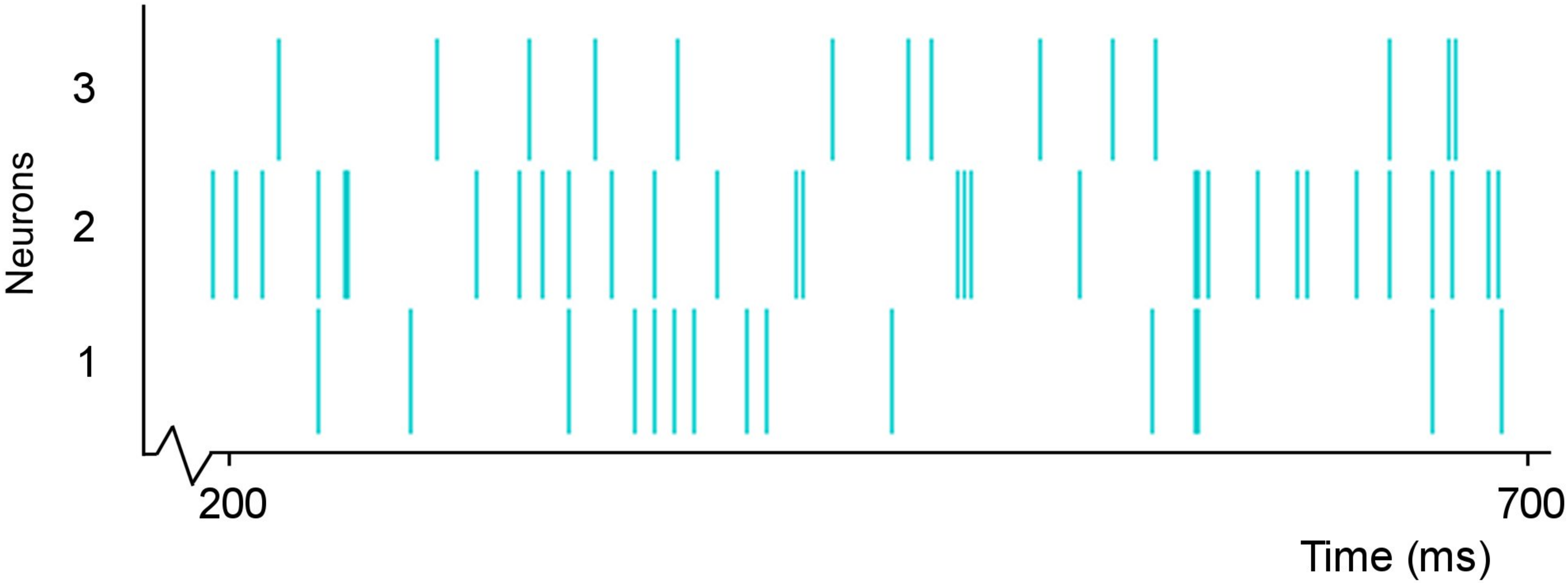
Context-dependent network estimation in high-dimensional auto-regressive models

Lili Zheng
University of Wisconsin-Madison

Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

Spike Train Data: Network among Neurons



Network structure is embedded in spike train data

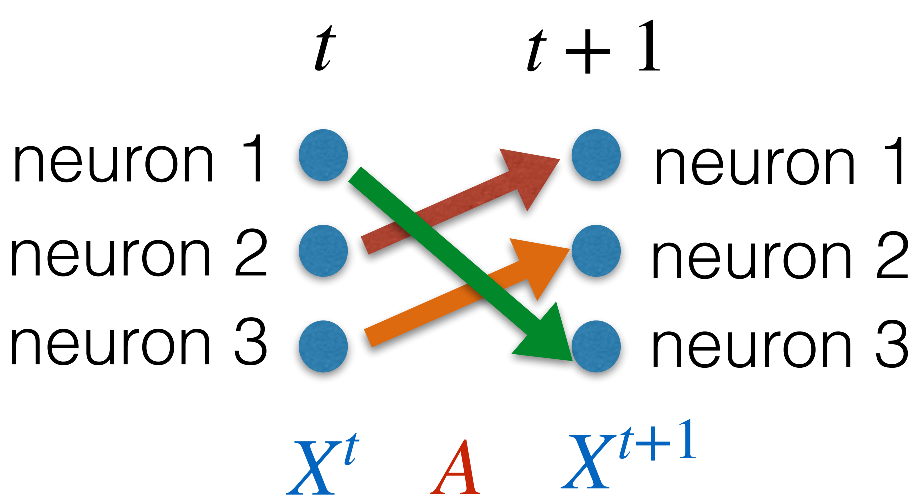
https://mathematica.stackexchange.com/questions/55845/spike-train-representation

Social Media Network

https://sloanreview.mit.edu/article/maximizing-the-impact-of-enterprise-social-media/

Modeling Framework

- Observation $\{X^t \in \mathbb{R}^M\}_{t=0}^T$, M nodes



$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0.5 \\ -0.8 & 0 & 0 \end{pmatrix} \in \mathbb{R}^{M\times M}$$

$A_{m,m'}$: influence of node m' upon node m

$$X_m^{t+1} \,|\, X^t \sim p(\langle A_m, X^t \rangle), m = 1, \ldots, M\,.$$

- Goal: estimate $A \in \mathbb{R}^{M\times M}$ from $\{X^t\}_{t=0}^T$, with certain $p(\,\cdot\,)$.

