Bayesian Spillover Graphs for Dynamic Networks

Grace Deng
David S. Matteson

1Department of Statistics & Data Science, Cornell University, Ithaca, NY, USA

Abstract

We present Bayesian Spillover Graphs (BSG), a novel method for learning temporal relationships, identifying critical nodes, and quantifying uncertainty for multi-horizon spillover effects in a dynamic system. BSG leverages both an interpretable framework via forecast error variance decompositions (FEVD) and comprehensive uncertainty quantification via Bayesian time series models to contextualize temporal relationships in terms of systemic risk and prediction variability. Forecast horizon hyperparameter $h$ allows for learning both short-term and equilibrium state network behaviors. Experiments for identifying source and sink nodes under various graph and error specifications show significant performance gains against state-of-the-art Bayesian Networks and deep-learning baselines. Applications to real-world systems also showcase BSG as an exploratory analysis tool for uncovering indirect spillovers and quantifying systemic risk.

1 INTRODUCTION

We consider the task of learning temporal interactions and important components over time in a dynamic network. Many real-world systems can be described by a multivariate time series (MTS) and a natural framework for analyzing temporal relationships is Granger causality [Granger, 1969], which tests for whether one time series is useful for forecasting another one. Network Granger causality (NGC) [Basu et al., 2015] extends this concept into the multivariate setting. NGC is useful for identifying one-step ahead predictive relationships within a system, and may be considered causal under very specific conditions [Pearl et al., 2000].

Many methods have been developed to estimate NGC. Vector Autoregression (VAR) [Sims, 1980] and its variants [Lütkepohl, 2005] remain a standard-bearer for macroeconomics and financial forecasting. Bayesian networks [Pearl, 2011] [Ben-Gal, 2008] are also a powerful collection of probabilistic graph models for learning NGC, usually via a directed acyclic graph (DAG). Dynamic Bayesian Networks (DBN) [Murphy, 2002] are particularly useful for modeling state changes and temporal structure learning, although it is restricted by acyclic representations. Alternative methods for estimating NGC adjacency matrices use deep learning variants, e.g., attention networks [Nauta et al., 2019], Statistical Recurrent Units (SRU) [Khanna and Tan, 2019], and sparse RNNs [Tank et al., 2018]. Recently, Generalized Vector Autoregression (GVAR) [Marcinkevičiūtė and Voge, 2021], which utilizes Self-explaining Neural Nets (SENN), also proposed aggregating model coefficients over lagged time series to estimate signs of NGC in addition to edge detection.

However, NGC has several drawbacks. First, it is not designed to capture cumulative interactions or multi-step ahead effects that evolve over longer forecast horizons [Marcinkevičiūtė and Voge, 2021], which may be particularly important in forecasting or inference for real-world systems [Diebold and Yilmaz, 2014] [Billio et al., 2012]. Spillovers, in particular, is an interesting subset of temporal relationships (graph edges) that can materialize beyond 1-step ahead forecasts [Diebold and Yilmaz, 2015] in the context of forecast variability and network connectivity. Furthermore, indirect spillovers between components can also manifest via intermediary nodes despite having no direct link via NGC. Estimating NGC via DAG constraints are hence not representative of true network interactions, which can be self-directed, bi-directional, or cyclic over time. Prior NGC methods also do not quantify strengths of temporal relationships [Marcinkevičiūtė and Voge, 2021] nor provide ample interpretation for related graph measures. Identification of important nodes relies on standard graph theory metrics [Kramer et al., 2009] [Yusoff and Sharifi, 2016] such as eigen-centrality [Bonacich, 1987] or in/out degrees [Freeman, 1978]. These metrics are also static point estimates based on NGC graphs.
And although methods such as GVAR offer sign estimation for temporal relationships, the actual coefficient values (edge weights) are not necessarily meaningful.

