Network Science **7** (3): 292–318, 2019
doi:10.1017/nws.2019.10

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

Bayesian dynamic modeling and monitoring of network flows

Xi Chen[1*], David Banks[2] and Mike West[2]

[1]LinkedIn Corporation, Sunnyvale, CA 94085, USA and [2]Department of Statistical Science, Duke University, Durham, NC 27708-0251, USA (e-mails: David.Banks@duke.edu, Mike.West@duke.edu)
[*]Corresponding author. Email: chenxi199008@gmail.com

Action Editor: Stanley Wasserman

Abstract

In the context of a motivating study of dynamic network flow data on a large-scale e-commerce website, we develop Bayesian models for online/sequential analysis for monitoring and adapting to changes reflected in node–node traffic. For large-scale networks, we customize core Bayesian time series analysis methods using dynamic generalized linear models (DGLMs). These are integrated into the context of multivariate networks using the concept of decouple/recouple that was recently introduced in multivariate time series. This method enables flexible dynamic modeling of flows on large-scale networks and exploitation of partial parallelization of analysis while maintaining coherence with an over-arching multivariate dynamic flow model. This approach is anchored in a case study on Internet data, with flows of visitors to a commercial news website defining a long time series of node–node counts on over 56,000 node pairs. Central questions include characterizing inherent stochasticity in traffic patterns, understanding node–node interactions, adapting to dynamic changes in flows and allowing for sensitive monitoring to flag anomalies. The methodology of dynamic network DGLMs applies to many dynamic network flow studies.

Keywords: Bayesian model emulation, decouple/recouple, dynamic network flow time series, dynamic generalized linear models, Internet traffic, parallel computing, state-space models

1. Introduction

Key areas of network science have successfully adopted and refined statistical modeling ideas and methods from other fields, and stimulated new statistical developments of broader use. The general area of dynamic networks has generated broad interest in questions of modeling changes in time of network (and graph) structures from statistics and machine learning perspectives; we comment on a number of recent contributions (including Hanneke et al., 2010; Richard et al., 2014; Newman, 2004, 2018; Holme & Saramäki, 2013; Holme, 2015; Giraitis et al., 2016; Bianchi et al., 2018), and relevant Bayesian approaches (including Kim et al., 2018; Sarkar et al., 2007; Xing et al., 2010; Xu & Hero, 2014) in Section 2.1.

Our interest is in studies of time-varying patterns of integer counts of traffic flowing between nodes in a given network (as well as into and out of the network). This topic has broad application but has not previously been well-addressed in statistical terms. Past work on so-called network tomography (e.g. Tebaldi & West, 1998) and physical traffic flow forecasting (e.g. Tebaldi et al., 2002; Queen & Albers, 2009; Anacleto et al., 2013a,b) are relevant, but our dynamic network flow contexts present challenges that require new modeling approaches. We exploit perspectives of Bayesian time series modeling and analysis to advance practicable methodology for characterizing and monitoring dynamic network flows. The motivating context is in Internet traffic in

e-commerce web-sites, where we present aspects of a case study based on the new modeling and analysis framework.

Increasing access to streaming traffic data on networks drives interest in models to quantify inherent levels of variation in flows of traffic between network nodes and into/out of the network from other sources. There are two aspects of this. The first is natural and unpredictable variation, the second is that of stochastic but sustained patterns in underlying trends over time. State-space models directly address this: they couple observation (noise) models for unpredictable variation with latent state processes representing the structure and relationships among nodes—and patterns of time-variation in these relationships. Understanding and appropriately characterizing normal patterns of variation in traffic flows is necessary to coherently address questions of monitoring flows for potential anomalies, and then to intervene or otherwise respond to interesting inferred changes.

Key technical challenges are to develop real-time/sequential analysis that is computationally efficient and scalable with network size and data sampling rates. We address these questions from a new perspective of Bayesian time series analysis, introducing to the network science literature approaches based on dynamic generalized linear models (DGLMs) of proven utility in other fields for many years (e.g. Migon & Harrison, 1985; West et al., 1985; West & Harrison, 1997, Chapter 14). The class of DGLMs integrates non-Gaussian sampling models of traditional generalized linear models with state-space evolution models for time series. The subclass of models based on conditional Poisson sampling distributions is especially relevant to dynamic network flow studies as it to other problems involving monitoring and forecasting multivariate systems of time series of counts (e.g. Berry & West, 2019; Berry et al., 2019).

We extend basic univariate DGLMs to the multivariate dynamic system generated by large-scale networks. We do this using the modeling concept of *decouple/recouple* originally introduced for multivariate time series in financial and economic applications (Gruber & West, 2016, 2017), and that has been recently extended for network data with simple network flow models (Chen et al., 2018). The latter reference represents our starting point here. The decouple/recouple idea is combined with the use of Bayesian model emulation, which maps inferences from collated sets of individual node–node flow models to an integrated multivariate network. This model enables us to explore dynamics in network-level activity intensity, node-specific flow dynamics, and node–node interactions over time with information from individual flow levels.

This new model class is quite general in admitting a range of possible state-space models for link-specific flow evolutions. In our case study, we utilize one of the most important special cases of (local) linear growth models for adapting to unpredictable changes in trends underlying flow patterns. Other applications will involve special cases customized to context—such as dynamic regressions on known predictors, seasonal structures, and other Bayesian state-space structures used in time series analysis (e.g. West & Harrison, 1997; Prado & West, 2010).

Our case study concerns Internet traffic in e-commerce, where the flow data are counts of visitors moving between nodes that are clusters of web pages corresponding to meaningful categories in a commercial website. The network example has over 56,000 node pairs and data for 288 time points within one day. Online advertisers are interested in many statistical issues related to traffic flow and site-segment content. The field has become quite sophisticated, and many methods have been employed to explore related problems, for instance, complex recommender systems (Koren et al., 2009), sentiment analysis (Pang & Lee, 2008), and text mining (Soriano et al., 2013), etc. However, basic questions of statistical modeling to characterize, monitor, and potentially understand dynamics in traffic across site-segments have not received as much attention as they warrant. In particular, there is potential commercial value as well as inherently interesting methodological concern in identifying and adapting to fluctuations in the changes of a site's popularity on short time scales. There is related interest in understanding the interactions between sites with respect to traffic, and in model-based monitoring for subtle anomaly detection in sub-networks. In such applications, one important focus is "ads pacing" which relates to the

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

speed with which advertisers deplete the budget assigned to a specific ad. A good understanding of traffic dynamics can help control the delivery of ads and budget spending, and thus improve the matching between demand-side and supply-side to maximize revenue.

We extend the recent work of Chen et al. (2018) in terms of network context and goals. We introduce the rich class of DGLMs for latent Poisson rates, with opportunities to substantially advance methodology for time-varying flow characteristics on larger networks. In addition to extending the statistical modeling methodology, analysis of historical patterns in flow rates is enabled with extensions of existing Bayesian time series analysis with DGLMs to include retrospective posterior sampling of state vectors over time. This underlies the ability to map to the time-varying parameters of so-called dynamic gravity models (DGMs) for inference on global network flow rates, dynamics in node-specific rates, and dynamics in node–node relationships. The application in this paper involves a large network that highlights the utility of the new models, along with the technical advances in modeling and computation for larger problems.

Section 3 describes the network context, basic statistical setting and notation. It outlines the concept of decoupling multivariate flows into those between univariate node–node pairs, as well as dynamic parameter mappings for inference on node-specific and node–node interactions. Section 4 details the class of conditionally Poisson DGLMs in a general setting, and then focuses on the specific example of local linear growth models (LLGMs) for latent flow rates underlying network counts. Linked to these models, the Appendix has three components. Appendix A summarizes technical details for mapping of DGLM inferences to the context of DGMs. Appendix B summarizes the standard online learning algorithm for DGLMs and a novel methodological extension required for approximate posterior simulation of full-time trajectories of latent state vectors. Appendix C provides further examples of inferred dynamics in node–node interaction effects, illustrating potential insights for online advertising.

Our case study involves data from the Fox News website, where flows represent individuals browsing web pages. Section 5 introduces the data and context, develops initial DGLM analysis and discusses some comparisons with the prior approach in Chen et al. (2018). Section 6 discusses the mapping of posterior distributions from the DGLM analysis results to the time-varying parameters of highly structured DGMs, the latter then defining inferences on node-specific and node–node interaction effects and how they vary over time. Several highlights are discussed using specific nodes from the news website. We give some concluding comments in Section 7.

2. Background

2.1 Related work

Dynamic network analysis is an increasingly active research area and includes a multitude of network contexts and analysis goals. In statistics, for example, Hanneke et al. (2010) proposed a family of statistical models extending traditional exponential random graph models (ERGMs) to a dynamic setting, with interest in potential problems of testing for changes as well as classification. Statistical and machine learning approaches have evolved to address the problem of link prediction in time-evolving graphs (e.g. Richard et al., 2014). Perspectives from different areas of the physical sciences have included approaches based on graph theory (e.g. Newman, 2004, 2018; Holme & Saramäki, 2013; Holme, 2015), while dynamic network models are increasingly visible in applied studies in areas including biology (e.g. Uetz et al., 2000; Giot et al., 2003), econometrics and marketing (e.g. Giraitis et al., 2016; Bianchi et al., 2018).

Various Bayesian statistical models have also been developed for dynamic network studies with state-space models (Kim et al., 2018). State-space models represent data with coupled observation (noise, natural random variation) components and latent process (underlying state variables with stochastic but sustained patterns over time). Conditionally linear, Gaussian models are amenable to analytic computation using Kalman filtering-style analysis, and more extensive analysis using broader classes of Bayesian state-space models (e.g. West & Harrison, 1997). However, many/most

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

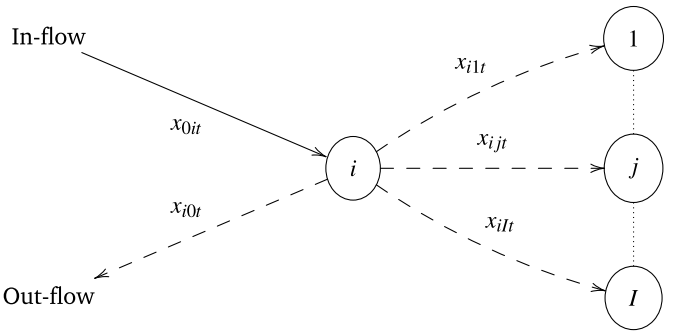

Figure 1. Network schematic and notation for flows at time t.

problems of network inference in dynamic contexts involve non-Gaussian data structures and non-linear effects, so one needs computational methods such as Markov chain Monte Carlo (Xu & Hero, 2014; Hoff, 2011). For example, Sarkar et al. (2007) map the dynamic co-occurrence data to dynamic embeddings in low-dimension Euclidean space, while Xing et al. (2010) and Xu & Hero (2014) extend well-known stochastic block models to study dynamic network tomography. More closely allied to some of our main interests here are past works on the so-called network tomography problem (e.g. Tebaldi & West, 1998) and state-space models for physical traffic flow forecasting (e.g. Tebaldi et al., 2002; Queen & Albers, 2009; Anacleto et al., 2013a,b). Our work presents a broader approach to such studies by exploiting rich classes of Bayesian state-space models to characterize and analyze complex, time-varying flows on networks.

This paper extends the recent work of Chen et al. (2018) in terms of network context and goals. Using an example of a small network (20 nodes), that work utilized a relatively simple smoothing model for latent time-varying Poisson rates on node–node pairs, and introduced the key idea of decouple/recouple to enable scaling. While based on a state-space approach, the simple smoothing model is often unable to adequately capture trends or abrupt changes in network flows, and the model structure is not flexible enough to incorporate other node/flow features. Such challenges are addressed by DGLMs whose utility has been proven in other fields for many years (e.g. Migon & Harrison, 1985; West et al., 1985; West & Harrison, 1997, Chapter 14).

2.2 Network structure and notation

Consider a closed network defined on I nodes with counts of traffic flowing between pairs of nodes observed sequentially over equally-spaced time $t = 1, 2, \ldots$. This defines a highly interdependent, multivariate count time series. The counts represent units who individually enter the network at one of the nodes at some specific time, transit to other nodes, may stay at a node for a period of time, and may exit the network at some time point. The case study involves IP addresses (indicating individuals) browsing a website comprised of disjoint sets of web categories (e.g., Arts & Entertainment, Weather) so that these sets are the nodes. In this and other applications, an additional node indexed by 0 is needed to represent flows into, or out from, nodes in the core network. At each time point $t = 1{:}T$, let n_{it} be the number of occupants of node i, and let x_{ijt} be the flow count from node i to j, including the in-flows x_{0it} and out-flows x_{i0t}, as shown in Figure 1.

3. Statistical Structure

3.1 Dynamic Poisson and multinomial flow distributions

The natural class of dynamic models has hidden Markovian structures with latent rates defining transitions between nodes over time, the rates themselves being time-varying. We adopt

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

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

$$x_{0it}|\phi_{0it} \sim Poi(\phi_{0it}) \quad \text{and} \quad \boldsymbol{x}_{it}|n_{i,t-1}, \boldsymbol{\Theta}_t \sim Mn(n_{i,t-1}, \boldsymbol{\theta}_{it}), \tag{1}$$

where $\boldsymbol{x}_{it}$ is the vector of outflows to all I nodes from node i at time t. In the notation here, $Poi(\phi_{0it})$ denotes a Poisson distribution with mean ϕ_{0it}, and $Mn(n_{i,t-1}, \boldsymbol{\theta}_{it})$ denotes a multinomial distribution with $n_{i,t-1}$ trials and $I + 1$ cell probabilities in the vector $\boldsymbol{\theta}_{it}$. The defining parameters of these distributions are time t values of underlying latent processes: ϕ_{0jt} is the time-varying rate process governing flows into node j from the external node 0; $\boldsymbol{\theta}_{it} = (\theta_{i0t}, \theta_{i1t}, \ldots, \theta_{iIt})'$ is the vector of time-varying transition probabilities from node i to all other nodes with $j = 0$ indicating departure from the network.