Prior work: Bernoulli autoregressive models

Observation: $\{X^t \in \mathbb{R}^M\}_{t=0}^T$

Network parameter: $A \in \mathbb{R}^{M \times M}$

Event intensity: $\mu_m^{t+1} = \langle A_m, X^t \rangle$

$$X_m^{t+1} \mid X^t \sim \text{Bernoulli}\left(\frac{e^{\mu_m^{t+1}}}{1 + e^{\mu_m^{t+1}}} \right)$$

Events and network are not associated with context or "marks"

Hall et al., 2016; Mark et al., 2018

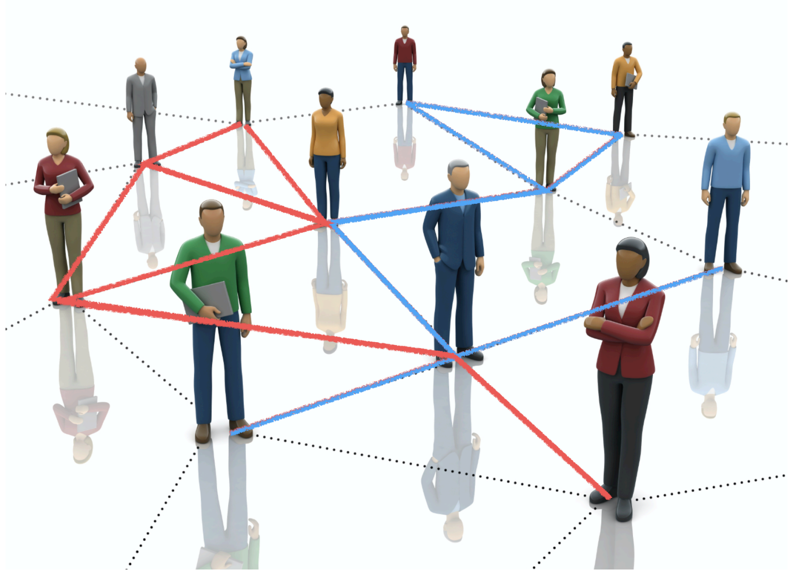
Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

Context-dependent Network Estimation (Zheng, Raskutti, Willett and Mark, 2020)
Motivation

- Network depends on the categories of events

- Examples:
  - Social media networks: family-focused, work-focused, and political, etc.



  - News media networks: business-related, sports-related

Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

Categorical Events: Multinomial Model

- Observation: $\{X^t \in \mathbb{R}^{M \times K}\}_{t=0}^{T}$

M= number of nodes: high-dimensional
K=number of categories: low-dimensional

e.g., K=2, left-leaning and right-leaning

$$X_m^t = \begin{cases} e_k, & \text{if node m has event type k at time t} \\ 0_K, & \text{if node m has no event at time t} \end{cases}$$

$$\mathbb{P}(X_m^{t+1} = e_k \,|\, X^t) = \frac{e^{\mu_{mk}^{t+1}}}{1 + \sum_{k'=1}^{K} e^{\mu_{mk'}^{t+1}}}, 1 \le k \le K$$

$$\mathbb{P}(X_m^{t+1} = 0_K \,|\, X^t) = \frac{1}{1 + \sum_{k'=1}^{K} e^{\mu_{mk'}^{t+1}}}$$

Event Intensity: $\mu_{mk}^{t+1} = \langle A_{mk}^{MN}, X^t \rangle + \nu_{mk}^{MN}$

Categorical Events: Multinomial Model

$$\mu_{mk}^{t+1} = \langle A_{mk}^{MN}, X^t \rangle + \nu_{mk}^{MN}$$

- Network parameter: $A^{MN} \in \mathbb{R}^{M \times K \times M \times K}$

$A_{mkm'k'}^{MN}$= (absolute) influence of (event at node m' in category k') on (event at node m in category k).

$A_{1,1,2,2}^{MN}$: the influence of the second twitter user's right-leaning tweets upon the first twitter user's left-leaning tweets

Categorical Events: Multinomial Model

Estimator

$$\widehat{A}^{MN} = \arg\min_{A \in \mathbb{R}^{M \times K \times M \times K}} L^{MN}(A) + \lambda \|A\|_R$$

Negative log-likelihood function:

$$L^{MN}(A) = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{m=1}^{M} \left[\log(1 + \sum_{k=1}^{K} e^{\mu_m^{t+1}}) - \sum_{k=1}^{K} \mu_{mk}^{t+1} X_{mk}^{t+1} \right]$$

$$\|A\|_R = \sum_{m,m'} \|A_{m,:,m',:}\|_F$$

This group-sparse regularizer promotes shared network structure across categories

Events with Mixed Membership in Multiple Categories: Logistic-normal Model

- News articles with mixed membership in multiple topics

- Observe $\{X^t \in \mathbb{R}^{M\times K}\}_{t=0}^{T}$

$$X_m^{t+1} \mid X^t = \begin{cases} Z_m^{t+1}, & \text{with probability } q_m^{t+1} (\text{event occurs}), \\ 0_K, & \text{with probability } 1 - q_m^{t+1} (\text{no event occurs}), \end{cases}$$

- $Z_m^{t+1} \in \triangle^{K-1}$: mixed membership/composition in K categories

Logistic-normal Model (Modeling Z^t)

Logisitic-normal distribution

Relative intensity

Noise $\epsilon_m^{t+1} \sim \mathcal{N}(0,\Sigma)$

$$\log \frac{Z_{mk}^{t+1}}{Z_{mK}^{t+1}} \,|\, X^t = \mu_{mk}^{t+1} + \epsilon_{mk}^{t+1}, \quad 1 \le k \le K-1$$

Baseline

$$\mu_{mk}^{t+1} = \langle A_{mk}^{LN}, X^t \rangle + \nu_{mk}^{LN}$$

Network parameter

Logistic-normal Model (Modeling Z^{t})

$$\mu_{mk}^{t+1} = \left\langle A_{mk}^{LN}, X^t \right\rangle + \nu_{mk}^{LN}$$

- Network parameter: $A^{LN} \in \mathbb{R}^{M \times (K-1) \times M \times K}$

  $A_{mkm'k'}^{LN}$ = relative influence of (event at node m' in category k')
  on (event at node m in category k
  compared to node m in category K)

Logistic-normal Model (Modeling q^t)

- q^t depends on past events through logistic link

$$\mathbb{P}(X_m^{t+1} \neq 0) = q_m^{t+1} = \frac{\exp\{\langle B_m^{Bern}, X^t\rangle + \eta_m^{Bern}\}}{1 + \exp\{\langle B_m^{Bern}, X^t\rangle + \eta_m^{Bern}\}}$$

- Network parameter: $B^{Bern} \in \mathbb{R}^{M \times M \times K}$

$B_{mm'k'}^{Bern}$ = overall influence of (event at node m' in category k') on (event at node m).