![Dynamic Systems](Image)

**Previous Methods**
- Network Granger Causality (NGC) Methods
  - Vector Autoregression
  - Dynamic Bayesian Networks
  - Generalized Vector Autoregression

**Bayesian Spillover Graphs (BSG)**
- Bayesian Model Estimation: \( f(\beta, \Sigma_h | Z_h, X) \) & Forecast Horizon \( h \)
- BSG Posterior Distribution: \( f(G(\beta, \Sigma_h | X)) \)

**Std. Network Measures**
- In/Out Degrees
- Eigen-centrality
- Closeness Centrality
- Betweenness Centrality

- Only applies to static NGC graphs
- No uncertainty quantification

**Forecast Error Variance Decomposition (FEVD) Framework**
1. Temporal Spillovers: Graph Edge \( \text{spillover} \)
2. Network Measures: Influence \( \text{influence} \)
3. Vulnerability \( \text{vulnerability} \)
- Dynamic temporal relationships + spillovers over multiple forecast horizons
- Interpretable network measures
- Ranks node influence + vulnerability
- Uncertainty quantification

Figure 1: Comparison of BSG vs. Prior NGC Methods. BSG combines Bayesian VAR estimation with interpretable FEVD framework over forecast horizons \( h \) to quantify strength of temporal interactions (BSG edge weights) and systemically important nodes over time.

To summarize, the major drawbacks of current methods are (1) lack of flexibility for observing network interactions over multiple forecast horizons, (2) lack of interpretable network measures that are contextualized, (3) lack of uncertainty quantification for strength of temporal relationships and node influence. To this end, a promising solution is to leverage forecast error variance decomposition (FEVD) from classic time series forecasting, which estimates the temporal effect of shocks to individual nodes in the system \cite{Barbaglia2020, Tsay2013, Diebold2015}, and Bayesian VAR models \cite{Rossi2012, Koop2010}, which provide comprehensive uncertainty quantification.

In particular, the formulae behind FEVD is a cornerstone of classic multivariate time series analysis when we are interested in relationships between time series components. It is commonly cited as (generalized) impulse response functions in statistical literature and multiplier analysis in economic literature \cite{Tsay2013}, and key applications include quantifying the effect of one time series component over forecast horizons, a key advantage over NGC. Under careful assumptions and conditions, it can also be a viable causal inference tool to analyze impact of specific policies \cite{Swanson1997}. The idea of standardizing FEVD as a measure of risk and connectivity has been motivated by macroeconomic and financial applications \cite{Diebold2015, Barbaglia2020}.

Formally, we define spillovers as the predicted impact of one component on all other components in a dynamic network with respect to forecast variability and forecast horizon \( h \). Intuitively, we are learning how unexpected shocks in one component cascades throughout the network to all other components, as well as examining how this impact evolves over time. Statistically, we can estimate \( h \)-step ahead spillovers based on normalized FEVD for one-step ahead forecasts and beyond after parameter estimation via Bayesian VAR; interpretation of resulting spillover effects is then contextualized by the input time series while also accounting for parameter estimation variability.

**Motivation.** We present Bayesian Spillover Graph (BSG) for analyzing temporal interactions over multiple forecast horizons, identification of systemic influential and at-risk nodes, and uncertainty quantification for novel network measures with interpretation beyond simple NGC. BSG is both a powerful exploratory data analysis and inference tool; key contributions include:

1. We model temporal relationships in a dynamic system based on a single observed MTS; forecast horizon hyperparameter \( h \) allows for flexibility in learning short-term vs. long-term spillover effects.
2. We propose interpretable network measures for contextualizing spillovers with respect to prediction variability and identifying sink and source nodes within a dynamic network. We demonstrate the robustness of these measures across various graph and error dependency specifications.
3. We provide uncertainty quantification for BSG measures through functionals of model parameter posterior distributions via Bayesian estimation, compared to point-estimates from baseline VAR and NGC retrieval methods. We showcase how BSG can quantify strengths of temporal interactions (including spillovers) and identify systemically vulnerable nodes in a wildfire risk application.