Modeling flexibility and computational efficiency are key needs in large-scale dynamic network analysis. In our case study, one goal of the end-user is to optimize online advertisement placement by analyzing browsing patterns, and these decisions must be made in less than a second, demanding fast and effective analysis. The theory underlying multinomial models allows us to address this using decoupling of flow transitions to node–node pairs, enabling use of univariate DGLMs in parallel, followed by theoretical recoupling for exact inference on the sets of multinomial probabilities. In the decoupling step, network flows originating from nodes $i = 1{:}I$ within the network are implicitly conditionally independent Poisson random variables, denoted by

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

where *(i)* the ϕ_{ijt} are latent Poisson rates governing the outflows from node i to all other nodes $j = 0{:}I$ at time t, and *(ii)* the *occupancy ratio* m_{it} provides appropriate scaling of outflows from node i as its level of occupancy counts changes over time.

Decoupling allows individual flows to be updated independently, to achieve fast parallel computing per unit of observation time. One aspect of recoupling is to then directly revert to the fundamental time-varying multinomial transition probabilities of Equation (1) via $\theta_{ijt} \propto m_{it}\phi_{ijt} \propto \phi_{ijt}$ subject to summing to 1 over $j = 0{:}I$. Given any set of the ϕ_{ijt} simulated from posterior distributions, this trivial computation provides inferences on the multinomial transition probability processes.

3.2 Mapping to dynamic gravity model

The recoupling aspect is to map inferences on the ϕ_{ijt} to those of a separate DGM (e.g. West, 1994; Sen & Smith, 1995; Congdon, 2000). This mapping enables evaluation of flow patterns at various levels of the network: *(i)* overall network level, *(ii)* the level of each individual node, and *(iii)* for all pairs of nodes. The details follow Chen et al. (2018), and are summarized here, with some additional technical detail in Appendix A. A DGM represents ϕ_{ijt} as a product of four terms, at each t and for each pair of nodes, viz

$$\phi_{ijt} = \mu_t \alpha_{it} \beta_{jt} \gamma_{ijt} \tag{3}$$

for each within-network node $i = 1{:}I$ and all $j = 0{:}I$ over all $t = 1{:}T$. Here the latent rates ϕ_{ijt} are mapped to: *(i)* a network-level flow rate process μ_t; *(ii)* a multiplicative *origin effect process* α_{it} for each node i; *(iii)* a multiplicative *destination effect process* β_{jt} for each node j; and *(iv)* an *affinity process* γ_{ijt}, a dynamic interaction term representing the directional attractiveness of node j as a destination for flows from node i relative to the contributions of baseline and main effects.

Given posterior samples of the ϕ_{ijt} over nodes and time, we can directly map to the DGM parameters for more incisive inference on node-specific and node–node interactions over time. In this sense, the flexible class of decoupled/recoupled DGLMs can be used as a Bayesian emulator for inference in the DGM. Technical details of the mapping are given in Appendix A.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

4. Dynamic Generalized Linear Models (DGLMs)

DGLMs are generalized linear models (McCullough & Nelder, 1989; West, 1985) with time-varying parameters defined by state-space evolutions of regression vectors. Time series observations are conditionally drawn from a sampling model in the exponential family, and the natural parameter of the distribution is regressed on a time-evolving state-vector. These models build on dynamic linear models and are central in Bayesian work in applied time series with non-Gaussian data (West & Harrison, 1997; Prado & West, 2010). We focus here on the special example of conditionally Poisson models for both the network inflow count time series and the decoupled within-network flow series, i.e., x_{ijt} for all node pairs $i, j \in 0{:}I$ and all times $t = 1{:}T$ for which the sampling models are $x_{ijt}|\phi_{ijt} \sim Poi(m_{it}\phi_{ijt})$ with $m_{0jt} = 1$ for inflows from outside the network. DGLMs are created via state-space models for the latent rate processes ϕ_{ijt}.

Ignoring the node indices i, j for clarity, consider a single Poisson rate process λ_t. Suppose that $\lambda_t = \log(\phi_t)$ is defined via an underlying linear, state-space, Markov model in which

$$\lambda_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t \quad \text{and} \quad \boldsymbol{\theta}_t = \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \quad \text{where} \quad \boldsymbol{\omega}_t \sim [0, \boldsymbol{W}_t], \tag{4}$$

with the following elements: *(i)* $\boldsymbol{F}_t$ is a known $p \times 1$ regression vector of constants and/or known values of predictors at time t; *(ii)* $\boldsymbol{\theta}_t$ is the corresponding $p \times 1$ dynamic regression parameter vector, known as the state-vector, at t; *(iii)* $\boldsymbol{G}_t$ is a known $p \times p$ state-evolution or transition matrix; *(iv)* $\boldsymbol{\omega}_t$ is a random $p \times 1$ innovation vector representing stochastic changes to the state at time t; *(v)* the $\boldsymbol{\omega}_t$ are independent over time, and the notation indicates $\boldsymbol{\omega}_t$ is zero mean and has known variance matrix $\boldsymbol{W}_t$.

Model specification depends on context, of course, and there are widely used subclasses in which $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ take specific forms (e.g. Prado & West, 2010, Chapter 4). Some examples include DGLMs when $\boldsymbol{F}_t$ includes values of known covariates (predictors), intervention indicators, and constants representing groups and design variables, in which cases the corresponding entries in $\boldsymbol{\theta}_t$ are dynamic regression coefficients. Natural evolution models then have corresponding rows of $\boldsymbol{G}_t$ as zero but for the implied column index, so that the model indicates a random walk time evolution for those parameters. Relevant to many applications are examples where both $\boldsymbol{F}_t \equiv \boldsymbol{F}$ and $\boldsymbol{G}_t \equiv \boldsymbol{G}$ are constant with specific forms chosen to define local smoothing and interpolation, such as in models M_1 and M_2 defined by

$$M_1{:} \quad \boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

$$M_2{:} \quad \boldsymbol{F} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{G} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

In model M_1, the latent state vector $\boldsymbol{\theta}_t = (\lambda_t, \rho_t)'$ consists of the current level of the latent λ_t process and the time t change in level (the discrete gradient, or "growth") term ρ_t. This is a local linear growth model (LLGM) and one of the most widely used DGLMs both alone or as a component of more elaborate models. The model defines local linear interpolation of time-varying trends that are otherwise regarded as unpredictable, and is key to retrospective smoothing of patterns in time series. Model M_2 is a more elaborate local quadratic model in which the third element of the state vector represents time-varying changes in gradient. More complicated local smoothing can be defined by higher-order polynomial DGLMs with obvious extension (West & Harrison, 1997, Chapters 7,10). The case study of this paper adopts the class of LLGMs defined by $\boldsymbol{F}, \boldsymbol{G}$ as in model M_1 above.

Summary details—including algorithms for implementation—of Bayesian analysis of general DGLMs are given in Appendix B. This includes details of sequential learning, i.e., forward filtering to process data as it arrives and sequentially update prior-to-posterior summary information

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

for the state vectors $\boldsymbol{\theta}_t$ over time. At any time t, this enables inference on the current state and forecasts of coming data. This online analysis is most relevant to sequential learning and monitoring of flows in many applications. Then, based on an observed time series of flows over a period of time $1{:}T$, key interests are addressed by retrospective analysis that examines inferences on historical trajectories of state vectors, and any functions of them of interest. Bayesian analysis here is best addressed using simulation of posteriors over historical trajectories, and the implied posteriors for past evolution in patterns of substantively interesting parameters such as those of DGMs than can be implied. Full technical and algorithmic details of this are summarized in Appendix B. One key element of model specification is the extent and nature of time-variation in the state vector as defined by the variance matrices $\boldsymbol{W}_t$. These are specified using the standard discount factors method; see Chapter 6 of West & Harrison (1997) and Section 4.3.6 of Prado & West (2010), and the additional technical details in Appendix B.

5. LLGM Analysis of Fox News Flow Data

5.1 Fox News flow data

The case study concerns flow data recording individual visitors (in terms of IP addresses) to well-defined nodes of the Fox News website. Our sample of data here concerns traffic on September 17, 2015 (a Thursday) segmented to IP addresses linked to visitors from only the Eastern Daylight Savings time zone. The website is structured by the Adex Category, a partition derived from text mining the webpage content and widely used in online advertising for webpage analysis. There are 2,208 predefined categories, including 26 main categories and different levels of subcategories. The 26 main categories are Arts & Entertainment, Computers & Electronics, Finance, Games, Home & 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. Although there can be web page content that combines different categories, the Adex Category tool enforces a strict partition. Exploratory study found little cluster structure among subcategories sharing the same main category.

Data are aggregated to 5-minute intervals, suggested by stability of exploratory analysis results across temporal levels of data aggregation. This defines a time series with $T = 288$ time points having the structure described in Section 3. At each time interval, over the directed network with nodes classified by Adex Category, the data include counts of transitions of visitors between each pair, incoming flows from outside the Fox News website to each node, as well as the total number of people visiting each node. There are no relevant additional covariates available, so the analysis focuses wholly on temporal trends in network, node-specific and node–node interactions as evidenced through analysis of flexible DGLMs that allow and adapt to changes over time. This is done using the special case of local linear growth DGLMS, i.e., the LLGM framework.

Most people spend no more than 5 minutes on a single web page (Jansen et al., 2007). Therefore, visitors who spend more than 5 minutes in a node are deemed inactive and handled as if they have left the website. Also, user information is unavailable before and after September 17, so the inactivity rule means the first and last 5-minute intervals are eliminated from the time series, which now has length $T = 284$. Then, there are some categories with little or no traffic during the entire 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 an initial superset of 2,208 categories are left for analysis.

5.2 Some initial DGLM analysis summaries

We focus on individual flows, and evaluate performance by comparing the accuracy of the one-step-ahead predictions against predictions from the Bayesian Dynamic Flow Model (BDFM) of Chen et al. (2018). In our LLGM analysis, for each individual flow, the latent state vector $\boldsymbol{\theta}_{ijt}$ has

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

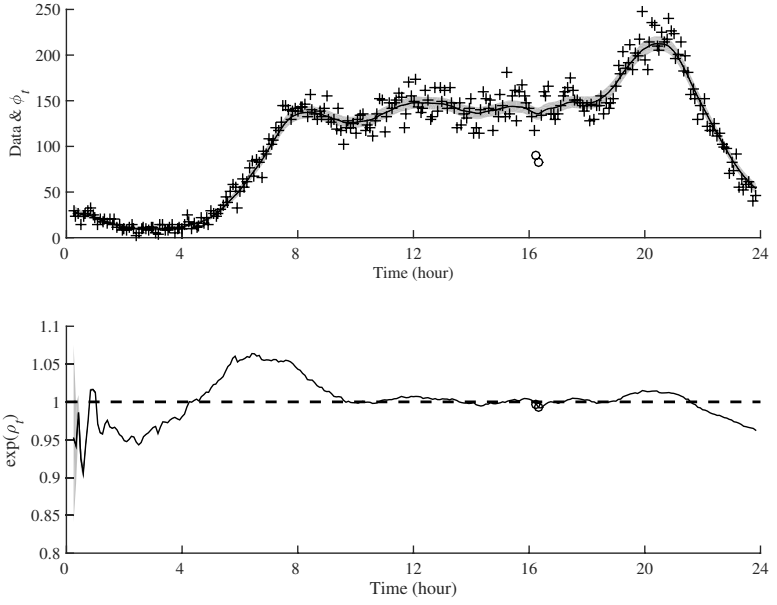

Figure 2. LLGM forward filtering of counts staying at node $i =$ Arts & Entertainment on September 17, 2015. *Upper*: data (crosses), filtered mean (solid) and 95% CI (shaded) of trajectories of ϕ_{iit}. *Lower*: filtered mean (solid) and 95% CI (shaded) of local linear growth term ρ_{iit}. Two low-count values denoted by circles are mentioned in the text.

two components: the rate parameter on the log scale λ_{ijt} and the local linear growth term ρ_{ijt}. The prior on λ_{ij0} is chosen with mean equal to the point estimate based on a simple average of data in the 5 minutes prior to the beginning of the time series, and the prior mean for ρ_{ij0} is 0. The initial prior variance matrix has zero covariances and diagonal entries 0.1. The levels of prior variance represent a relatively diffuse initial prior that allows for swift adaptation to the data over the initial few time points. The discount factor for all reported analyses is set as $\delta = 0.9$, for all node pairs. This level of discounting encourages smoothness but also allows the model to flexibly adapt to changes, and variations (that have been explored) might marginally improve local descriptions for some node pairs but are secondary to the interests and emphases here.

Decoupling allows us to apply LLGMs simultaneously to all of the flows. The LLGM analysis of flows staying at the category Arts & Entertainment is used to illustrate the model performance. Figure 2 shows the forward filtering results for both the rate parameter and the local linear growth term, while Figure 3 shows the results as smoothed by backward sampling. In general, both the sequential and retrospective analyses capture patterns of change over time well, and the efficacy of retrospective analysis is highlighted. We note that the intervals representing uncertainty about trajectories are very tight, indicating a high level of precision in estimating the underlying transition processes. This is true in several of the following graphs for other model parameters.