Logistic-normal Model
Estimator

Loss functions for A^{LN} and B^{Bern}:

$$L^{LN}(A) = \frac{1}{2T} \sum_{m} \sum_{t \in \mathcal{T}_m} \| Y_m^{t+1} - \mu_m^{t+1}(A_m) \|_2^2$$

$$\mathcal{T}_m = \{t : X_m^{t+1} \neq 0\} \qquad Y_m^{t+1} \in \mathbb{R}^{K-1}, Y_{mk}^{t+1} = \log \frac{X_{mk}^{t+1}}{X_{mK}^{t+1}}$$

$$L^{Bern}(B) = \frac{1}{T} \sum_{t,m} \log(1 + e^{\langle B_m, X^t \rangle + \eta_m^{Bern}}) - (\langle B_m, X^t \rangle + \eta_m^{Bern}) 1_{\{X_m^{t+1} \neq 0\}}$$

Logistic-normal Model
Estimator

Joint Estimation for A^{LN} and B^{Bern}:

$0 \leq \alpha \leq 1$
adjust for unknown
noise level Σ

Estimator:

$$(\widehat{A}^{LN}, \widehat{B}^{Bern}) = \min_{A,B} \alpha L^{LN}(A) + (1-\alpha) L^{Bern}(B) + \lambda R_{\alpha}(A, B)$$

$$R_{\alpha}(A,B) = \sum_{m,m'} \left(\alpha \|A_{m,:,m',:}\|_F^2 + (1-\alpha) \|B_{m,m',:}\|_2^2 \right)^{\frac{1}{2}}$$

promotes shared network structure between A^{LN} and $\mathrm{B}^{\mathrm{Bern}}$

Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

Statistical Guarantees

Assumptions:

- Group sparsity:
  - $S_m := \{m' : \|A_{m,:,m',:}\|_F > 0\}$, $\rho_m = |S_m|$
  - $\rho := \max_{1 \le m \le M} \rho_m$, $s := \sum_{m=1}^{M} \rho_m$
- Boundedness:

$$\|A\|_{\infty,\infty,1,\infty} = \max_{m,k} \sum_{m'} \max_{k'} |A_{mkm'k'}| \le C$$

for $A = A^{MN}, A^{LN}, (A^{LN}, B^{Bern})$

Number of categories K, Offset parameters, covariance matrix $\Sigma \in \mathbb{R}^{K \times K}$ all bounded.

Statistical Guarantees

Multinomial Model and Estimator

If $\lambda \asymp \sqrt{\frac{\log M}{T}}$, $T \gtrsim \rho^2 \log M$, with high probability,

$$\|\widehat{A}^{MN} - A^{MN}\|_F^2 \lesssim \frac{s \log M}{T}$$

$$\|\widehat{A}^{MN} - A^{MN}\|_R \lesssim s\sqrt{\frac{\log M}{T}}$$

ρ = maximum in-degree of any node

s= sparsity, number of network edges

M = number of network nodes

$\|\cdot\|_{\mathrm{R}}$: group $\ell_{1,2}$ norm

Statistical Guarantees
Logistic-normal Model and Estimator

If $\lambda \asymp C(\alpha)\sqrt{\frac{\log M}{T}}$, $T \gtrsim \rho^2 \log M$, with high probability,

$$\alpha \| \widehat{A}^{LN} - A^{LN} \|_F^2 + (1-\alpha) \| \widehat{B}^{Bern} - B^{Bern} \|_F^2 \lesssim \frac{s \log M}{T}$$

$$R_\alpha(\widehat{A}^{LN} - A^{LN}, \widehat{B}^{Bern} - B^{Bern}) \lesssim s\sqrt{\frac{\log M}{T}}$$