We emphasize the distinction between Bayesian DAGs versus BSG, which models temporal, bi-directional relationships that can potentially amplify spillovers over multi-step horizons. DAG structure is a popular assumption in causal inference and can be viewed as a special case of BSG. BSG learns important edges (temporal interactions) and nodes (time series components) directly from estimated statistical network metrics. It also accounts for various dependencies in error terms that deviate from standard Gaussian noises, which are more descriptive of real-world systems. A brief overview of BSG vs. prior methods is shown in Figure 1.
2 METHODOLOGY

2.1 VECTOR AUTOREGRESSION (VAR)

Let \( z_t \) be a stationary \( d \)-dimensional multivariate time series, and \( \{ z_{jt} \} \) be the \( j \)-th component of this time series at time \( t \). A VAR(p) model with order \( p \) is defined as:

\[
z_t = \phi_0 + \sum_{i=1}^{p} \phi_i z_{t-i} + a_t \tag{1}
\]

where \( \phi_0 \) is a \( d \)-dimensional constant, \( \phi_i \) is the \( d \times d \) lag \( i \) coefficient matrix for \( i \geq 0 \), and \( a_t \) is a sequence of \( i.i.d \) random vectors with mean 0 and covariance matrix \( \Sigma_n \).

**Bayesian Estimation.** We utilize a Bayesian approach [Tsay, 2013] for estimating unknown model parameters \( \{ \beta', \Sigma_n \} \) for a VAR(p) time series with length \( T \), where \( \beta' = [\phi_0, \phi_1, ..., \phi_p] \):

\[
Z = X\beta + A \tag{2}
\]

where \( Z \) and \( A \) are \( (T-p) \times d \) matrices, and the \( i \)th row is \( z'_{p+i} \) and \( a'_{p+i} \). \( \beta' \) is a \( d \times (dp+1) \) matrix, and \( X \) is a \( (T-p) \times (dp+1) \) design matrix with \( i \)th row as \((1, z'_{\beta p+i-1}, z'_{i})\). The likelihood function for the data is

\[
f(Z|\beta, \Sigma_n) \propto |\Sigma_n|^{-n/2} \exp\left[-\frac{1}{2} tr(\{Z-X\beta'(Z-X\beta)\Sigma_n^{-1}\})\right] \tag{3}
\]

where \( n = T - p \) is the effective sample size. We utilize Normal-inverse-Wishart conjugate priors \( f(\beta, \Sigma_n) = f(\Sigma_n) f(\beta|\Sigma_n) \):

\[
f(\Sigma_n) \sim W^{-1}(V_0, n_0) \tag{4}
\]

\[
f(vec(\beta)|\Sigma_n) \sim N(vec(\beta_0), \Sigma_n \otimes C^{-1}) \tag{5}
\]

where hyperparameters \( V_0 \) is a \( d \times d \) matrix, \( n_0 \) is some real number, \( C \) is a \( (dp+1) \times (dp+1) \) matrix, and \( \beta_0 \) is a \((dp+1) \times d \) matrix. The posterior distribution is then:

\[
f(\Sigma_n|Z, X) \sim W^{-1}(V_0 + S, n_0 + n) \tag{6}
\]

\[
f(vec(\beta)|Z, X, \Sigma_n) \sim N(vec(\bar{\beta}), \Sigma_n \otimes (X'X + C^{-1})^{-1}) \tag{7}
\]

where \( \bar{\beta} = (X'X + C^{-1})^{-1}(X'X\bar{\beta} + C\beta_0) \) and \( S = (Z - X\bar{\beta})'(Z - X\bar{\beta}) + (\bar{\beta} - \beta_0)'C(\bar{\beta} - \beta_0) \) based on hyperparameter choices from the prior; \( \bar{\beta} \) is the least-squares estimate of \( \beta \). Usually, \( V_0 \) is set to identity \( I_d \) and \( n_0 \) is a small number; as sample size \( n \) increases, the choice of \( n_0 \) has very little effect on the final posterior. Similarly, we can choose vague priors for \( vec(\beta) \) by letting \( vec(\beta_0) = 0 \) and \( C^{-1} = c_0 I_{dp+1} \), where \( c_0 \) is some large real number, and hence the posterior distribution \( f(vec(\beta)|Z, X, \Sigma_n) \) is also mainly updated via the data \( X \).

Although \( \Sigma_n \) is unknown, we can sample \( M \) i.i.d samples from the joint posterior distribution by iterative sampling from \( f(\Sigma_n|Z, X) \) and \( f(vec(\beta)|Z, X, \Sigma_n) \), replacing \( \Sigma_n \) with posterior estimate \( \Sigma^{(m)}_n \).