There are periods when volatile patterns in the data challenge model adaptability. For example, in the morning (8:00–10:00) and late at night (20:00–24:00), forward filtering analysis without any intervention underestimates the swings in the rate parameter. The sequential analysis is more sensitive to outliers. For instance, at around 16:00, there are two data points with low counts compared to the data before and after; see the circles in Figures 2 and 3. These two data points drive the rate down in the forward filtering. That said, the retrospective analysis is able to resolve these two issues by using information from the later in the series. Looking back, retrospective posterior analysis provides smoother and more accurate inference on the rate parameter.

Forward filtering and backward smoothing for the local linear growth term give insight into how trends vary during the day. For example, at first the Poisson rate for people staying in Arts &

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

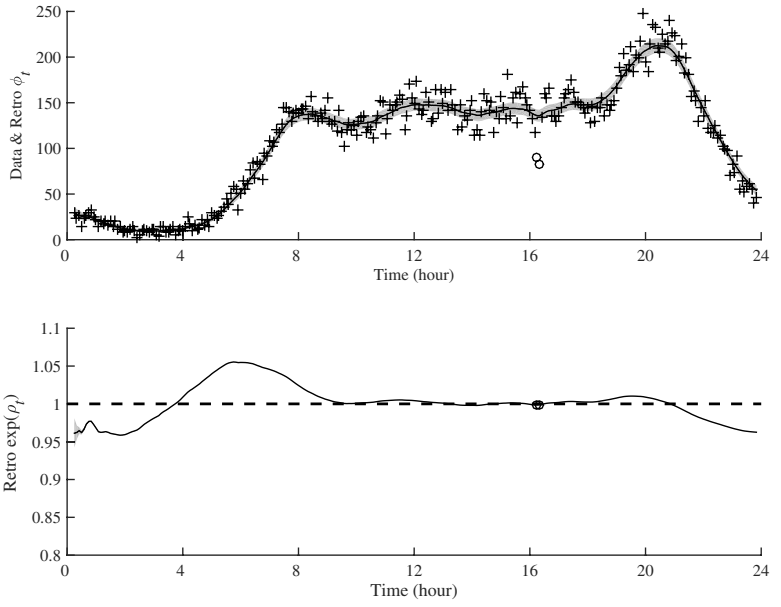

Figure 3. As in Figure 2, but now showing full retrospective analysis.

Entertainment decreases at 0:00–4:00, but its decrease slows, reaching the minimum at around 4:00. Afterwards, the rate increases rapidly and reaches the first peak of the day at around 8:00 a.m., then maintains a steady, high level until around 19:00. The rate then reaches its second peak at around 20:00, and after that, the activity level declines rapidly until midnight.

5.3 Comparison with BDFM

Some comparison of LLGM and the simple smoothing model (Chen et al., 2018), referred to as a Bayesian dynamic flow model (BDFM), was made; one example concerns counts staying at category Arts & Entertainment. The discount factor controlling adaptability is $\delta = 0.9$ for both. Summaries here focus on forward filtering and one-step ahead forecasting for comparison. Both models perform well when the trend is stable—between 8:00 and 19:00, the one-step ahead forecasts by both models agree closely with the true data. However, the BDFM tends to underestimate the rate when the trend is rising and to overestimate when the trend is declining, as during 4:00–8:00 and 20:00–24:00, respectively. In contrast, the LLGM still provides good point-wise prediction. Across nearly all flows, the LLGM outperforms the BDFM in terms of one-step ahead forecast accuracy, illustrated in Figure 4.

6. Model Mapping Analysis

6.1 Daily fox news flows example

We now recouple, mapping the retrospective results for the log rate parameters λ_{ijt} to the DGM. The results provide insight into four aspects of flow dynamics: *(i)* the baseline process μ_t that characterizes general activity intensity; *(ii)* the origin effects α_{it} that relate to activity of outgoing flows from node i; *(iii)* the destination effects β_{jt} that relate to the attractiveness for incoming flows to node j; and *(iv)* the directed pairwise affinity effects γ_{ijt}, for interactions that impact the rate functions governing flows from nodes i to j.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

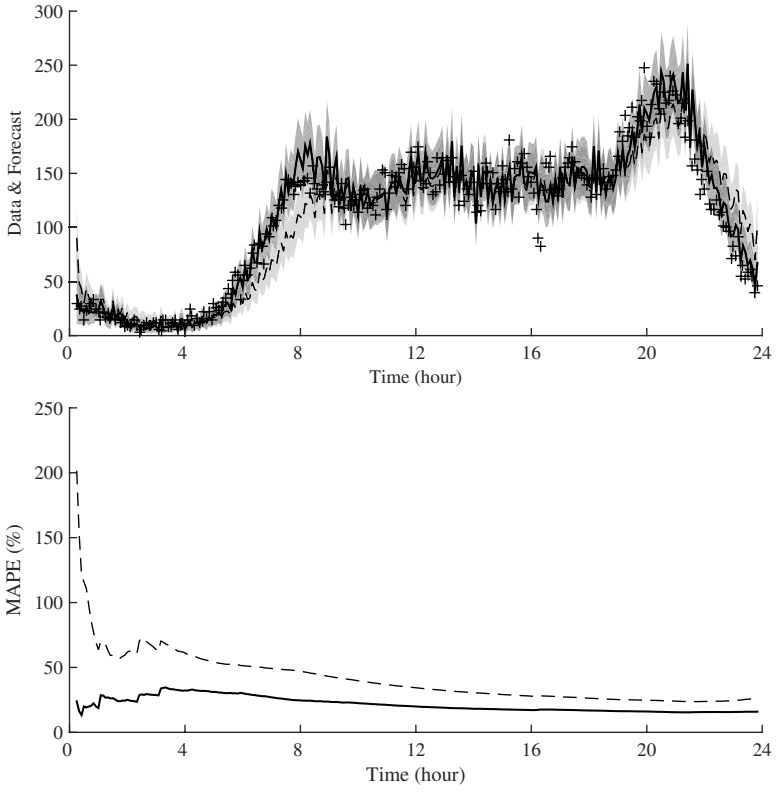

Figure 4. Summary of analyses of counts staying at node $i =$ Arts & Entertainment using both LLGM and BDFM. Data are from September 17, 2015. *Upper*: Data (+) with one-step forecast means and 95% intervals from the LLGM analysis (solid/dark gray) compared to the analysis using the standard BDFM (dashed/light grey). *Lower*: Mean absolute percentage error (MAPE) over time from the LLGM analysis (solid) compared to the analysis using the standard BDFM (dashed).

6.1.1 Baseline level

We apply the DGM decomposition to the Fox News data on the network defined by Adex Category for September 17, 2015. As expected, the baseline activity level reflects human routine—high traffic during the day and early evening, and low traffic late at night, as shown in Figure 5.

The day starts at midnight when the overall mean intensity is 1.025. As users go offline, this overall activity level decreases, reaching a minimum at around 4:00. Then the mean increases until around 8:00, when the trend becomes flat. During the day (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, which are typical times for work breaks. There is a slight decreasing trend from 16:00 to 18:00, presumably as people travel home. After dinner time, the trend increases to a peak at around 20:00, and then declines as people retire.

6.1.2 Origin and destination effects

For most categories, the origin and destination trends are similar so we focus here on the former. Trajectories of origin effects α_{it} exhibit several general patterns, varying by Adex Category. Three examples of each appear in Figure 6.

First, categories such as Arts & Entertainment, News, and Shopping show trends similar to the overall activity level μ_t. The origin effect increases to a high level during the early morning, is steady during the day and early evening, and then declines. These categories are among the most popular in the network.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

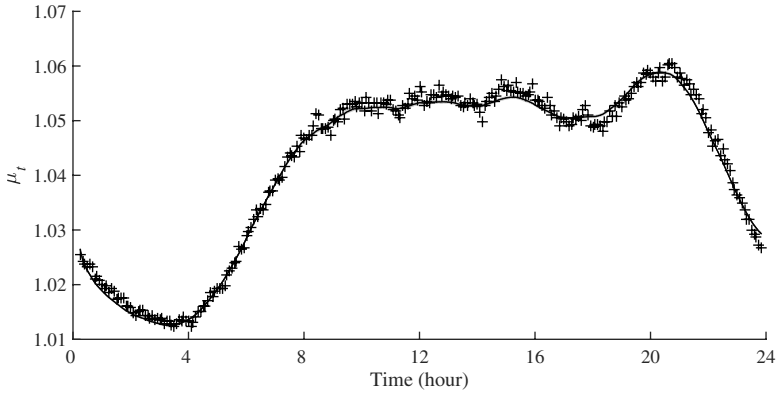

Figure 5. DGM-based inference on smoothed trajectories of the baseline activity level μ_t for Fox News data on September 17, 2015 with a 95% credible interval. The + symbols indicate empirical values from the raw data.

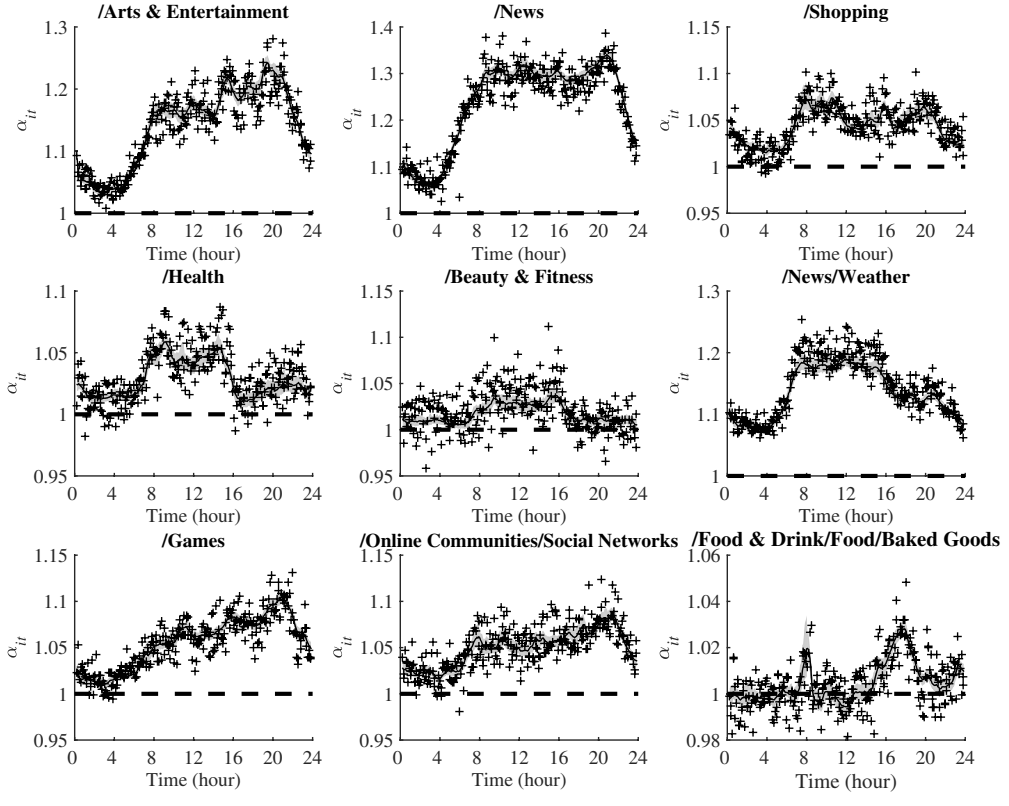

Figure 6. Smoothed trajectories of selected node-specific origin effects $\alpha_{i,1:T}$ with 95% CIs. The + symbols indicate empirical values from the raw data.

Second, categories such as Health, Beauty & Fitness, and News/Weather show high activity during the day, but drop to a low level in the evening and at night. This is reasonable for weather, since many people want to know the forecast before leaving for work. It also seems reasonable for the other categories, which are oriented towards a narrow focus that is not entertainment or relaxation.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

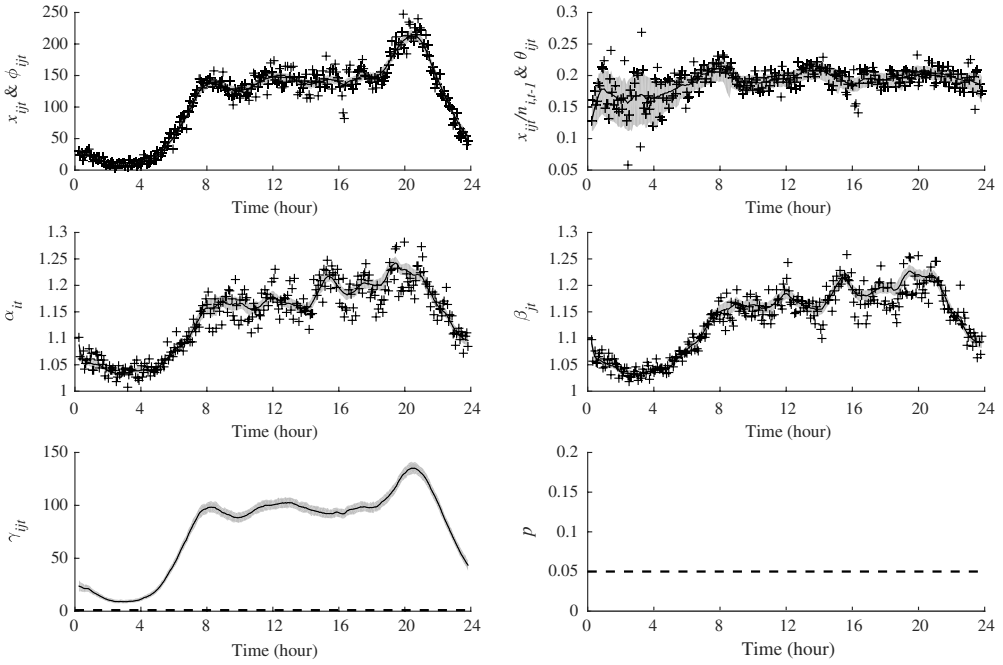

Figure 7. Posterior summaries for DGM parameters for transitions staying at node i = Finance/Investing. *Upper left:* Posterior trajectory for the ϕ_{iit} with raw counts (crosses). *Upper right:* Posterior trajectory for the θ_{iit} with raw frequencies (crosses). *Center left:* Posterior trajectory for the origin (outflow) process α_{it}. *Center right:* Posterior trajectory for the destination (inflow) process β_{it}. *Lower left:* Posterior trajectory for the self-affinity process γ_{iit}. *Lower right:* Corresponding trajectories of Bayesian credible values assessing support for γ_{iit} near 1.

