and hugely affected the US economy.

As in Figure 7.1, by some exploratory analysis, most sectors co-move a lot. For example, the “Industrials” sector and “Information Technology” go up/down simultaneously more than 80% of the time during 2008. However, the “Energy” sector seems not to co-move with other sectors a lot. For instance, it only co-moves with the “Financials” sector for about 40% of the time. This simple heat map offers limited information about the structure of this co-movement network, so a more detailed analysis is of interest, especially concerning dynamics over time.

### 7.4.2 Analysis by LTM with MCMC

For the latent threshold model, the MCMC is run by drawing $M = 10,000$ samples and $B = 2,000$ burn-ins. As for the priors, I put informative priors estimated from the year 2007 on $\mu$ and $\Psi$, while for $\Phi$, the prior for each individual proposal $c^*$ is chosen as $Beta(2, 2)$, which is centered at 0.5 and ensures the stability for all $\phi_1, \cdots, \phi_P$.

As in Figure 7.2, the sector pairs “Consumer staples - Information Technology”, “Industrials - Consumer staples”, “Industrials - Information Technology” and “Industrials - Financials” have higher posterior means of the latent processes $y_{ijt}$, while
“Financials - Energy” has the lowest latent values among all pairs. This result also shows interesting fluctuations in February and July-August simultaneously in several pairs, which is a strong indication for the volatile market environment in 2008.

Figure 7.3 shows the histogram of the posterior samples of auto-regressive coefficients $\phi_{1,p}$. As in this plot, the model gives confident estimates for the auto-regressive coefficients for most categories, as most of them have highly centered distributions; however, the coefficient $\phi_{ij}$ of pair “Consumer staples-Information Technology” is quite diffuse compared with others. Moreover, the auto-regressive coefficient posteriors of most sector pairs are centered around 0, while that of pair “Industrials-Information Technology” is significantly positive.

Figure 7.4 shows the heat map of posterior mean of correlations from $\Psi$ between each pair of sectors. They are clearly classified into two clusters: those involving “Information Technology” and those not, which can be explained by the inherent nature of sector classification.

### 7.4.3 Analysis by DGLM with FFBS

DGLMs provide a flexible structure for efficient forward filtering and backward smoothing. The analysis is initialized with a neutral probability with mean 0.5 and a relatively tight covariance structure. The discount factor $\delta$ is chosen as 0.95 which ensures stability of the analysis but also allows for a certain level of flexibility to adapt to the data.

The sequential analysis of the binomial probability shows how it adapts to data, while the retrospective analysis is a smoothed version of the result utilizing all the data, as in Figures 7.5 and 7.6. For example, for the sector pair “Industrials-Financials”, in the forward filtering step, the binomial probability of co-movement is drifting up since most of the observations indicate strong co-movement pattern between them, while we can also see that it goes down a little bit when there are
Figure 7.1: Frequencies of S&P 500 Index co-movement between five sectors in 01/02/2008 - 12/31/2008. Due to the symmetry property of relational data, I am only modeling the upper triangle of adjacency matrix, and thus set the diagonal and lower triangle entries as 0.

Figure 7.2: Posterior mean by MCMC for LTM of latent processes $y_{1:T}$ in 01/02/2008 - 12/31/2008 with fluctuations in February and July-August.
Figure 7.3: Posterior samples by MCMC for LTM of auto-regressive coefficients $\phi_{1:p}$ with dashed vertical red line as posterior mean.

Figure 7.4: Posterior mean by MCMC for LTM of correlations between sector pairs.
weeks when “Industrials” and “Financials” go opposite in June to September. In the backward smoothing result, such inconsistencies are removed and we see a smoothed trajectory of the binomial probability at around 0.7.

7.5 Summary Comments

In this section, dynamic relational data is analyzed from two perspectives: different modeling strategies for streaming binary data, and the utilization of network structure. Both latent threshold models and dynamic generalized linear models are analyzed and implemented with an application of stock index co-movement data. By not directly modeling the actual time series of stock indices, this enables targeted modeling while limiting the concerns about model misspecification, and also provides insights into dynamics and structure in the financial market.

The analyses provide insights into the time-varying covariance structure between nodes as well as the dynamic changes of individual links. However, computational efficiency is a big concern for LTMs. Due to the complex modeling structure of LTM, it can only be solved by MCMC which is neither scalable for large networks nor appropriate for online analysis. An approximate FFBS algorithm for LTMs would be of interest in future work. Moreover, as inspired by Gruber and West (2016, 2017) and Zhao et al. (2016), simultaneous graphical dynamic linear models provide ideal structures for incorporating contemporary network information into modeling, and future research to adapt these models to dynamic relational data is suggested as a very likely relevant direction.