ρ = maximum in-degree of any node
s = sparsity, number of network edges

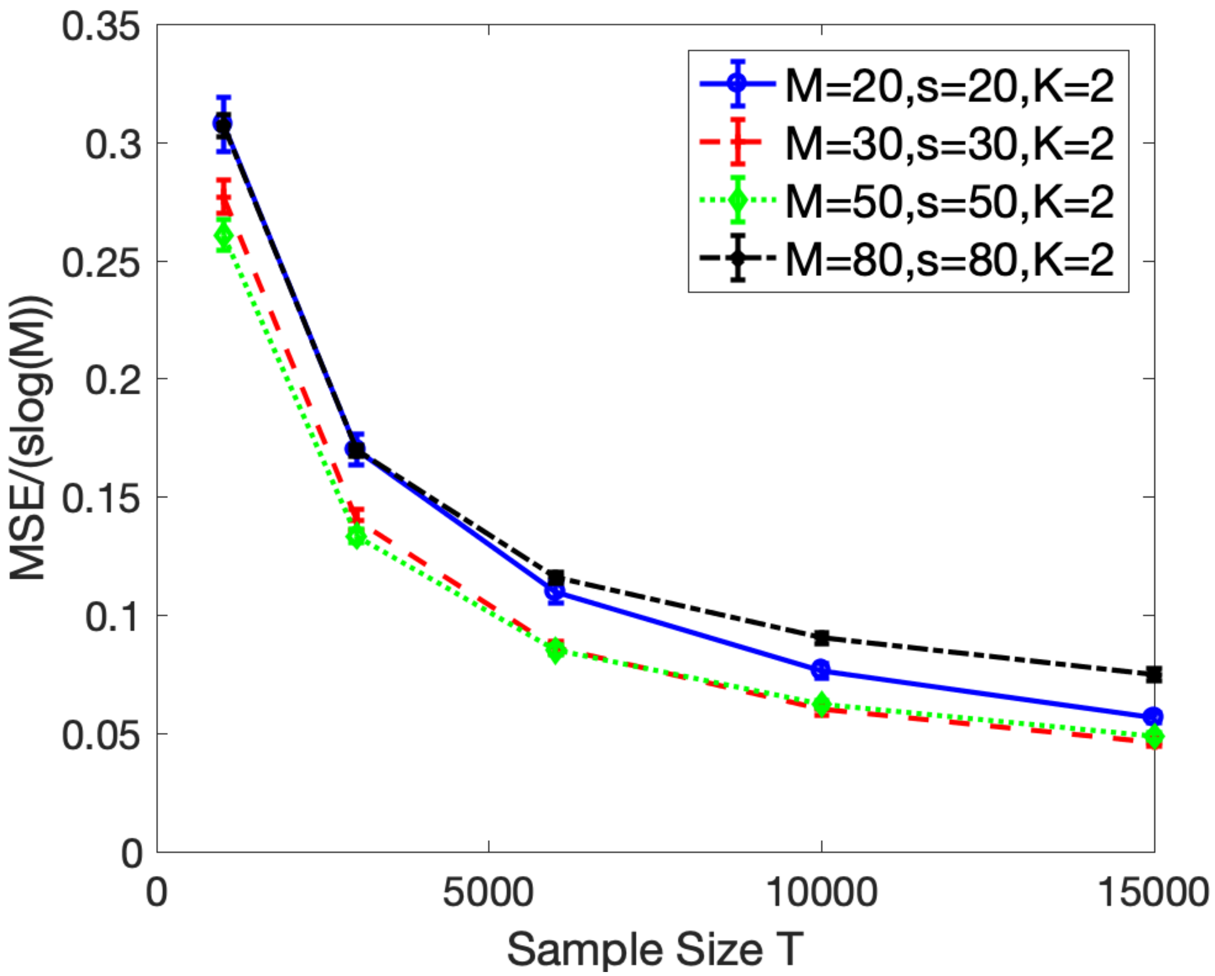
M= number of network nodes

Rate Plots

How does mean squared error scale with s, M, T?

theoretical bound: $\dfrac{s \log M}{T}$

- $M \in \{20, 30, 50, 80\}$
- $K = 2$
- $T \in [1000, 150000]$
- $\rho \in \{1, 2, 3\} \Rightarrow s = \rho M$
- $\Sigma = I_K$
- non-zeros element of network parameters sampled from U(-2,2)
- $\lambda = C\sqrt{\dfrac{\log M}{T}}$ where C is chosen by cross-validation



Multinomial: Mean squared error/(s log M)

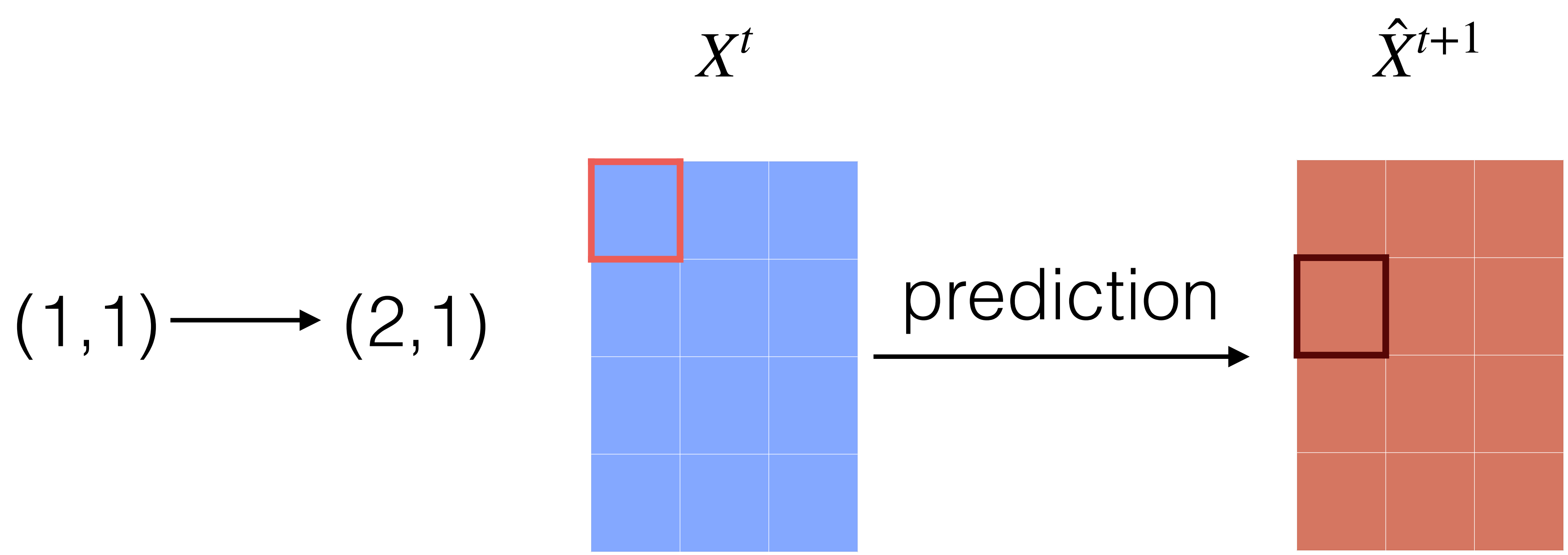
Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