Third, categories such as Games, Online Communities/Social Networks, and Food & Drink/Food/Baked Goods show a different trend in origin effect—a pattern that increases from about 4:00. to late evening, peaks at around 22:00, and exhibits local peaks at other times. Topics on games and social networks pertain to relaxation, which is a reasonable evening activity. The food category has two peaks, in the morning and the late afternoon, which may reflect people looking at recipes in the morning to plan grocery purchases, and then later when preparing dinner.

6.1.3 Affinity effects

Affinity effects capture interaction between pairs of categories. Pairs with strong affinity may indicate that people interested in one tend to be interested in the other, which would be potentially valuable in computational advertising. There are four kinds of affinity effects: *(i)* staying at a certain category γ_{iit}; *(ii)* entering the network γ_{0jt}; *(iii)* leaving the network γ_{i0t}; and *(iv)* moving between two different categories γ_{ijt}. We discuss these separately. For all four kinds of affinities, we show representative trend patterns, with more extensive examples summarized in Appendix C.

Self-affinities: A high self-affinity implies that users tend to stay at that category. A trend that shows times of day when people linger is useful to advertisers since it suggests readers have more leisure time and thus could be tempted to click on ads.

A representative trend pattern features high activity during the business day, and then lower level in the evening. Such categories include Finance/Investing and Computers & Electronics/Software (Figures 7 and 8, respectively). The self-affinity of Finance/Investing has three peaks: one around 10:00, one around 15:00, and one around 20:30. Computers & Electronics/Software is interesting since, for most categories, self-affinity drops a bit after 8:00

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

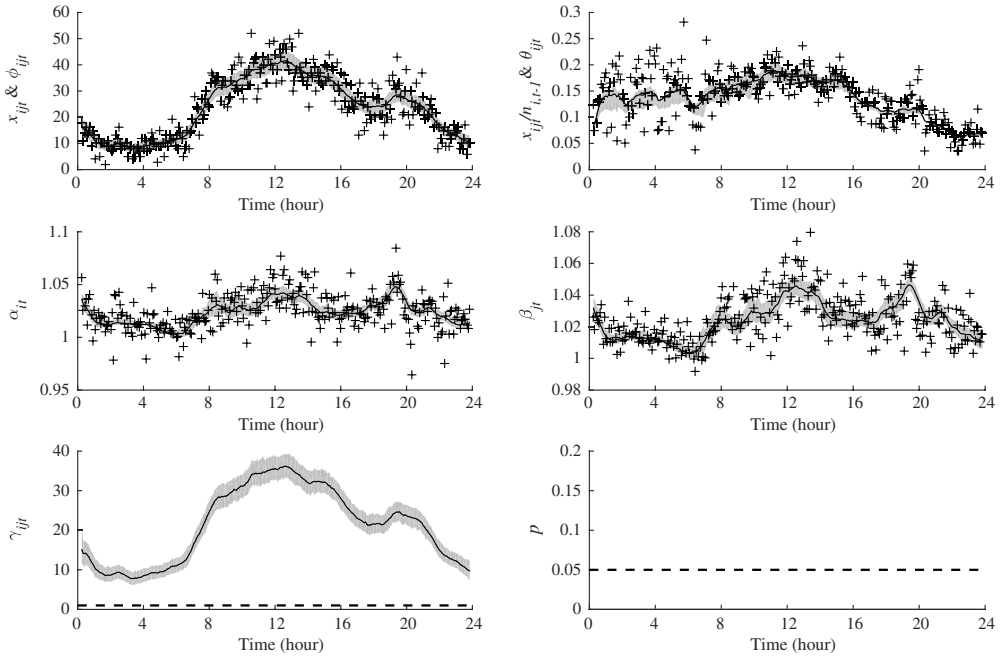

Figure 8. Posterior summaries for DGM parameters for transitions staying at node i = Computers & Electronics/Software, with details as in Figure 7.

and is not very high at noon, while the affinity of Computers & Electronics/Software increases over the morning and peaks at noon. This trend indicates people spent more time reading related contents compared with other times of the day, and should inform ad buy decisions during those peak times.

Entering affinities: A category with high entering affinity draws users from outside the Fox News website. Ads shown on such categories may be more cost-effective. The probability θ_{0it} of users entering the network at node i is a measure of a category's overall popularity.

Entering affinities show interesting patterns identified in trajectories. One such pattern peaks in the early morning; e.g., flows into News/Weather and News/Politics. The peak for both is around 6:00, as in Figures 9 and 10. For News/Weather the insight is obvious. For News/Politics, September 17 is the day after a US national political debate, which may drive extra interest. That said, this peak pattern is sustained on other weekdays, indicating that many people access news sites in the morning. Advertisers should generally avoid placing ads at this time, since people are not in shopping mode.

Exiting affinities: Categories with high exiting affinities tend to be the last stop for users navigating the Fox News website. Such categories are the last chance to show an ad, and if grocery store checkout lines are any guide, a good opportunity for impulse purchases.

One interesting trend increases over the morning, peaks at about 8:00 and then stays stable with high intensity until evening. Often, these categories are the only categories visited by a user. Examples include Arts & Entertainment and Food & Drink/Cooking & Recipes (Figures 11 and 12). Though the affinity intensity is stable, there is an interpretable bump during 16:00–18:00 for Food & Drink/Cooking & Recipes which is probably related to dinner preparation.

Distinct node pair affinities: High levels of affinity between distinct category pairs indicate interaction, which could reasonably influence advertising strategy.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

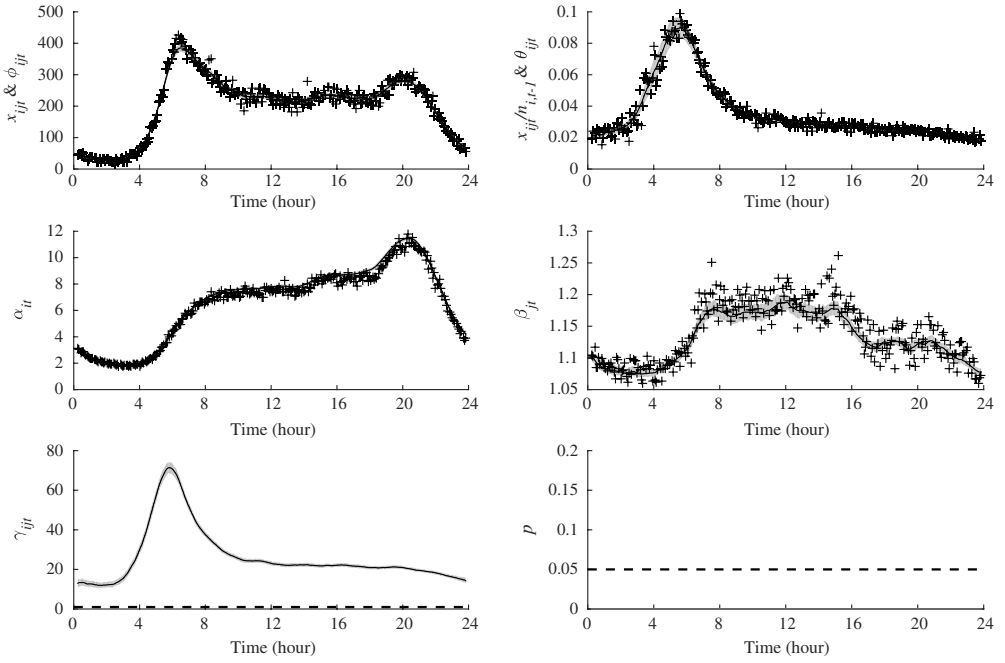

Figure 9. Posterior summaries for DGM parameters for transitions entering node $i =$ News/Weather. *Upper left:* Posterior trajectory for the ϕ_{0it} with raw counts (crosses). *Upper right:* Posterior trajectory for the θ_{0it} with raw frequencies (crosses). *Center left:* Posterior trajectory for the external origin (outflow) process α_{0t}. *Center right:* Posterior trajectory for the News/Weather destination (inflow) process β_{it}. *Lower left:* Posterior trajectory for the News/Weather entering affinity process γ_{0it}. *Lower right:* Corresponding trajectories of Bayesian credible values assessing support for γ_{0it} near 1.

In all patterns of such affinities, the one with a bump is the kind in which we are most interested. The bump indicates that users are only active in moving between those categories in a short window, while they remain inactive during the rest of the day, and thus this short window is the best time that related ads should be displayed. Examples include Online Games to Video Games and Home & Garden to Reference/General Reference/How to DIY & Expert Content.

The affinity from Online Games to Video Games (Figure 13) has a 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 strongly indicates that users who read about online games are also interested in computer and video games during this period.

The affinity from Home & Garden to Reference/General Reference/How to DIY & Expert Content (Figure 14) has a bump around 8:00–12:00. Moreover, both of 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 between them is large, which indicates strong interaction. Obviously, people plan home projects in the morning, and seek information on how to implement them.

7. Closing Comments

The case study demonstrates the value of this new class of dynamic network flow models on a large network. The specific special case of DGLMs adopted—the LLGM—is just one example of the broader class, but the analysis examples highlight the ability of this family to characterize and adapt to quite heterogeneous patterns of change over time in latent flow patterns. An important element of the analysis is that full Bayesian inference, based on computationally efficient

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

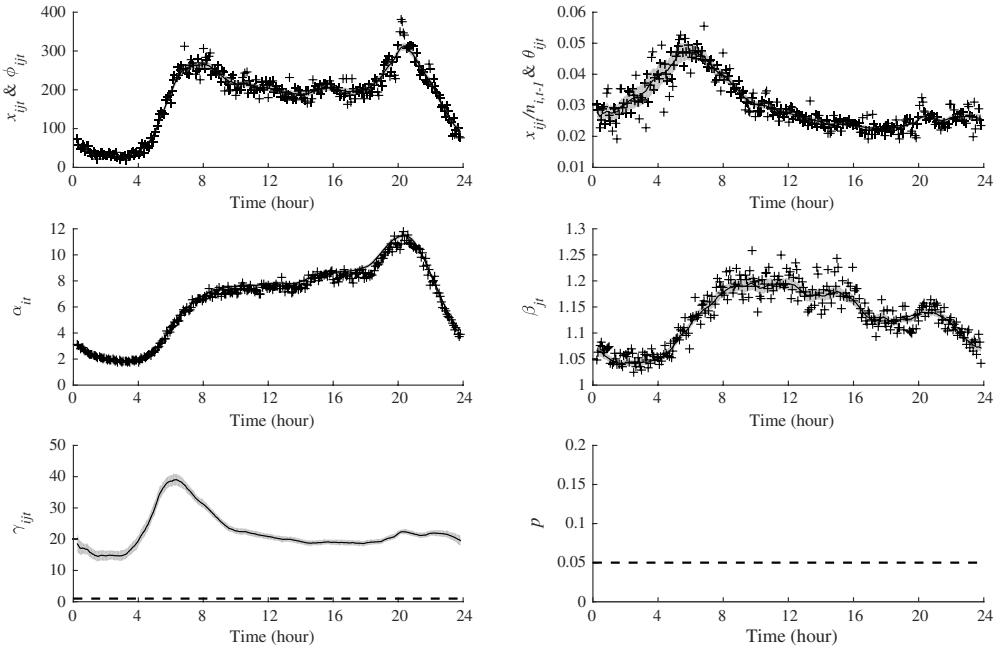

Figure 10. Posterior summaries for DGM parameters for transitions entering node $i =$ News/Politics with details as in Figure 9.

retrospective sampling analysis, defines not only point estimates of trajectories of key dynamic parameters, but also uncertainties about them from posterior samples.

A critical component of the analysis is the strategy of decoupling/recoupling. This has two aspects. First, the individual univariate DGLMs for dynamic Poisson flows are learned by forward filtering updates at each time step, and then for all within-network nodes these are theoretically recoupled to define inferences on the dynamic transition probability vectors in the inherent conditional multinomial distributions governing flows, so inferring node dependencies. The decoupling/recoupling strategy also provides an ideal structure to perform parallel computation, enabling scalable and efficient analysis for increasingly large dynamic networks. Using a computer with K cores, the computational demands for a series of length T on a network with I nodes scales only as $\mathcal{O}(I^2T/K)$. A 2018 Matlab implementation running on a standard laptop (2.3 GHz CPU, 16 Gb memory) took 52 minutes to run our analysis with $T = 286$, $I = 2{,}208$.

The second aspect of recoupling analysis is the use of the set of DGLM-based models as an emulator for a multivariate DGM that is explicitly parametrized in terms of network-wide, node-specific and node pair interaction effects, all potentially time-varying. The Fox News web study shows a number of examples of the use of this mapping to uncover—again with full Bayesian posterior uncertainty measures—interpretable patterns in time trajectories of node main effects (impacting both incoming and outgoing traffic) and node–node pair interactions (in terms of affinities of one node for another that impact traffic flows between them). We have noted the potential for such information to be exploited by online advertisers in the case study, and formal inferences will be of interest in other applications.