2.2 BAYESIAN SPILLOVER GRAPHS

In brief, we adopt Bayesian estimation for Vector Autoregressions (VAR) to estimate posterior distribution for model parameters \([\beta', \Sigma_n] \) from a single realized MTS. We then construct \( G_h(\beta, \Sigma_n|Z) \), the BSG for forecast horizon \( h \), with components of MTS as nodes and temporal interactions as directed, weighted edges. Specifically, we can estimate BSG edge weights by computing \( h \)-step ahead normalized spillovers between two nodes via FEVD for \( M \) posterior samples of \([\beta', \Sigma_n] \), and taking averages over \( M \). Consequently, BSG is an interpretable graph where both magnitude and specific values of edges are meaningful.

We also introduce three network measures based on functionals of BSG: the spillover index, vulnerability score, and influence score. These measures describe systemic-wide behavior over time and are useful for monitoring influential and-at-risk nodes for a dynamic network. With a Bayesian framework, we can quantify uncertainty for both BSG edges and network measures. Under stationarity assumptions, estimated normalized spillovers are finite after some fixed forecast horizon \( h \).

**Interpretable BSG Edges from Forecast Error Variance Decomposition.** We adopt generalized FEVD for analyzing \( h \)-step ahead spillover effects [Diebold and Yilmaz, 2014] [Diebold and Yilmaz, 2015]; the accuracy of a forecast can be measured by its forecast error. Let \( \sigma_{kk} \) be the \( k \)-th diagonal of \( \Sigma_n \), and \( \psi_{jk} \) be the coefficient matrix for a non-orthogonalized VAR under an infinite moving-average representation. The \( j-k \)-th entry of the \( h \)-step ahead forecast error variance is

\[
\sigma_{h,jk} = \frac{\sigma_{kk}^{-1} \psi_{jk}^{'h-1}[\psi_{jk}]^2}{\sum_{i=0}^{h-1}[\psi_{jk}]^{'i}} \tag{8}
\]

which measures the amount of information of the \( h \)-step ahead forecast error variance for variable \( j \) accounted for by innovations/exogenous shocks to variable \( k \). The \( h \)-step ahead normalized spillover from component \( k \) to \( j \) is:

\[
s_{h,k}^{k\rightarrow j} = 100 * \tilde{w}_{h,jk}, \quad \tilde{w}_{h,jk} = \frac{w_{h,jk}}{\sum_{k=1}^{M} w_{h,jk}} \tag{9}
\]

where \( \tilde{w}_{h,jk} \) is the normalized variance decomposition. \( s_{h,k}^{k\rightarrow j} \) is the proportion of the \( h \)-step ahead forecast error variance for node \( j \) attributed to changes in node \( k \), and becomes the weight for a directed edge from node \( k \) to \( j \) for BSG. \( G_h(\beta, \Sigma_n|Z) \). This definition makes BSG an interpretable graph with respect to forecast errors, with direct explanation of edge weight meaning. Prior methods such as GVAR would only estimate the sign of a temporal relationship [Marcinkevics and Vogl, 2021]. See Algorithm [ ] for details on estimating BSG edges from posterior distributions of Bayesian VAR parameters.
BSG Network Measures as Systemic Risk Indicators.

We propose novel BSG network measures based on functionals of BSG edges over forecast horizon $h$ that can describe system-wide behavior and node importance over time. The goal is to quantify cumulative temporal interactions and spillovers within a system, as well as identify strongly influential or vulnerable nodes.

We define the $h$-spillover index as the magnitude of $h$-step normalized spillovers across all components, which describes the total spillover effect experienced over the full graph. The $h$-spillover index can be viewed as a measure of cumulative risk within the system after $h$ time periods; the higher it is, the more fragile the system is to innovations in any individual node.

$$S(\cdot) = S_h = \sum_{j=1}^{d} \sum_{k=1}^{d} s_{h}^{k\rightarrow j}(10)$$

We may then be interested in identifying specific nodes at high risk over the full graph. For example, say we wanted to rank the individual nodes by the magnitude of spillovers from all other components to a specific component $j$.

$$V(\cdot) = V_h = \sum_{s_{h}^{k\rightarrow j}}(11)$$

$s_{h}^{k\rightarrow j}$ can be viewed as the vulnerability score for a specific node at $h$-steps ahead, and can theoretically take on values between $[0, 100]$. The vulnerability score for node $j$ can be interpreted as the proportion of FEVD not attributed to innovations to $j$ itself. In particular, nodes with higher vulnerability are more susceptible to shocks and cascading effects from other components within the system.