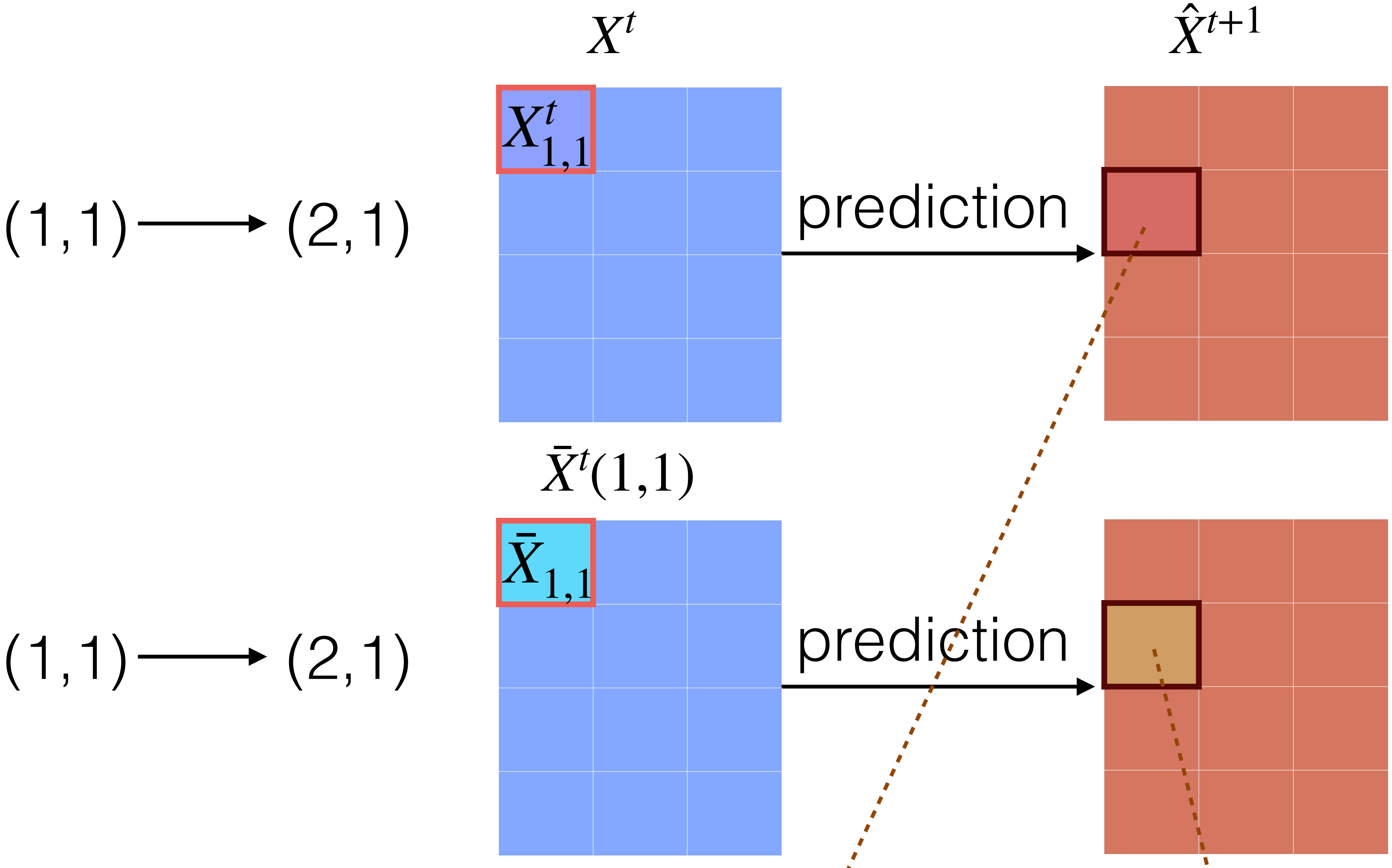
Shared Interpretation: Variable Importance Network

- Network parameters have different interpretations under the two models
- Different parameterization can lead to drastically different network estimates

Variable importance network:
prediction-based, non-parametric



Variable Importance Network



$$\frac{1}{T}\sum_{t=0}^{T-1} \operatorname{sgn}(X^t_{1,1} - \bar{X}_{1,1})\Big(\mathbb{E}(X^{t+1}_{2,1} \,|\, X^t) - \mathbb{E}(X^{t+1}_{2,1} \,|\, \bar{X}^t(1,1))\Big)$$

Variable Importance Network

Variable importance network parameter $V \in \mathbb{R}^{M \times K \times M \times K}$

$$V_{m_1,k_1,m_2.k_2} = \frac{1}{T}\sum_{t=0}^{T-1} \mathrm{sgn}(X^t_{m_2,k_2} - \bar{X}_{m_2,k_2})\Big(\mathbb{E}(X^{t+1}_{m_1,k_1} \,|\, X^t) - \mathbb{E}(X^{t+1}_{m_1,k_1} \,|\, \bar{X}^t(m_2,k_2))\Big)$$

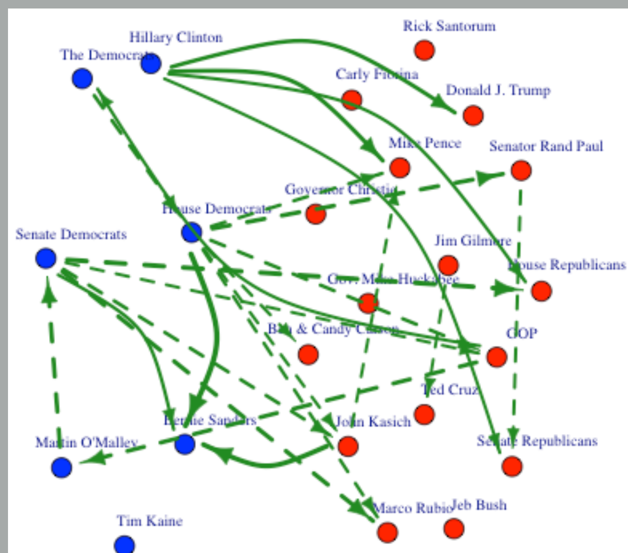
substituted by one-step ahead prediction of the fitted model

estimated version $\widehat{V}$

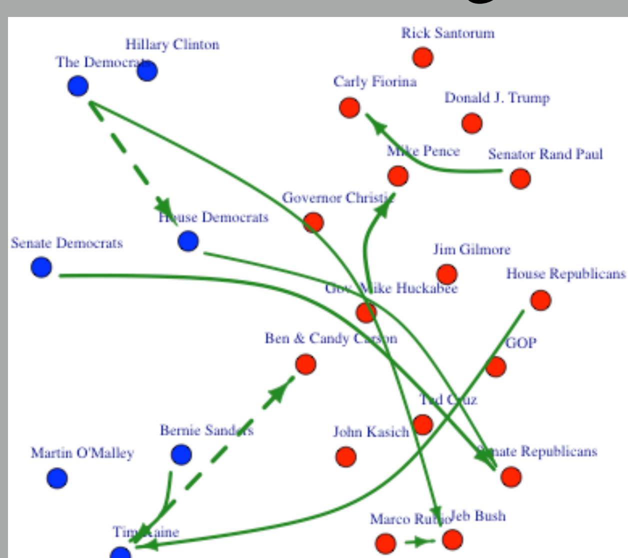
Under proposed models: $\| \widehat{V} - V\|_F^2 \lesssim \dfrac{\rho^2 s \log M}{T}$