The DGM mapping also enables investigation of network grouping structures over time by clustering nodes using the dynamic node–node pair affinities, with potential for deeper future study. With the overall network intensity and node main effects removed by DGM, the affinity between a pair of nodes is an inherently interesting measure of the "closeness" of nodes; at one level, it can be regarded as a coherent statistical summary of node–node relationship and their

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

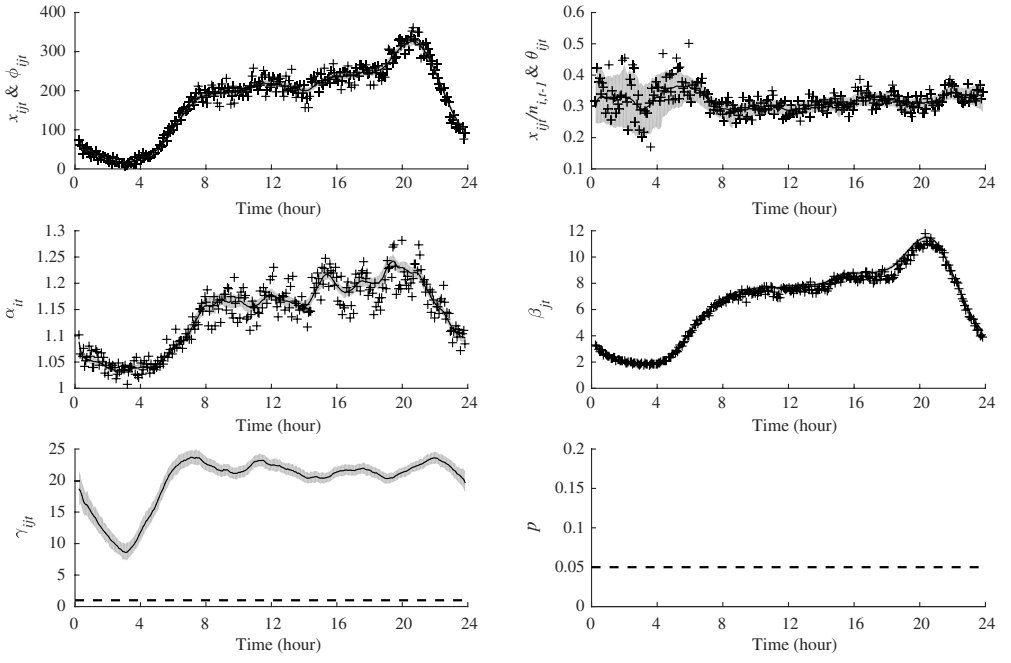

Figure 11. Posterior summaries for DGM parameters for transitions exiting from node $i =$ Arts & Entertainment. *Upper left:* Posterior trajectory for the ϕ_{i0t} with raw counts (crosses). *Upper right:* Posterior trajectory for the θ_{i0t} with raw frequencies (crosses). *Center left:* Posterior trajectory for the Arts & Entertainment origin (outflow) process α_{it}. *Center right:* Posterior trajectory for the external destination (inflow) process β_{0t}. *Lower left:* Posterior trajectory for the Arts & Entertainment leaving affinity process γ_{i0t}. *Lower right:* Corresponding trajectories of Bayesian credible values assessing support for γ_{i0t} near 1.

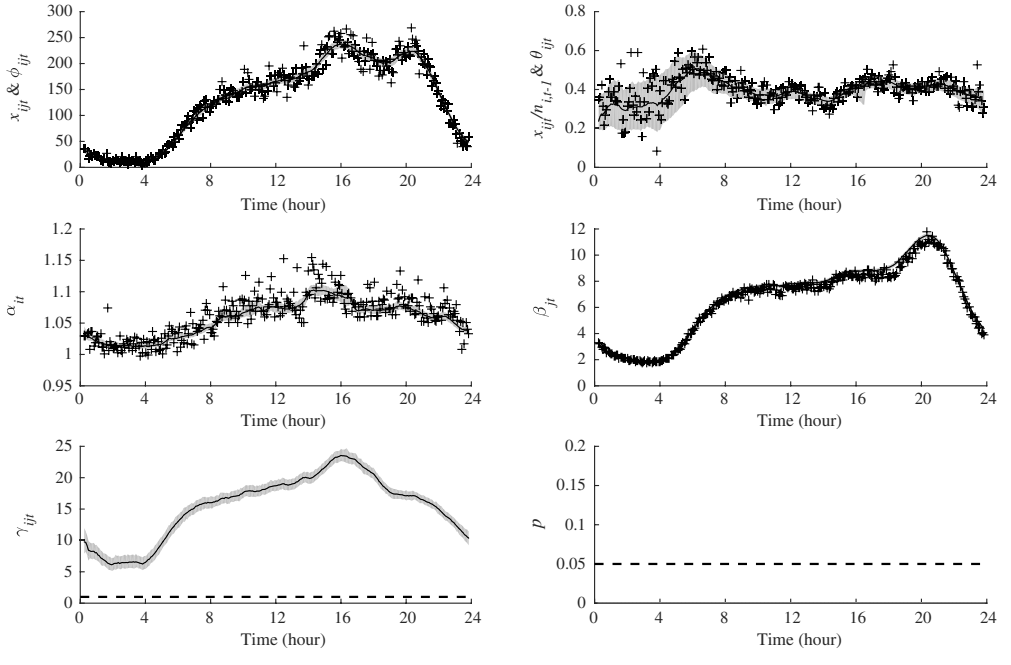

Figure 12. Posterior summaries for DGM parameters for transitions exiting from node $i =$ Food & Drink/Cooking & Recipes with details as in Figure 11.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

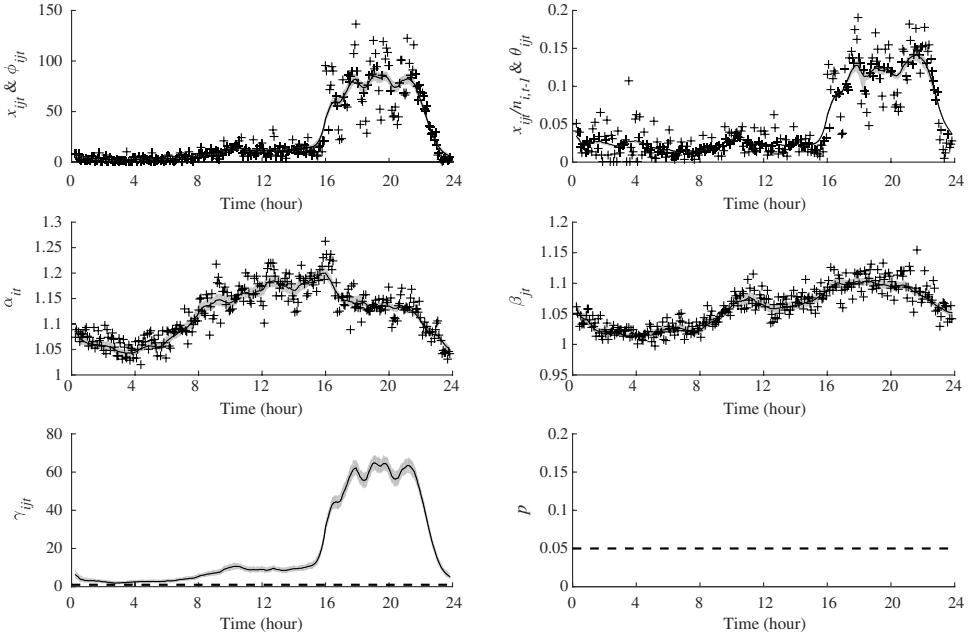

Figure 13. Posterior summaries for DGM parameters for transitions from node $i =$ Games/Online Games $\rightarrow j =$ Games/Computer & Video Games. *Upper left:* Posterior trajectory for the ϕ_{ijt} with raw counts (crosses). *Upper right:* Posterior trajectory for the θ_{ijt} with raw frequencies (crosses). *Center left:* Posterior trajectory for the Games/Online Games origin (outflow) process α_{it}. *Center right:* Posterior trajectory for the Games/Computers & Video Games destination (inflow) process β_{jt}. *Lower left:* Posterior trajectory for the Games/Online Games : Games/Computer & Video Games affinity process γ_{ijt}. *Lower right:* Corresponding trajectories of Bayesian credible values assessing support for γ_{ijt} near 1.

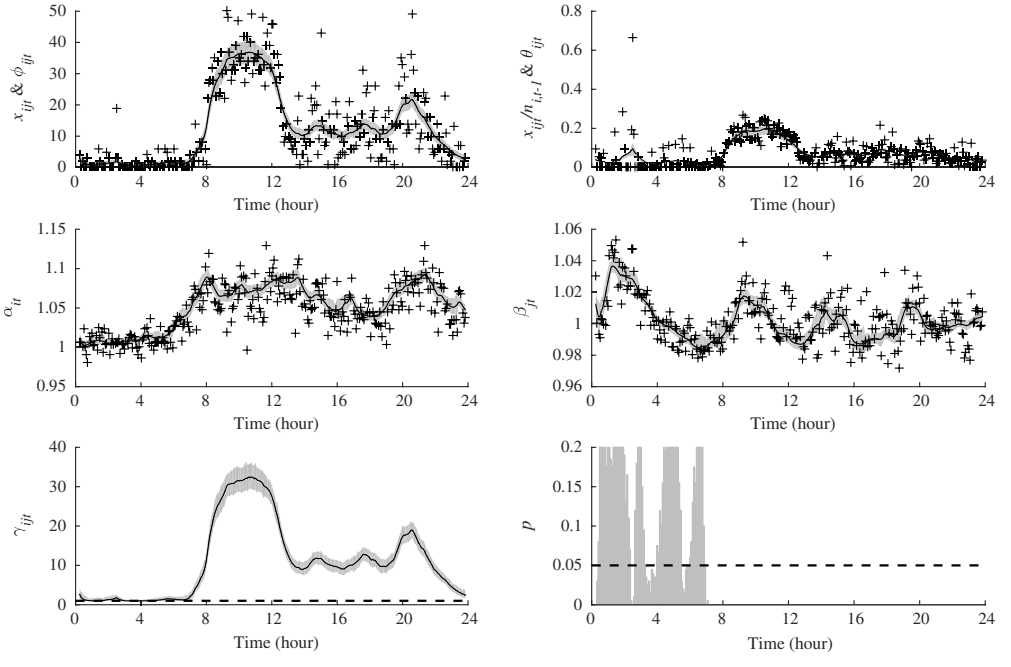

Figure 14. Posterior summaries for DGM parameters for transitions from node $i =$ Home & Garden $\rightarrow j =$ Reference/General Reference/How to DIY & Expert Content with details as in Figure 13.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

patterns over time (and much preferred to any raw data summary). We note, in addition, that the node-specific main effects for inflows can be regarded as statistical quantities of node importance or centrality; inflows and outflows together reflect the scale of importance of a node in terms of numbers and flow intensities with neighbors in the network. Traditional summary network topology measures may be of interest in other contexts, but for dynamic network traffic analysis these model-derived quantities are foremost, and their inferred trajectories over time are of primary interest.

The case study exploits and highlights the adaptability and utility of the LLGM class of models as a key driver of the overall decouple/recouple and emulation analysis. This is, of course, just one special subclass of the full class of DGLMs, and other forms are applicable in other areas. Future studies with available covariate information might include, for example, Internet traffic studies with known interventions, dynamic traffic flow with geographical and structural information, or brain network data with known or hypothesized connectivity information. Such studies can be expected to exploit more general DGLMs that include node-specific covariates or dummy variables representing intervention effects (e.g., for known ad placements in e-commerce examples, or network structure changes in others). This work offers benefits in other applications through the flexibility to customize the DGLMs to the context. The overall model framework and approach needs only customization of details in the specification of the state-space elements $\boldsymbol{F}_t$, $\boldsymbol{G}_t$ (which will in some extensions be specific to pairs of nodes as well as time-varying), as the analysis described here and illustrated in the case study applies generally to the full DGLM class.

Acknowledgments. The authors are grateful for discussions with Mark Lowe and Jewel Thomas on aspects of the research reported here. Maxpoint Interactive Inc. (now Valassis Digital) provided the network flow data of the reported case study as well as partial support for the research. We also acknowledge the constructive comments of the editor, associate editor and three anonymous referees on the original manuscript.

Conflict of interest. Authors have nothing to disclose.

References

Anacleto, O., Queen, C., & Albers, C. J. (2013a). Forecasting multivariate road traffic flows using Bayesian dynamic graphical models, splines and others traffic variables. *Australian and New Zealand Journal of Statistics*, *55*, 69–86.

Anacleto, O., Queen, C., & Albers, C. J. (2013b). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society (Series C: Applied Statistics)*, *62*, 251–270.

Berry, L. R., & West, M. (2019). Bayesian forecasting of many count-valued time series. *Journal of Business and Economic Statistics* (in press).

Berry, L. R., Helman, P., & West, M. (2019). Probabilistic forecasting of heterogeneous consumer transaction-sales time series. *International Journal of Forecasting* (in press).

Bianchi, D., Billio, M., Casarin, R., & Guidolin, M. (2018). Modeling systemic risk with Markov switching graphical SUR models. *Journal of Econometrics*, *210*, 58–74.

Chen, X., Irie, K., Banks, D., Haslinger, R., Thomas, J., & West, M. (2018). Scalable Bayesian modeling, monitoring and analysis of dynamic network flow data. *Journal of the American Statistical Association*, *113*, 519–533.

Congdon, P. (2000). A Bayesian approach to prediction using the gravity model, with an application to patient flow modeling. *Geographical Analysis*, *32*, 205–224.