Alternatively, we may be interested in pinpointing the sources of risks to the system. We define the influence score for a specific node, $s_{h}^{k\rightarrow s}$, as:

$$I(\cdot) = I_h = \sum_{s_{h}^{k\rightarrow j} \forall k, k \neq j}(12)$$

Note that the numerator of this expression quantifies the total spillover effect on the graph originating from component $k$, which is then standardized by the $h$-spillover index. This allows us to interpret the influence score for node $k$ as the proportion of total spillover effect on the entire system attributed to innovations in $k$, which again takes on values between $[0, 100]$ and is comparable across different networks. In particular, nodes with higher influence leads to greater impact on the entire system if there is a shock or change to the node. Collectively, these BSG network measures have wide applicability for describing real-world systems and as systemic risk indicators (SRI), which captures holistic risk arising from overall network connectivity [Che-Castaldo et al., 2021; De Bandt and Hartmann, 2000].

BSG Estimation & Uncertainty Quantification. Given a single realized MTS $Z_t$, we can construct BSG $G_h(\beta, \Sigma_n|Z)$ directly via Bayesian VAR estimation. We first draw $M$ samples, $\{\beta^{(m)}, \Sigma_n^{(m)}\}$, from the posterior distribution of model parameters. For fixed forecast horizon $h$, we compute $w_{h,jk}$ the $h$-step ahead forecast error variance, for each sample. BSG edges are then constructed by averaging over $M$, where $s_{h}^{k\rightarrow j} = 1/\sqrt{M} \sum_{m=1}^{M} s_{h}^{k\rightarrow j}(m)$ is a weighted directed edge from node $k$ to node $j$. BSG nodes are the individual components of $Z_t$. BSG network measures can also be computed directly by averaging over $M$ samples, e.g., the influence score for node $k$ would be estimated via $\bar{s}_{h}^{k\rightarrow s} = \sum_{m=1}^{M} \sum_{j,j \neq k} s_{h}^{(m),k\rightarrow j}/s_{h}^{(m)}$. See Algorithm 1. This process also allows for uncertainty quantification for any BSG edge or network measure by constructing credible intervals over $M$ estimates. We can also leverage the simplicity of Highest Posterior Density Interval (HPDI) or Bayes Factor [Kass and Raftery, 1995]. See Section 5 for an example with California wildfire data.

Stationarity and Optimal $h^*$ for Equilibrium BSG. A VAR(1) model can be written with an infinite sum as:

$$z_t = \mu + \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

where $\psi_1 = \phi_1^*$ for $i \geq 0$ and $\mu$ is a $d$-dimensional constant. See Appendix A for details. If the series is stationary, then the absolute value of the eigenvalues of $\phi_1$ will be strictly less than 1. Various transformations, including detrending, removing seasonality, or differencing the series [Granger and Newbold, 2014] are recommended to ensure stationarity before parameter estimation. MTS with DAG temporal network structures can be viewed as a subset of VARs with restrictive assumptions on $\beta$. In the special case of a VAR(1) model where the temporal network structure of $z_t$ can be described by a DAG, $z_t$ is stationary; see Theorem 1 and proof in Appendix B.

Theorem 1. If $\phi_1$ is a DAG, then (1) no component-wise autocorrelation exists, (2) $\phi_1$ can be specified by a strictly triangular matrix, (3) all eigenvalues of $\phi_1$ are 0 and hence $z_t$ is stationary.

Under stationarity, BSG can reliably model cumulative response functions if shocks are not persistent and the system will return to equilibrium. See Algorithm 1 for choosing the optimal $h^*$-step. The horizon $h$ can be interpreted as a tuning parameter that controls the trade-off between learning immediate versus cumulative effects for BSG.