Variable Importance Network (on tweets example)

parametric networks

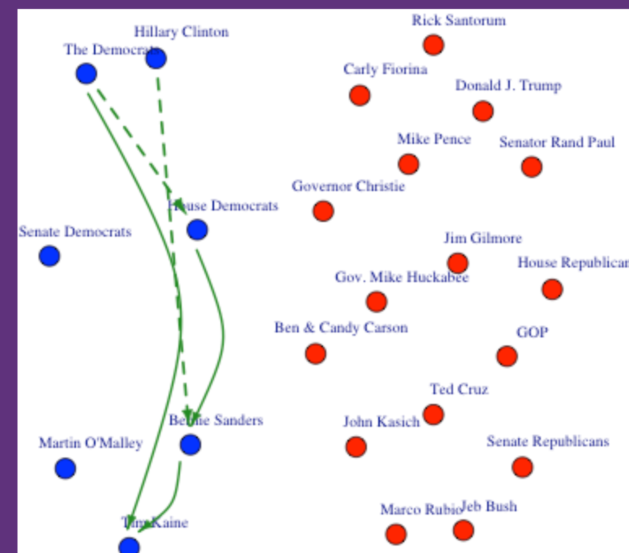


MN: $\text{left} \rightarrow \frac{\text{left}}{\text{right}}$

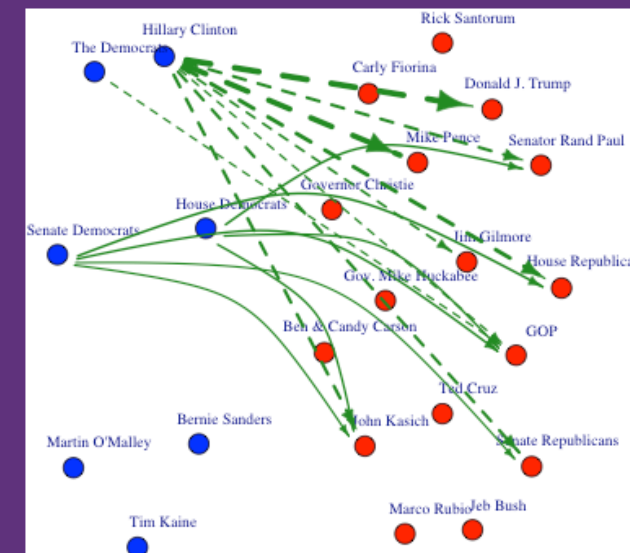


LN: $\text{left} \rightarrow \frac{\text{left}}{\text{right}}$

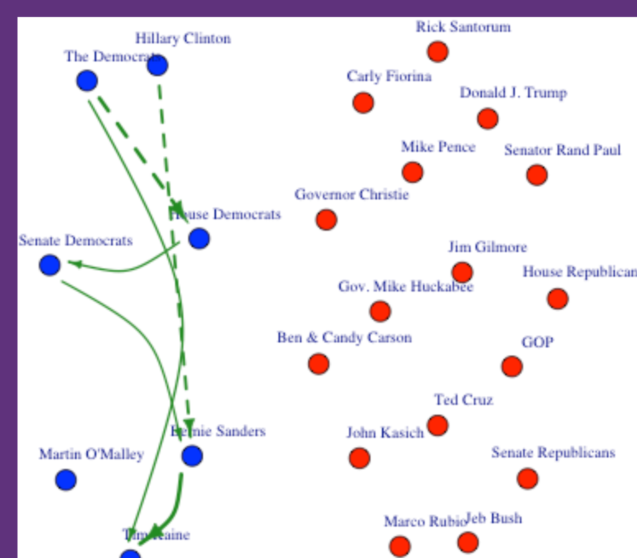
variable importance networks



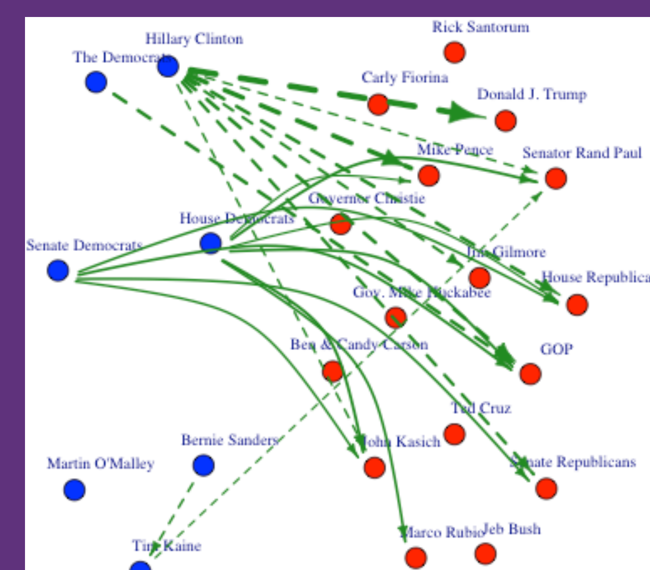
MN: $\text{left} \rightarrow \text{left}$



MN: $\text{left} \rightarrow \text{right}$



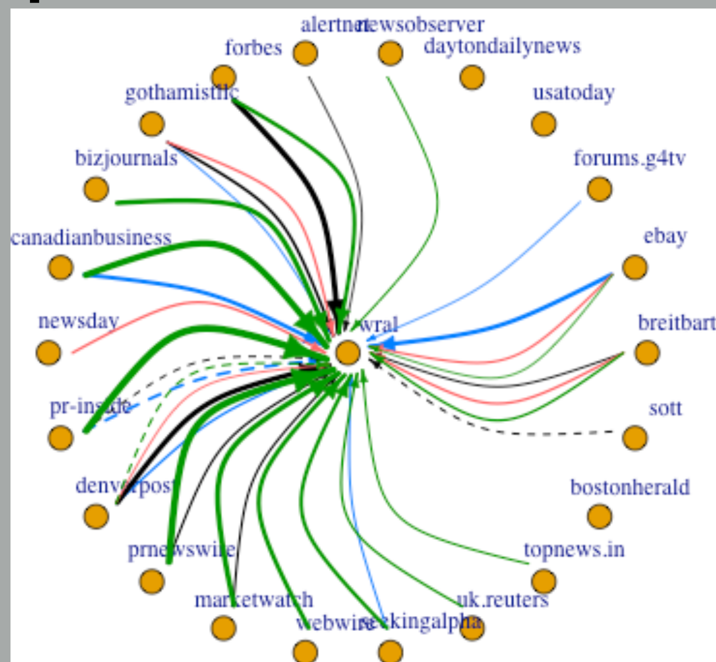
LN: $\text{left} \rightarrow \text{left}$