Giot, L., Bader, J. S., Brouwer, C., Chaudhuri, A., Kuang, B., Li, Y., Hao, Y. L., Ooi, C. E., Godwin, B., Vitols, E. and Vijayadamodar, G. (2003). A protein interaction map of Drosophila Melanogaster. *Science*, *302*, 1727–1736.

Giraitis, L., Kapetanios, G., Wetherilt, A., & Žikeš, F. (2016). Estimating the dynamics and persistence of financial networks, with an application to the Sterling money market. *Journal of Applied Econometrics*, *31*, 58–84.

Goldstein, M. (1976). Bayesian analysis of regression problems. *Biometrika*, *63*, 51–58.

Gruber, L. F., & West, M. (2016). GPU-accelerated Bayesian learning in simultaneous graphical dynamic linear models. *Bayesian Analysis*, *11*, 125–149.

Gruber, L. F., & West, M. (2017). Bayesian forecasting and scalable multivariate volatility analysis using simultaneous graphical dynamic linear models. *Econometrics and Statistics*, *3*, 3–22.

Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics*, *4*, 585–605.

Hartigan, J. A. (1969). Linear Bayesian methods. *Journal of the Royal Statistical Society (Series B: Methodological)*, *31*, 446–454.

Hoff, P. D. (2011). Hierarchical multilinear models for multiway data. *Computational Statistics and Data Analysis*, *55*, 530–543.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

Holme, P. (2015). Modern temporal network theory: A colloquium. *European Physical Journal B*, *88*, 234.
Holme, P., & Saramäki, J. (2013). *Temporal Networks*. Springer.
Jansen, B. J., Spink, A., & Kathuria, V. (2007). How to define searching sessions on web search engines. *Pages 92–109 of:* Nasraoui, O., Spiliopoulou, M., Srivastava, J., Mobasher, B., & Masand, B. (eds), *Advances in Web Mining and Web Usage Analysis: Eighth International Workshop on Knowledge Discovery on the Web, WebKDD 2006*. Lecture Notes in Computer Science. Springer.
Kim, B., Lee, K. H., Xue, L., & Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics Surveys*, *12*, 105–135.
Koren, R., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, *8*, 30–37.
McCullough, P., & Nelder, J. A. (1989). *Generalized Linear Models*. New York: Chapman & Hall.
Migon, H. S., & Harrison, P. J. (1985). An application of non-linear Bayesian forecasting to television advertising. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian Statistics 2* (pp. 681–696). North-Holland, Amsterdam: Valencia University Press.
Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, *70*, 056131.
Newman, M. E. J. (2018). Network structure from rich but noisy data. *Nature Physics*, *14*, 542.
Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*, 1–135.
Prado, R., & West, M. (2010). *Time Series: Modeling, Computation and Inference*. Chapman & Hall/CRC Press.
Queen, C. M., & Albers, C. J. (2009). Intervention and causality: Forecasting traffic flows using a dynamic Bayesian network. *Journal of the American Statistical Association*, *104*, 669–681.
Richard, E., Gaïffas, S., & Vayatis, N. (2014). Link prediction in graphs with autoregressive features. *Journal of Machine Learning Research*, *15*, 565–593.
Sarkar, P., Siddiqi, S. M., & Gordon, G. J. (2007). A latent space approach to dynamic embedding of co-occurrence data. *Artificial Intelligence and Statistics*, 420–427.
Sen, A., & Smith, T. (1995). *Gravity Models of Spatial Interaction Behavior*. Springer.
Soriano, J., Au, T., & Banks, D. (2013). Text mining in computational advertising. *Statistical Analysis and Data Mining*, *6*, 273–285.
Tebaldi, C., & West, M. (1998). Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, *93*, 557–573.
Tebaldi, C., West, M., & Karr, A. F. (2002). Statistical analyses of freeway traffic flows. *Journal of Forecasting*, *21*, 39–68.
Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P. and Qureshi-Emili, A.. (2000). A comprehensive analysis of protein-protein interactions in saccharomyces cerevisiae. *Nature*, *403*, 623.
West, M. (1985). Generalized linear models: Scale parameters, outlier accommodation and prior distributions (with discussion). In J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith (Eds.), *Bayesian Statistics 2* (pp. 531–558). North-Holland, Amsterdam: Valencia University Press.
West, M. (1994). *Statistical inference for gravity models in transportation flow forecasting*. Discussion Paper 94-20, Duke University, and Technical Report #60, National Institute of Statistical Sciences.
West, M., & Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. 2nd edn. Springer.
West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, *80*, 73–83.
Xing, E. P., Fu, W., & Song, L. (2010). A state-space mixed membership block model for dynamic network tomography. *Annals of Applied Statistics*, *4*, 535–566.
Xu, K. S., & Hero, A. O. (2014). Dynamic stochastic block models for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, *8*, 552–562.

Appendix A. Recouple Mapping to Dynamic Gravity Model

The mapping from the DGLM flow model to DGM parameter processes of Section 3.2 follows developments in Chen et al. (2018), summarized as follows. We work on logged DGM parameters $h_t = \log(\mu_t)$, $a_{it} = \log(\alpha_{it})$, $b_{jt} = \log(\beta_{jt})$, and $g_{ijt} = \log(\gamma_{ijt})$. To ensure identification and a one-one mapping between the models, the traditional zero-sum constraint is adopted for the main effects and interactions at each time. Then, given the full set of ϕ_{ijt} values in the DGLM flow model, define $f_{ijt} = \log(\phi_{ijt})$ for each $i = 1{:}I, j = 0{:}I$, at each time $t = 1{:}T$. The recoupling step is then implemented by computing DGM parameters as below. In the notation, subscript + indicates summation over the relevant index i or j. For each t,

- compute $\mu_t = \exp(h_t)$ where $h_t = f_{++t}/I(I+1)$; then,
- for each $i = 1{:}I$, compute $\alpha_{it} = \exp(a_{it})$ for $a_{it} = f_{i+t}/(I+1) - h_t$; then,

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

- for each $j = 0{:}I$, compute the destination node main effect $\beta_{jt} = \exp(b_{jt})$ for $b_{jt} = f_{+jt}/I - h_t$; finally,
- for each $i = 1{:}I$ and $j = 0{:}I$, compute the affinity (interaction) effect $\gamma_{ijt} = \exp(g_{ijt})$ where $g_{it} = f_{ijt} - h_t - a_{it} - b_{jt}$.

This is applied to all simulated ϕ_{ijt} values from the posterior analysis under the DGLMs to create implied posteriors for the DGM parameter processes.

Appendix B. Bayesian Analysis of Poisson DGLMs

Analysis is based on sequential Bayesian computation that combines variational Bayes approximation with linear Bayes updates (West & Harrison, 1997; Hartigan, 1969; Goldstein, 1976), along with retrospective sampling of posteriors for state vector trajectories, extending earlier algorithms for DGLM analysis.

As in Section 4, we focus on one node pair but ignore the node indices, so that we have a Poisson DGLM for x_t with mean ϕ_t where $\lambda_t = \log(\phi_t)$ is defined via the underlying linear, state-space, Markov model of Equation (4).

Sequential analysis: Forward filtering and learning

At time $t = 0$, specify a prior mean vector $\boldsymbol{m}_0$ and variance matrix $\boldsymbol{C}_0$ for the pre-initial state vector, denoted by $\boldsymbol{\theta}_0 \sim [\boldsymbol{m}_0, \boldsymbol{C}_0]$. Then, over every future time point $t > 0$, the evolution over $t-1$ to t, prediction of x_t from time $t-1$ and posterior update based on observing x_t follows the standard evolve/predict/update cycle of Bayesian state-space model analyses. In DGLMs, we use two approximations in the cycle. First, as the evolution noise distribution is specified only in terms of first and second order moments, the modeler is free to constrain implied state and predictive distributions to chosen forms; DGLMs use the variational Bayes concept to constraint to conjugate forms enabling fast and efficient computation, as well as defining predictive distributions of contextually appropriate (for count time series) negative binomial forms. This is coupled with the use of decision-theoretic linear Bayes approximations to feed back data information in the prior-posterior update (filtering) step at each time step, appropriately conditioning the mean vector and variance matrix for the state vector as new data is processed. Details are summarized below.

1. At time $t-1$, given all the previous data and information $\mathscr{D}_{t-1}$ the mean vector and variance matrix of the posterior for $\boldsymbol{\theta}_{t-1}$ are available as $\boldsymbol{\theta}_{t-1}|\mathscr{D}_{t-1} \sim [\boldsymbol{m}_{t-1}, \boldsymbol{C}_{t-1}]$.
2. By the state evolution equation, the implied time $t-1$ prior for $\boldsymbol{\theta}_t$ has moments $\boldsymbol{\theta}_t|\mathscr{D}_{t-1} \sim [\boldsymbol{a}_t, \boldsymbol{R}_t]$, where $\boldsymbol{a}_t = \boldsymbol{G}_t\boldsymbol{m}_{t-1}$ and $\boldsymbol{R}_t = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t$.
3. This implies that $\lambda_t = \log\phi_t$ has prior mean and variance given by $f_t = \boldsymbol{F}_t'\boldsymbol{a}_t$ and $q_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t$, respectively.
4. The implied prior for the latent Poisson rate is constrained to a conjugate gamma form based on the above information—a variational Bayes decision and constraint. That is, the modeling choice is made to specify $\phi_t|\mathscr{D}_{t-1} \sim Ga(r_t, c_t)$ with defining parameters consistent with the prior information, i.e., consistent with the moment constraints about on $\lambda_t = \log\phi_t$. Matching these moments to the gamma prior implies that r_t, c_t are given as solutions to the equations $f_t = \gamma(r_t) - \log c_t$ and $q_t = \dot{\gamma}(r_t)$, respectively, where $\gamma(\cdot)$ is the digamma function and $\dot{\gamma}(\cdot)$ is the trigamma function. These equations are easily solved numerically (most efficiently using the Newton–Raphson method) to give the values of r_t, c_t.
5. Forecasting x_t from time $t-1$, the conjugate Poisson/gamma structure implies a negative binomial predictive distribution $p(x_t|\mathscr{D}_{t-1})$.
6. On observing x_t at time t, the implied posterior for the latent Poisson rate is the conjugate form $\phi_t|\mathscr{D}_t \sim Ga(r_t + x_t, c_t + m_t)$ (where m_t is the relevant occupancy correction factor). For the natural parameter $\lambda_t = \log\phi_t$, this implies posterior moments $\lambda_t|\mathscr{D}_t \sim [f_t^*, q_t^*]$ given by
$$f_t^* = \gamma(r_t + x_t) - \log(c_t + m_{it}) \quad \text{and} \quad q_t^* = \dot{\gamma}(r_t + x_t)$$
and these are trivially calculated.
7. Using linear Bayes decision theory arguments, the posterior mean vector and variance matrix for the state vector $\boldsymbol{\theta}_t$ are conditioned on the new information x_t via the predictor-corrector forms that adjust the prior moments based on forecast accuracy. Specifically, the time $t-1$ to t posterior update of moments $[\boldsymbol{m}_*, \boldsymbol{C}_*]$ required to complete the time t filtering steps are given by
$$\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t(f_t^* - f_t)/q_t \quad \text{and} \quad \boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{A}_t'(q_t - q_t^*)$$
where $\boldsymbol{A}_t$ is the adaptive coefficient vector $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t/q_t$.

Discount specification of evolution variance matrices

Specification of the evolution variance matrix $\boldsymbol{W}_t$ at each time uses the standard single-discount factor approach in which $\boldsymbol{W}_t = (\boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t')(1-\delta)/\delta$. This corresponds to the time $t-1$ posterior variance matrix $V(\boldsymbol{\theta}_{t-1}|\mathscr{D}_{t-1}) = \boldsymbol{C}_{t-1}$ evolving to

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

the prior variance matrix $V(\boldsymbol{\theta}_t|\mathscr{D}_{t-1}) = \boldsymbol{R}_{t-1} = \boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t' + \boldsymbol{W}_t = (\boldsymbol{G}_t\boldsymbol{C}_{t-1}\boldsymbol{G}_t')/\delta$. That is, following the deterministic component (defined by $\boldsymbol{G}_t$) of the state evolution, the stochastic innovation term $\boldsymbol{\omega}_t$ in the evolution increases uncertainties about the state by "discounting" historical information at a rate defined by δ. The discount factor δ is a tuning parameter to be chosen. See, for example, West & Harrison (1997, Chapter 6) and Prado & West (2010, Section 4.3.6).

Retrospective analysis

The above analysis provides for sequential learning, i.e., forward filtering to process data as it arrives and sequentially update prior and posterior mean and variance matrices for the state vector $\boldsymbol{\theta}_t$ over time. At any time t, this enables inference on the current state and forecasts of coming data. For most network studies this is most relevant to online learning and monitoring of flows. Then, having processed data up to any time T, a key interest is in looking back over time to update inferences on the historical trajectories of state vectors, and any functions of them of interest. This is called retrospective analysis and can be best addressed in terms of simulation of posterior distributions over the full past history of states. This is enabled using DGLM extensions of the standard "backward sampling" algorithm for conditionally Gaussian, dynamic linear models. That is, exploiting the retrospective extrapolation of posterior mean vectors and variance matrices, and adopting the variational Bayes concept again to constrain the implicit backward innovations to multivariate normal distributions with moments defined by the linear Bayes retrospection, we impute and simulate historical *trajectories* of sets of states $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_T$ by recursing backwards in time as follows.

- At time $t = T$, sample the approximating normal posterior $\boldsymbol{\theta}_t|\mathscr{D}_T \sim N(\boldsymbol{m}_T, \boldsymbol{C}_T)$.
- Recurse back over times $t = T-1, T-2, \ldots, 1$, at each stage sampling the variational Bayes normal approximation to $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathscr{D}_T)$.
- At $t = 1$, save the sampled trajectory of states.
- Repeat to generate a random sample of trajectories.