3 BSG FOR QUANTIFYING INDIRECT SPILOVERS

We showcase how BSG models temporal spillovers that materialize after multiple periods. Consider a 5-dimensional
VAR(1) time series represented by the directed graph of temporal interactions ($\phi_1$) in Figure 3 with true parameters:

$$\phi_1 = \begin{bmatrix} 0.8 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.5 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.8 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.7 & 0.8 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.4 & 0.8 \end{bmatrix}$$

Eigen-decomposition of $\phi_1$ indicates that all eigenvalues have magnitude $\leq 1$ and this network is stationary with standard independent error terms. Nodes 3 and 1 are analogous to source nodes with high out-degree centrality, and 5 and 3 to sink nodes with high in-degree centrality [Borgatti 2005; Bollobás 2012; Goldberg et al. 1989]. Node 5 will experience spillovers from Node 3 via Node 4 after multiple time periods, but this relationship is omitted in a simple NGC. This limitation is suitably addressed with a BSG with $h > 1$; see Figure 2 where indirect spillover (red arrow from 3 to 5) becomes stronger as $h$ increases.

In Figure 4, we plot average BSG directed edge weights ($h$-step ahead normalized spillover) from Nodes 1-4 into Node 5. The indirect spillover effect through intermediary Node 4 manifests after 2-steps ahead forecast and significantly amplifies as the forecast horizon increases (turquoise line) before flattening after $h = 17$. We can directly interpret this edge: the posterior mean for $s_{20 \to 5}$ is 80.1% with 95% HPDI of (71.9%, 87.7%), which predicts that after 20 periods, roughly 80.1% of forecast variability for node 5 can be attributed to changes in node 3. In contrast, the edge from Node 4 to Node 5 rapidly declines past $h = 4$. With prior methods of only estimating static NGC, we would not be able to observe nor quantify these spillover effects that evolve over longer forecast horizons.

### 4 BSG FOR IDENTIFYING NETWORK SOURCE & SINK NODES

We illustrate how BSG network measures accurately ranks and identifies nodes of interest compared to baselines with...
simulated MTS. Since relative order matters, this is a ranking instead of prediction task. Performance is evaluated by Normalized Discounted Cumulative Gain (NDCG) [Valizadegan et al., 2009]. NDCG measures ranking quality of a node ordering by BSG network measures or other graph measures, e.g., source nodes are ranked highly influential. NDCG is between [0, 1] and directly comparable across methods; see Appendix C.

**Identifying Nodes Across Network Specifications.** 3 stationary network specifications ($\phi_1$) are used for simulating 5 MTS replicates: (1) a DAG, (2) a directed cyclic graph with autocorrelation = 0.5, and (3) a bi-partite graph. Networks (1) and (2) have 5 source and sink nodes and Network (3) has 10 source and sink nodes; all have independent Gaussian noise for $\Sigma_a$. Edge weights are sampled from a Unif(0,1) distribution; $T = 500$ and $d = 20$ for each network. We construct BSGs $G_h(\beta, \Sigma_a|Z)$ with edges $\bar{s}^{k\to j}_h$ for $h = \{1, 5, 10\}$, and use influence and vulnerability scores for ranking source and sink nodes respectively. The first set of baselines are 4 standard graph measures on a NGC graph: in/out degree distributions, eigen-centrality, betweenness centrality, and closeness centrality. NGC is constructed from a VAR(1) model fitted via the MTS package, and significant edges are identified via multiple-testing with Benjamini-Hochberg procedure [Benjamini and Hochberg, 1995]. Another set of baselines is DBN and GV AR combined with the 4 graph measures above, because these methods are designed only to retrieve NGC graphs. For fairness of comparison, GV AR lag is restricted to 1 and run with default hidden units/layer (50), hyperparameters $\lambda = 0.1$ and $\gamma = 0.01$, and 500 epochs in PyTorch. DBN uses default settings with the dbnR package.

Average NDCG are reported in Table 1 for each combination of network specification and h-step-ahead forecast horizon.