7.6 Appendix: MCMC Algorithm for LTMs

This appendix summarizes the full conditionals/MH used in the MCMC algorithm for latent thresholding model with an underlying linear, Gaussian AR(1) process. The model is
Figure 7.5: Sequential analysis by DGLM of binomial probability $\mu_{1:T}$ with 95% credible interval in 01/02/2008 - 12/31/2008. The ‘*’s are real data.

Figure 7.6: Retrospective analysis by DGLM of binomial probability $\mu_{1:T}$ with 95% credible interval in 01/02/2008 - 12/31/2008. Notations are the same as in Figure 7.5 but now for retrospective results.
Observation: \( z_{ijt} = 1_{\{y_{ijt} > 0\}} \) for \( i = 1 : I, j = (i + 1) : I \),

System evolution: \( y_t = \mu + \Phi(y_{t-1} - \mu) + \epsilon_t, \epsilon_t \sim N(0, \Psi) \).

With chosen conjugate multivariate normal prior \( MVN(m_{0\mu}, M_{0\mu}) \) on \( \mu \), and inverse-Wishart prior \( IW(s_0, S_0) \) on \( \Psi \), starting from any set of initial values

\[
\{\mu^{(0)}, \Phi^{(0)}, \Psi^{(0)}, y_{1:T}^{(0)}\},
\]

run the MCMC algorithm using the following full conditionals/MH.

- Sample \( y_t | \mu, \Phi, \Psi, D_T \setminus y_t \).
  - Sample \( y_1 | \mu, \Phi, \Psi, D_T \setminus y_1 \sim MVN(m_1, M_1) \) with
    \[
    M_1 = (\Phi^{-1} \Psi^{-1} (I - \Phi)^T + \Phi \Psi^{-1} \Phi)^{-1}; \tag{7.9}
    \]
    \[
    m_1 = M_1 \Phi \Psi^{-1} (y_2 - \mu). \tag{7.10}
    \]
  - For \( 2 \leq t \leq T - 1 \), sample \( y_t | \mu, \Phi, \Psi, D_T \setminus y_t \sim MVN(m_t, M_t) \) with
    \[
    M_t = (\Psi^{-1} + \Phi \Psi^{-1} \Phi^T)^{-1}; \tag{7.11}
    \]
    \[
    m_t = M_t \Phi \Psi^{-1} (y_{t-1} + y_{t+1} - 2\mu). \tag{7.12}
    \]
  - Sample \( y_T | \mu, \Phi, \Psi, D_T \setminus y_T \sim MVN(m_T, M_T) \) with
    \[
    M_T = \Psi; \tag{7.13}
    \]
    \[
    m_T = \mu + \Phi(y_{T-1} - \mu). \tag{7.14}
    \]
- Sample \( \mu | \Phi, \Psi, D_T \sim MVN(m_\mu, M_\mu) \) with
  \[
  M_\mu = (M_{0\mu}^{-1} + T(I - \Phi) \Psi^{-1} (I - \Phi)^T)^{-1}; \tag{7.15}
  \]
  \[
  m_\mu = M_\mu (M_{0\mu}^{-1} m_{0\mu} + ((I - \Phi)^{-1})^T \Psi (I - \Phi)^{-1} y_1 + \sum_{t=2:T} (I - \Phi)^{-1} (y_t - \mu - \Phi(y_{t-1} - \mu))). \tag{7.16}
  \]
• Sample $\Phi|\mu, \Psi, D_T$. For $i = 1, \cdots, P$, sample $c^* \sim Beta(a_0, b_0)$, $\phi_i^* = 2c^* - 1$, $\Phi^* = \text{diag}(\phi_1^*, \cdots, \phi_P^*)'$ and accept with probability

$$
\alpha(\phi, \phi^*) = \min\{1, \prod_{t=2:T} \frac{N(y_t|\mu + \Phi^*(y_{t-1} - \mu), \Psi)}{N(y_t|\mu + \Phi(y_{t-1} - \mu), \Psi)} \}. \quad (7.18)
$$

• Sample $\Psi|\mu, \Phi, D_T \sim IW(s_n, S_n)$ with

$$s_n = s_0 + T; \quad (7.19)$$

$$S_n = S_0 + ((I - \Phi)(y_1 - \mu))((I - \Phi)(y_1 - \mu))^T$$

$$+ \sum_{t=2:T} (y_t - \mu - \Phi(y_{t-1} - \mu)(y_t - \mu - \Phi(y_{t-1} - \mu))^T. \quad (7.21)$$
In studies of dynamic network flow count data, both BDFMs and LLGMs adapt to time-varying flows within networks and can coherently quantify non-stationary changes in network flow rate processes. This work extends and customizes non-stationary process models for count data flows in a network. Novelties include the use of occupancy factors to appropriately scale Poisson rates in sets of decoupled models, and the introduction of discount factor scheduling to address problems with, in particular, low flow rates. Sequential analysis of the resulting BDFM is fast, scalable, and efficient. Compared with BDFMs, LLGMs are able to adapt more quickly to trend change in data, and thus has better performance when there are more variations in flows. The analysis of the Fox News network time series data shows the value of both model in making initial inferences on flow rate processes, in discovering differences across days, and in flagging potential advertising (or content provision) opportunities.