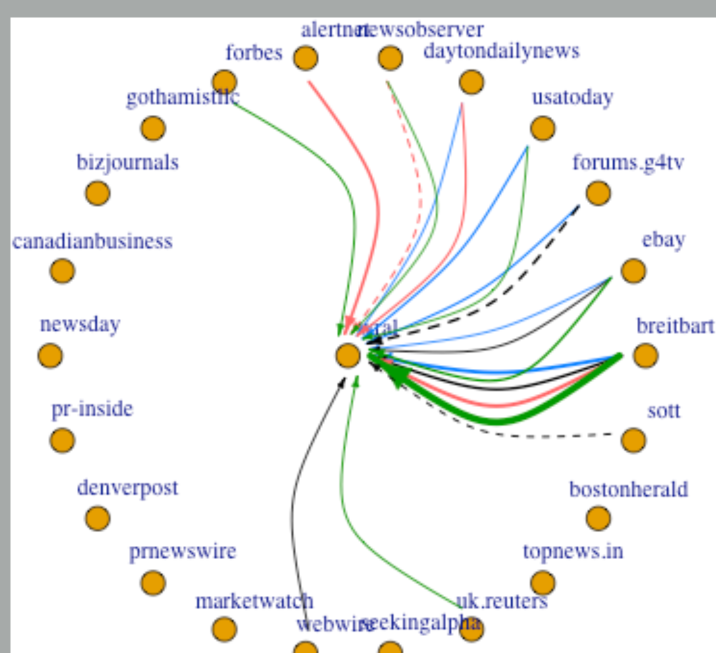
LN: $\text{left} \rightarrow \text{right}$

Variable Importance Network (on MemeTracker example)

parametric networks

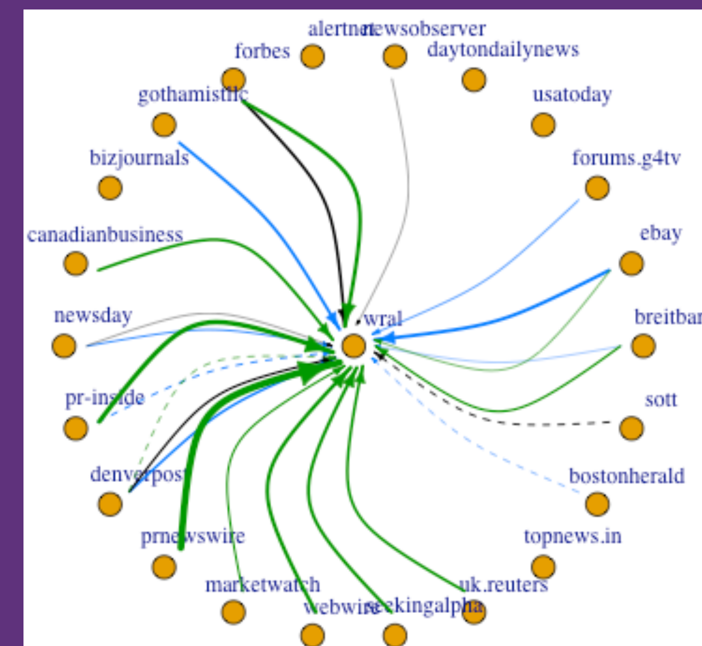


MN

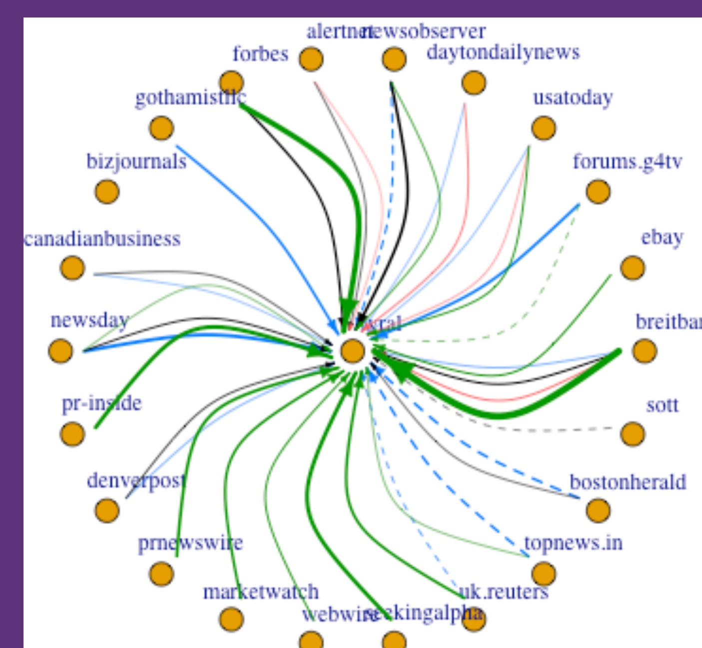


LN

variable importance networks



MN



LN

Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

Choice between the Two Approaches?

- Real world situation: we have **compositional vectors for all events**
- Some nodes' events exhibit mixed membership while others' events mostly fall in one category

Contaminated mixture model

Understand the performance of the two methods and develop a better approach

Contaminated Mixture Model

Mixture

- Non-overlapping sets $\mathcal{N}_1 \cup \mathcal{N}_2 = [M]$,
- $X_m^{t+1} \mid X^t \sim \begin{cases} \text{MN model,} & m \in \mathcal{N}_1, \\ \text{LN model,} & m \in \mathcal{N}_2 . \end{cases}$

Contamination

Observe $\{\tilde{X}^t\}_{t=0}^{T}$:

$$\tilde{X}_m^t = X_m^t \text{ if } m \in \mathcal{N}_2 \text{ or } X_m^t = 0$$

If $m \in \mathcal{N}_1$ and $X_m^t = e_k$,

$$\log \frac{\tilde{X}_{m,1:(K-1)}^t}{\tilde{X}_{mK}^t} \sim \mathcal{N}(\mu_k, \sigma^2 I_{K-1}) .$$