---

**Algorithm 1 Estimating Bayesian Spillover Graph with Optimal $h^*$**

1. Draw $M$ posterior samples for $\beta = [\phi_0, \phi_1, ..., \phi_p], \Sigma_a$
2. while $m < M$ do sample
3. $\Sigma_a(m) \sim W^{-1}(V_0 + \tilde{S}, n_0 + n)$
4. $vec(\beta(m)) \sim N(vec(\beta), \Sigma_a(m) \otimes (X'X + C)^{-1})$
5. end while
6. Iterate over $h$ until converge
7. for $h$ in 1, 2, ..., $H$ and $\epsilon > 0$ do
8. Compute $w(m)_{h,jk}$ from $\Sigma_a(m), \beta(m)$
9. Compute $s(m)_{h,jk}$ from $w(m)_{h,jk}$
10. Compute posterior mean $\bar{s}^{k\to j}_h = \frac{1}{M} \sum_{m=1}^{M} s^{(m)\to j}_h$
11. if $|\bar{s}^{k\to j}_h - \bar{s}^{k\to j}_{h-1}| < \epsilon, \forall j, k$ then
12. $h^* = h$
13. end if
14. end for
15. Construct BSG $G_h(\beta, \Sigma_a|Z)$ with edges $\bar{s}^{k\to j}_h$.

---

1Example code at https://github.com/gdeng96/bsg
2GV AR code available at https://github.com/i6092467/GV AR

---

**Figure 5** BSG Accuracy for identifying source nodes via influence scores, w.r.t. h-step ahead forecast horizon and different $\sigma_{jk}$ strengths.
We compare BSG influence/vulnerability scores vs. benchmarking with Network (2), which allows for bi-directional action strengths. We generate 5 MTS replicates using α, β, γ, δ = 1, and pairwise covariance is {σ_{kk} = 1}, and standard V AR breaks down and produces a degenerate graph (i.e., `'marks for correctly identifying nodes as predator (source) and prey (sink). Results and example MTS simulation is reported in Table 2 and Figure 1 in Appendix D.2; BSG at all forecast horizons outperforms baselines for T = 50 and T = 200. For T = 1000, BSG performs consistently well for identifying source nodes, but has lower accuracy for identifying sink nodes, likely due to long-range dependence for a longer MTS. GVAR-Closeness has marginally higher accuracy (+0.014) for identifying predators compared to BSG (h = 1) but very low accuracy (0.554) for identifying prey. Meanwhile, standard VAR after FDR adjustment produces degenerate graphs. On average, BSG still performs well on between both source and sink node identification; in practice, it may be useful to first difference MTS with higher-order autocorrelation.

5 BSG FOR UNDERSTANDING REAL-WORLD SYSTEMS

Non-Linear Dynamic Systems Recent works have also focused on dynamic systems with non-linear or higher-order temporal relationships. A prime example is the Lokta-Volterra predator-prey model Bacaër [2011]. Four parameters \{α, β, γ, δ\} correspond to prey → itself, predator → prey, predator → itself, and prey → predator interaction strengths. We generate 5 MTS replicates using the same parameter specifications \{(1,2,0,2,1,1,0,05)\} as Marcinkevičs and Vogt [2021], with T = {50, 200, 1000}. We compare BSG influence/vulnerability scores vs. benchmarks for correctly identifying nodes as predator (source) and prey (sink). Results and example MTS simulation is

![Figure 6: BSG for Kincade Fire, h=12 hours ahead. Red indicates source and blue indicates sink nodes. Arrow width is prop. to BSG edge weight. See Figure 4 in Appendix E for 95% HPDI of spillovers.](Image 322x302 to 526x506)

Inferring Spillovers from California Wildfires. The Kincade Fire was the largest California wildfire in 2019, burning a total of 77,758 acres. It originated in Sonoma County and dangerous PM10/PM2.5 particles in the air posed a serious public health risk spillover for nearby counties with high population density. We use BSG to investigate spillovers and rank at-risk nodes (counties) as measured by hourly PM2.5 particle concentrations from Oct 22-Nov 7. We have a reasonable ground-truth for underlying network structure with Sonoma County as the single source node. Therefore, any strong BSG edges detected between Sonoma and non-adjacent counties, or two counties that does not include Sonoma, can be considered indirect spillover effects.

Data Description. Using public data from EPA (Environ-
mental Protection Agency), hourly PM 2.5 concentrations are extracted for 10 counties within 50 miles of Sonoma County in Northern California; Yolo, Sutter, and Lake counties had no data available. See Figure 2 in Appendix E for MTS plot. No visible trend or seasonality effects are observed; autocorrelation plots show evidence of long memory for some counties and we also observe prominent spikes, particularly initially in Sonoma and later with time lag in other counties. To ensure stationarity, we proceed with the first order difference of the MTS.