From a Monte Carlo sample of trajectories, we can then directly map to any functions of the state vectors for inference. Centrally, this includes mapping to the sampled Poisson rate parameters and then to the network-wide, node-specific and node–node interaction parameter processes of the DGMs.

In terms of implied computation, the retrospective theory simplifies considerably in the single discount DGLM. Summary computations for simulation at time t of the state vector $\boldsymbol{\theta}_t$ from the implied conditional normal for $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathscr{D}_T$ has mean vector and variance matrix given by simplified versions of the general expressions in dynamic linear models (e.g. Prado & West, 2010, Sections 4.3.5 and 4.3.6). In detail, the conditional moments are simply

$$E(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathscr{D}_T) = (1-\delta)\boldsymbol{m}_t + \delta\boldsymbol{G}^{-1}\boldsymbol{\theta}_{t+1} \quad \text{and} \quad V(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathscr{D}_T) = (1-\delta)\boldsymbol{C}_t,$$

and the implied approximate normal distribution is trivially simulated.

Appendix C. More Examples of Affinity Effects

Self-affinities

Another pattern in self-affinities that is common across nodes resembles the overall activity μ_t; category Arts & Entertainment is a good example (Figure A1). Levels are high during the day and evening (8:00–20:00) with a large bump around 20:00. We also see a pattern that increases throughout the day and peaks at night; the category Arts & Entertainment/TV & Video is one of the few examples (Figure A2).

Entering affinities

Beyond the examples in Section 6.1, another pattern in entering affinities, exemplified by the categories Health/Pharmacy/-Drugs & Medications and Finance/Investing (Figures A3 and A4), has several bumps during the workday (8:00–16:00), and is relatively low otherwise; this again has implications for selecting ad content.

A further pattern in entering affinities has high intensity in the morning and then is stable during the day. The entering affinity of Shopping is a good example—see Figure A5. But things may be complex—perhaps someone on this website at a non-standard time is more apt to purchase than a casual browser.

One additional pattern in entering affinities is notable for multiple pronounced peaks. Beauty & Fitness/Fashion & Style (Figure A6) has three peaks, at about 7:00–8:00, at about 12:00–16:00, and at about 20:00–24:00. September 17 is during the New York Fashion Week, which may lead to atypical behavior.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

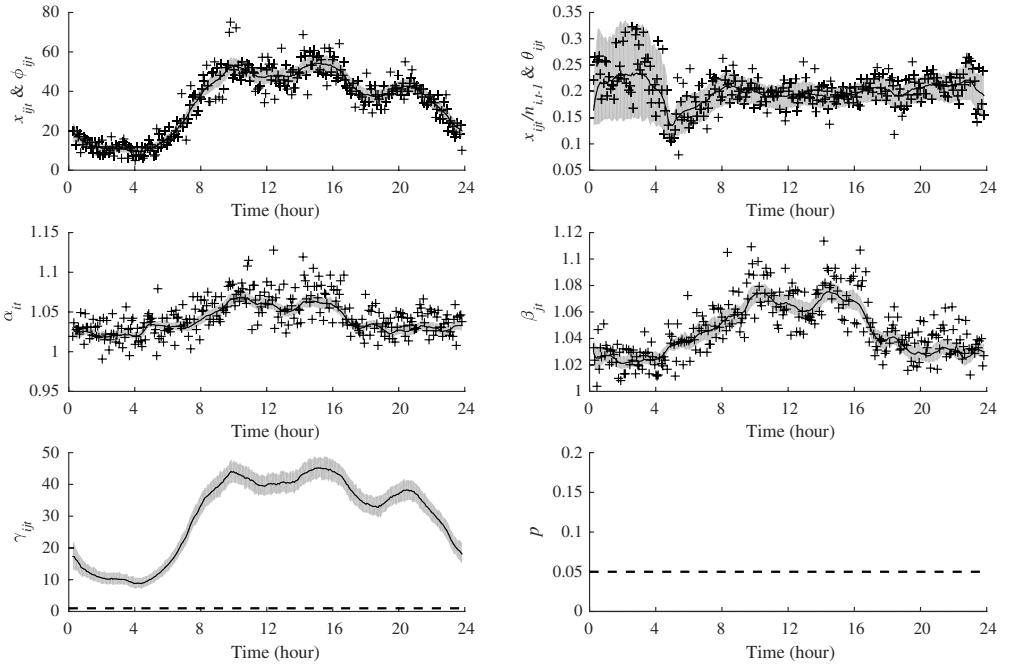

Figure A1. Posterior summaries for DGM parameters for transitions staying at node i = Arts & Entertainment with details as in Figure 7.

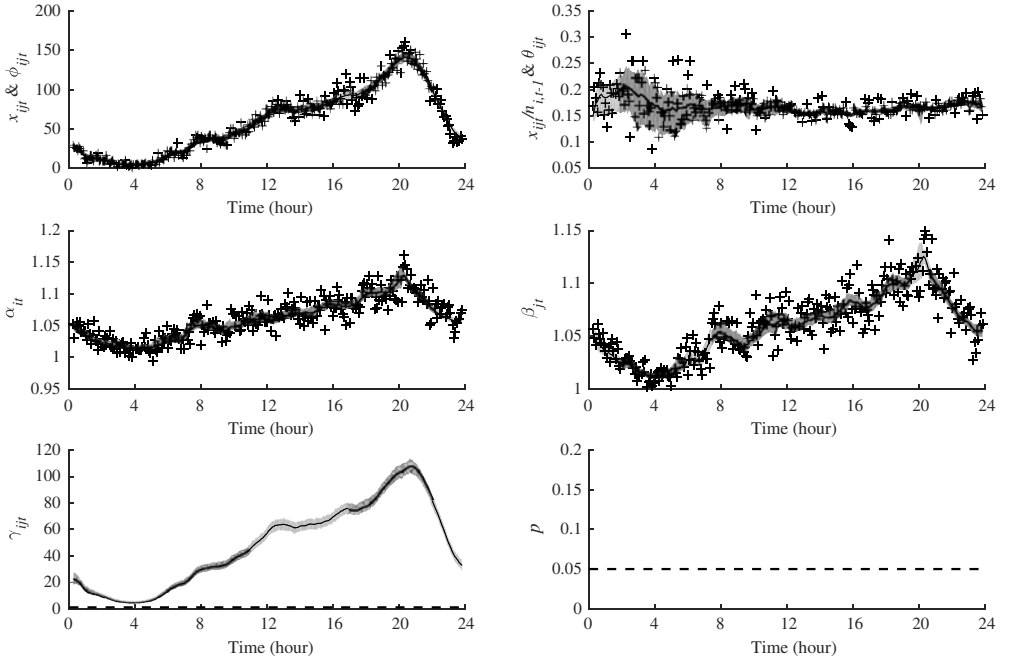

Figure A2. Posterior summaries for DGM parameters for transitions staying at node i = Arts & Entertainment/TV & Video, with details as in Figure 7.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

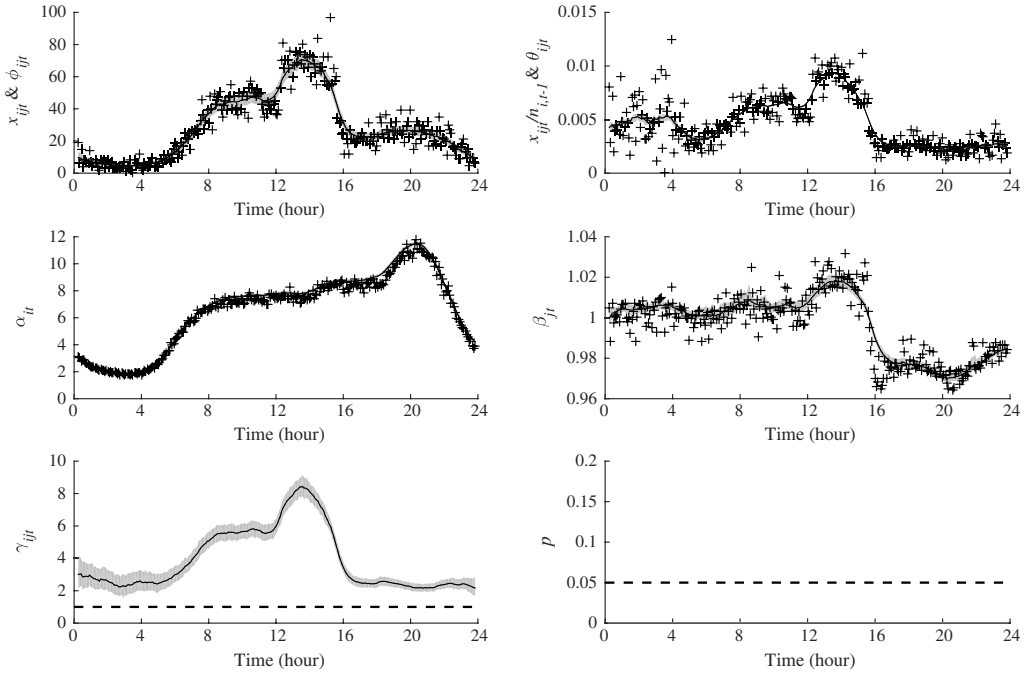

Figure A3. Posterior summaries for DGM parameters for transitions entering node i = Health/Pharmacy/Drugs & Medications with details as in Figure 9. An interesting bump is noted in the afternoon (about 12:00–16:00).

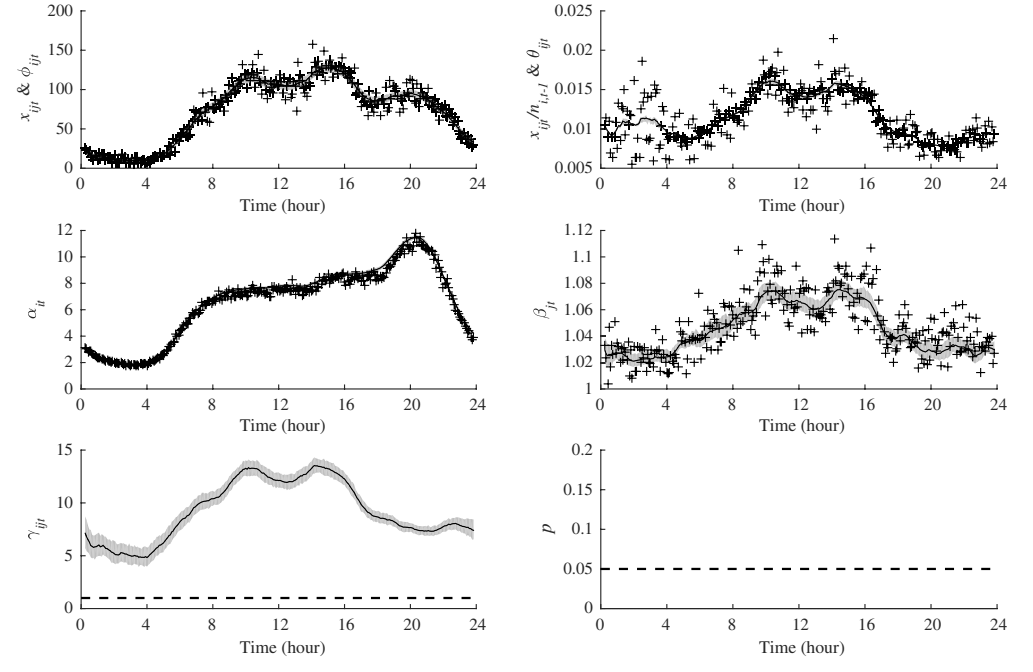

Figure A4. Posterior summaries for DGM parameters for transitions entering node i = Finance/Investing with details as in Figure 9. Note a maintained high level during the day (8:00–16:00).

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

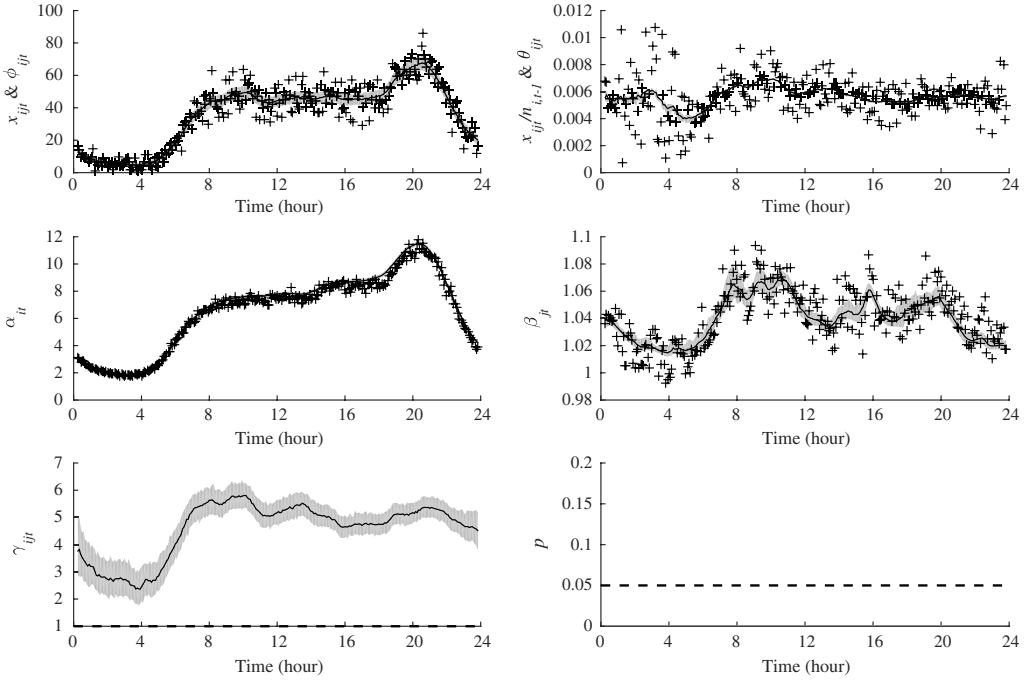

Figure A5. Posterior summaries for DGM parameters for transitions entering node i = Shopping with details as in Figure 9.