The emulation map from the BDFM/LLGM to the DGM is an important new strategy to achieve efficient computation and scalability. I fit a flexible, adaptive model in a set of decoupled analyses, and then directly map posterior samples to
the more substantively interesting parameter processes in a model that is otherwise challenging. With the Fox News flow data, I show some (of many) examples of how this isolates dynamics in node-specific and interaction effects. In some such cases, there is “time-varying sparsity” in node-node interaction effects over time, identified by Bayesian credible values; i.e., some interaction effects are sometimes present, sometimes not.

Computational demands for the full analysis scale as $T(I + 1)^2M$ where $M$ is the Monte Carlo sample size. The modeling and emulation strategy makes analysis fast. For example, a 2016 Matlab implementation of BDFM and its DGM mapping running on a standard laptop (2.3GHz cpu, 16Gb memory) took less than 5 minutes, with $T = 110, I = 22$ and $M = 5,000$, while with proper edge cutting, a LLGM implementation on the same computer for a network with $T = 284, I = 237$ and $M = 5,000$ takes about 1 hour.

Also, there is novel methodology that allows formal sequential Bayesian model monitoring based on Bayes factor tests. This is accompanied by automated intervention to allow models to adapt at times of change in the flow parameter processes. The decoupled/recoupled models for within-network transitions allow development of efficient sequential Bayesian testing based on Bayesian factors related to short-term prediction of each node-node flow (in parallel). At every time point, the decoupled models are monitored, and significant departures from predictions may be linked across node pairs to find dependencies. For example, significant increase in flow to “Entertainment” may come via increased transitions from “Homepage” alone, or represent an “Entertainment” effect seen in flows from other nodes as well. Change detection in the decoupled BDFMs allows intervention to modify one or more of the decoupled posteriors for the $\phi_{ijt}$, but the mapping to the DGM parameters then quantifies its relevance for node and interaction effects.

The BDFM, LLGM and DGM have many other applications. The ability to
rapidly signal potential anomalies in network traffic in real-time arises beyond computational advertising and the traditional origin-destination studies in transportation science (e.g. Tebaldi and West, 1998) and physical traffic (e.g. Tebaldi et al., 2002; Queen and Albers, 2009; Anacleto et al., 2013a,b). It is also relevant to cybersecurity, neural spike train experiments, count data in social networks (e.g., grooming activity among baboons, telephone calls among friends), and flows between financial institutions. The work provides new Bayesian dynamic modeling methodology in an application whose structure has general application.

A new question that grows from the current study is how to incorporate covariate information to help improve inferences about change and structure in network flows. Often there are useful covariates that should inform the analysis. As a special kind of dynamic generalized linear model, LLGM provides a flexible structure for incorporating time-varying covariate information. A next step is to incorporate such information on nodes to improve interpretability and decision making, as was exemplified with some vignettes from the Fox News study (e.g., the probable impact of the Academy Awards on flow to “Entertainment”). Dynamic traffic flow with geographical information and brain network data with cell connection information would also be appropriate applications.

As for dynamic relational data, both the LTM and DGLM are able to provide insights about the time-varying covariance structure between nodes as well as the dynamic change of individual links, while there are still two issues that should be solved in future studies. The first one is computational efficiency. Due to the complex structure of LTMs, it can only be solved by MCMC. Though it is trivial to do straightforward inference using MCMC posterior samples, the computational complexity of this algorithm is \( O(TI^2M) \) for a network data of \( I \) nodes and length \( T \) using \( M \) MCMC samples, which is very inefficient and not ideal for sequential analysis where the MCMC must be rerun at each time point as each new observation...
arises. An approximate FFBS algorithm for LTM would be of interest in future work. The other issue is that more network structure information could be incorporated into modeling. As developed in Gruber and West (2016, 2017) and Zhao et al. (2016), simultaneous graphical dynamic linear models (SGDLMs) provide ideal structure for incorporating contemporary network information, and some extension of SGDLMs to relational data and dynamic networks should be expected in future.
Bibliography


140


Biography

Xi Chen was born on August 12, 1990 in Qingdao, China. In July 2013, she received her B.S. in Mathematics & Applied Mathematics with Honors and B.A. in English Language & Literature from Chu Kochen Honors College, Zhejiang University. She then moved to Durham, NC to pursue doctoral studies in the Department of Statistical Science at Duke University under the supervision of Professor Mike West. She graduated with her Ph.D. in Statistical Science in May 2017.