**Quantifying Spillover & At-risk Nodes.** In Figure 6, we illustrate all BSG edges \( h = 12 \) greater than the 80th percentile in magnitude for simplicity, with arrow width proportional to edge weights. The top source node Sonoma (by BSG influence score) is shaded in red, and top sink nodes (by vulnerability score) is shaded in blue. The BSG neatly captures the Kincade Fire in that Sonoma has the majority of all outgoing edges, while further away, non-adjacent counties (sink nodes) such as Colusa and Alameda have strong spillovers both directly from Sonoma and indirectly via other counties as well. In particular, note the cycle from Sonoma \( \rightarrow \) Contra Costa \( \leftrightarrow \) Alameda where sink nodes also interact and amplify spillover effects. We can further quantify downstream spillovers via BSG edge weights for counties to the southeast of Sonoma; see Figure 7 for county map with spillovers. Roughly 10% of FEVD for each county can be attributed to changes in Sonoma’s PM 2.5 concentration. One possible explanation is downsloping winds from the north [Mass and Ovens 2019], which is particularly concerning due to the far higher population density of impacted counties. Two other notable indirect spillovers not involving Sonoma include those from San Mateo to Contra Costa (12.3%) and Alameda (9.3%).

BSG influence and vulnerability scores for each county are reported in Figure 3 in Appendix E. Sonoma County is the most influential node, accounting for more than 40.9% of total spillover effect across all 10 counties on average, with the 95% HPDI as (17.9%, 62.7%). BSG accurately identifies the origin of the Kincade Fire while also showing Sonoma itself is the least vulnerable node. Locations most at risk to the fire, by vulnerability score, are Alameda and Contra Costa followed by San Francisco, Solano, and Colusa. None of these 5 counties are adjacent to Sonoma; they incur higher risk via spillovers from intermediary Marin and Napa counties, accumulated over multiple time periods. These risk quantifications from BSG have practical implications for policies with respect to wildfire relief and public health. For example, although FEMA allocated nearly 60 million dollars in federal relief [FEM 2019], the funds were strictly designated for Sonoma County. Meanwhile, BSG as an exploratory tool clearly identifies much broader spillovers and at-risk counties.

**6 DISCUSSION**

BSG is a novel framework for modeling temporal interactions and identifying important nodes within a dynamic system based on a single realized multivariate time series. BSG combines interpretable forecast error based network measures with uncertainty quantification via sampling from posterior graph distribution, and demonstrates robust performance across various graph specifications and error dependency structures. The hyperparameter \( h \) allows for custom learning of both short and long-term temporal relationships, including indirect spillovers, which are better suited for understanding how real-world systems evolve over time. Careful choice of horizon \( h \) can help model equilibrium state of systems and optimize proper ranking of sink and source nodes.

A key application of BSG could be for analyzing spillover impact in response to new regulations and economic policies. For example, consider when a significant event occurs in a particular city, e.g., a new tax policy is passed or a local manufacturer is shut-down and off-shored. Prior works have utilized impulse response functions to analyze policy interventions [Sims 1980; Ericsson et al., 1998; Lütkepohl 2005]; we propose leveraging BSG to examine and quantify both positive and negative externalities (spillover effects) in terms of employment statistics, traffic congestion, local rent, wages, etc., for neighboring cities or counties. Inference via BSG can be for both short-term and long-term impact based on forecast horizon, and used to inform both the public and policymakers.
Another potential BSG application is in time series analysis of fMRI data in healthcare and medicine [Penny et al., 2005]; for example, we can examine individual brain fMRI time series where each component are atlas based regions of interest, i.e. aggregated behavior from sets of voxels, which represent smaller unit regions in the brain. The time series could measure brain activity in response to some stimuli or treatment, and a BSG can illustrate cumulative effect of temporal interactions between different brain regions over time. The novel BSG network measures (influence score, vulnerability score) can also pinpoint critical components of brain connectivity, analogous to sink or source nodes.

Future work can dive deep into applying BSG for some of these datasets aforementioned, as well as extending the BSG framework for Bayesian networks with time-varying coefficients [Kowal et al., 2019] or latent state-space representations.

Acknowledgements

The authors gratefully acknowledge financial support from the National Science Foundation Awards 1934985, 1940124, 1940276, and 2114143.

References