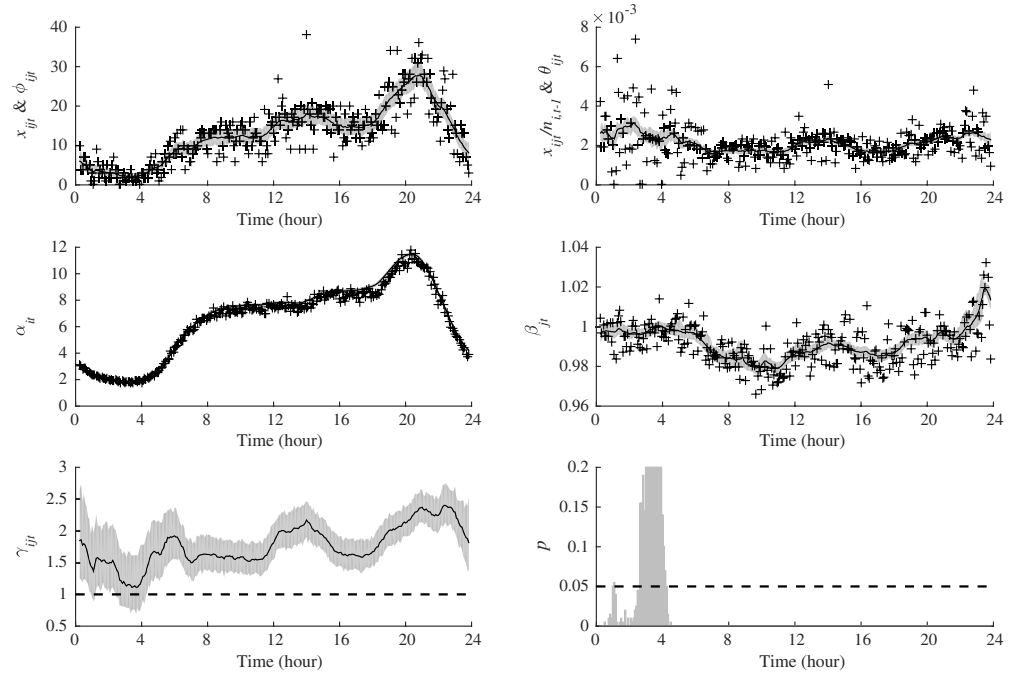

Figure A6. Posterior summaries for DGM parameters for transitions entering node i = Beauty & Fitness/Fashion & Style with details as in Figure 9.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

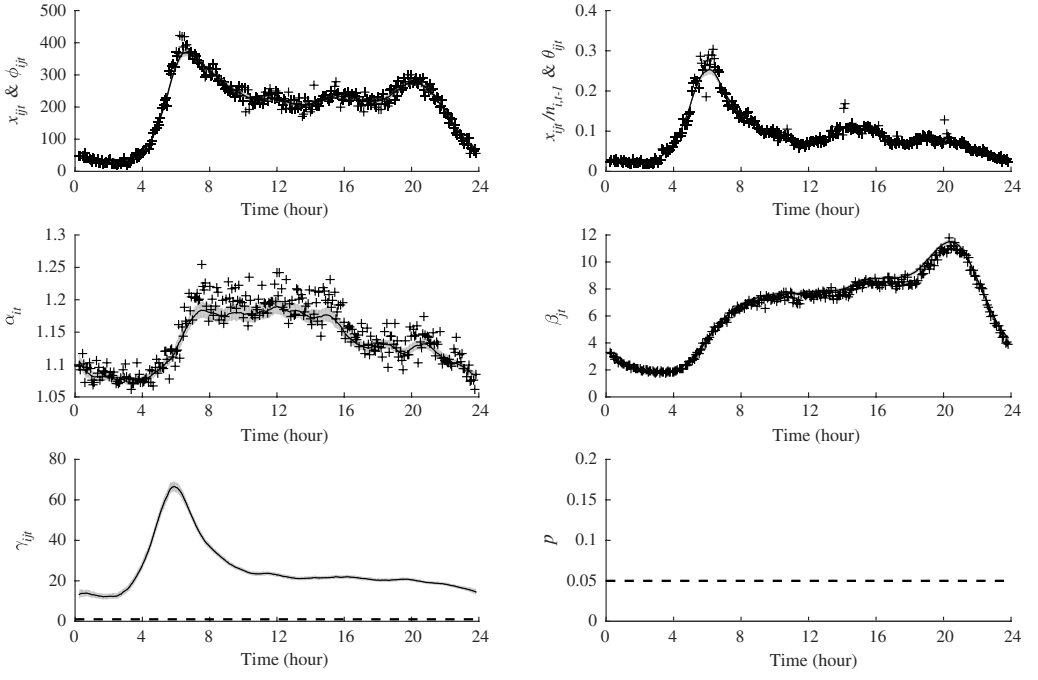

Figure A7. Posterior summaries for DGM parameters for transitions exiting from node $i =$ News/Weather with details as in Figure 11.

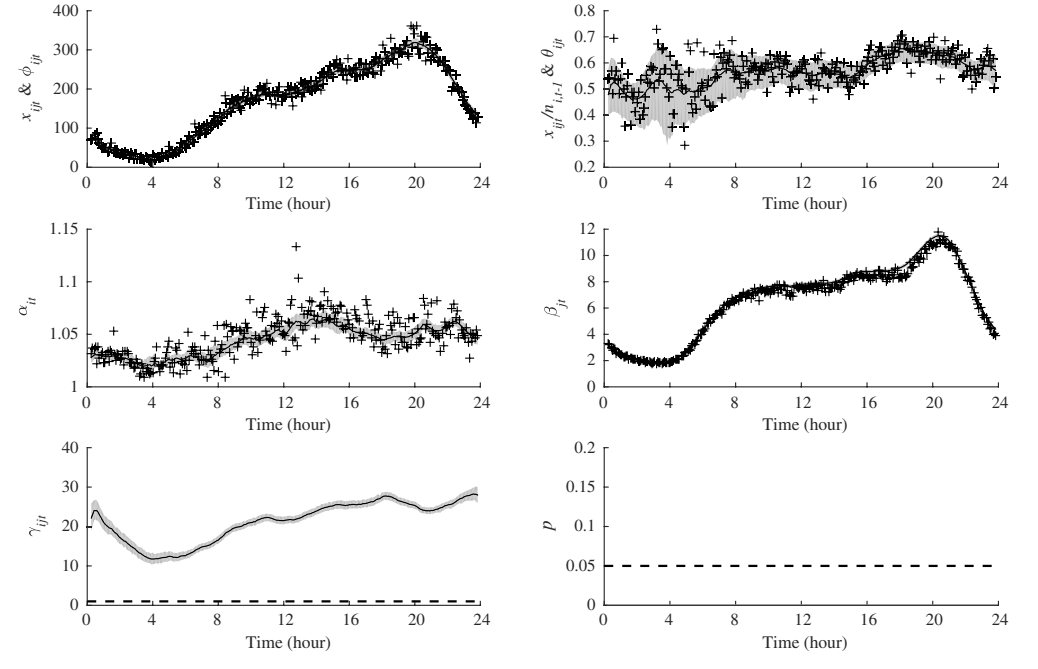

Figure A8. Posterior summaries for DGM parameters for transitions exiting from node $i =$ Arts & Entertainment/Music & Audio with details as in Figure 11. Note the peak at night.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

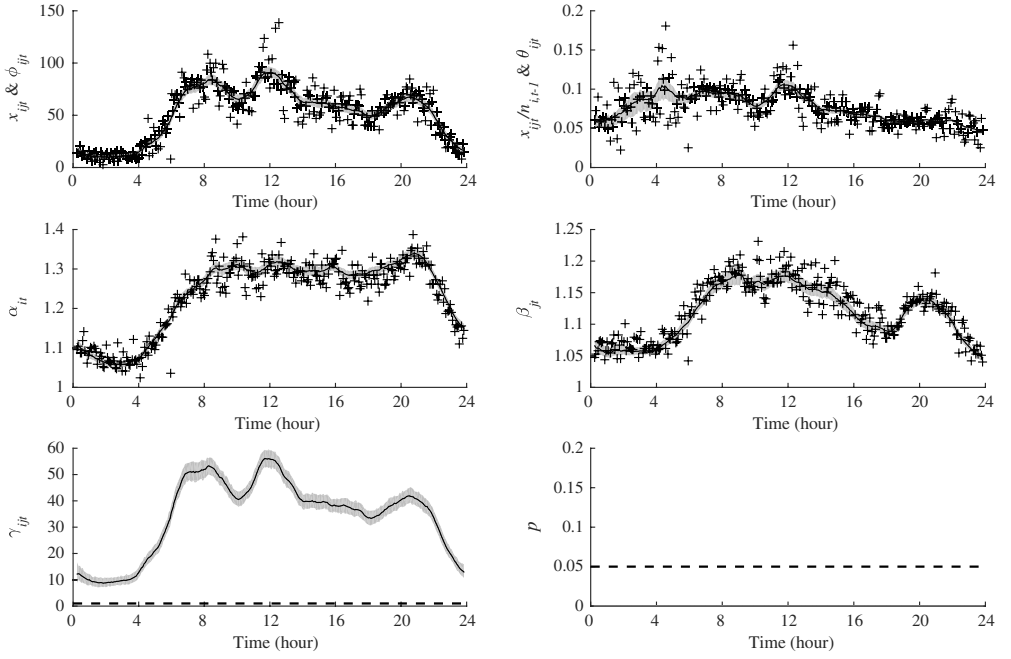

Figure A9. Posterior summaries for DGM parameters for transitions from node i = News → j = News/Local News with details as in Figure 13.

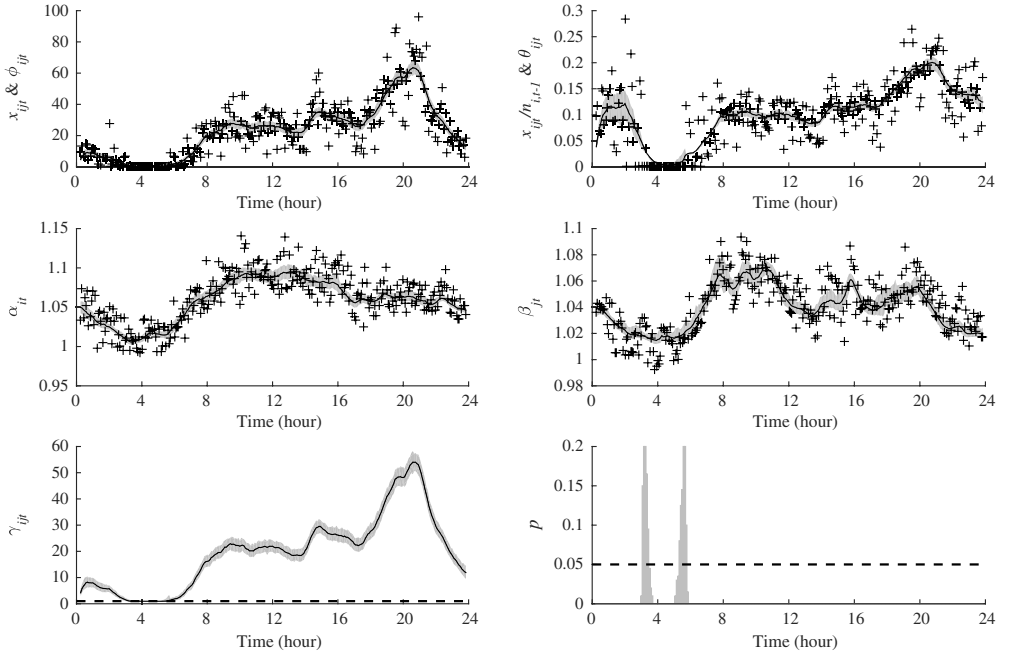

Figure A10. Posterior summaries for DGM parameters for transitions from node i = News/Technology News → j = Shopping with details as in Figure 13.

https://doi.org/10.1017/nws.2019.10 Published online by Cambridge University Press

Exiting affinities

Other patterns have been detected in trajectories of exiting affinities. One has a peak, but is otherwise low, as in the exiting affinity for News/Weather (Figure A7). Clearly, people are visiting this site only to learn about the weather forecast, and then leave for work.

A third pattern in trajectories of exiting affinities increases from 4:00 in the morning until the evening, indicating that these categories are increasingly losing users throughout the day. Such categories include Arts & Entertainment/Music & Audio (Figure A8).

Distinct node pair affinities

The affinity from News to News/Local News (Figure A9) has three peaks in the day. The first is around 8:00, the second around noon, and last around 20:00. The peaks indicate that people who are interested in national news also check the local news, and that the timing coincides with leisure.

The last example is the affinity from News/Technology News to Shopping (Figure A10). This peaks around 20:00, which suggests the users who have read technology news start to explore technology purchase, and should be a clear signal for ad display.

Cite this article: Chen X., Banks D., and West M. (2019). Bayesian dynamic modeling and monitoring of network flows. *Network Science* **7**, 292–318. https://doi.org/10.1017/nws.2019.10