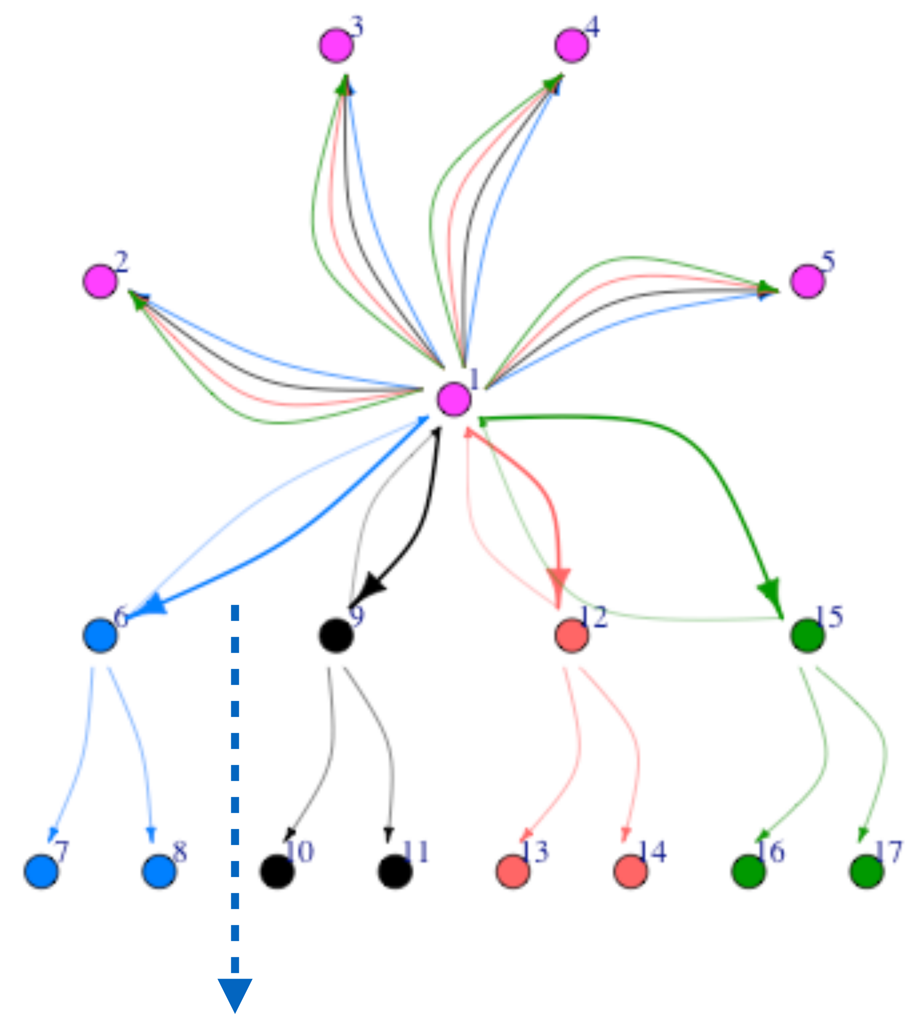
Synthetical Example

- Blue/black/red/green nodes(single category):

focus on one of the first 4 categories
multinomial model

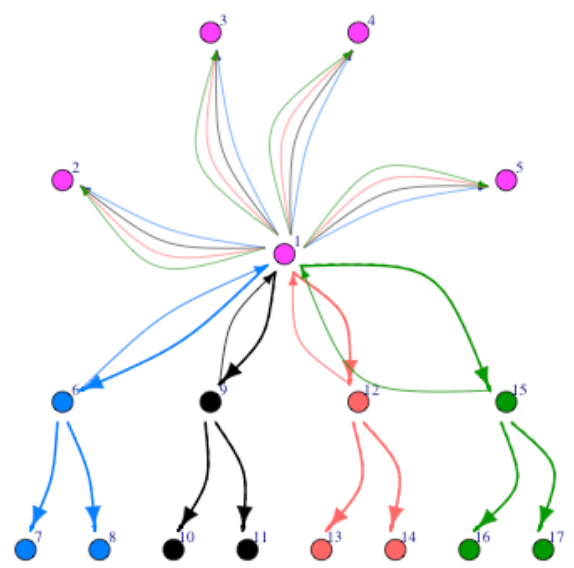
- Purple nodes (mixed membership):

uniform preference,
in the first 4 categories
logistic-normal model

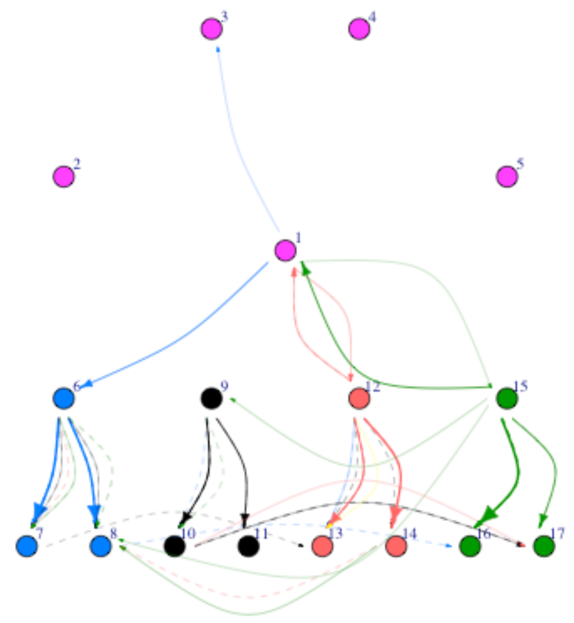


Influence of node 1 on node 6 in blue category

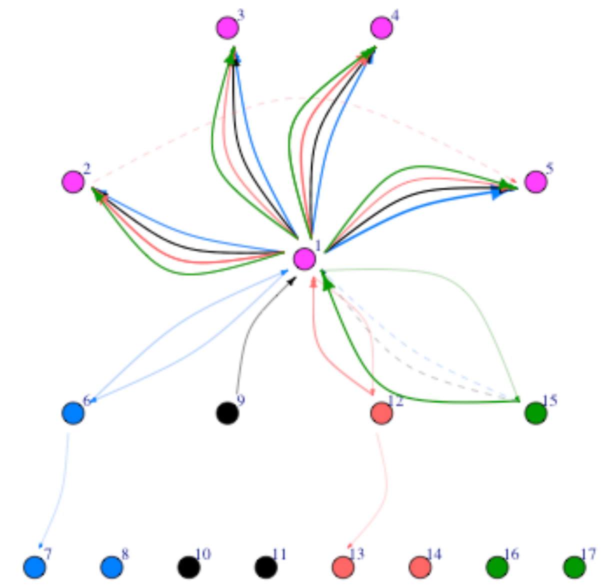
Synthetical Example (Estimates Comparison)



True network



Multinomial network estimate



Logistic-normal network estimate

More accurate for the first type of nodes

More accurate for the second type of nodes

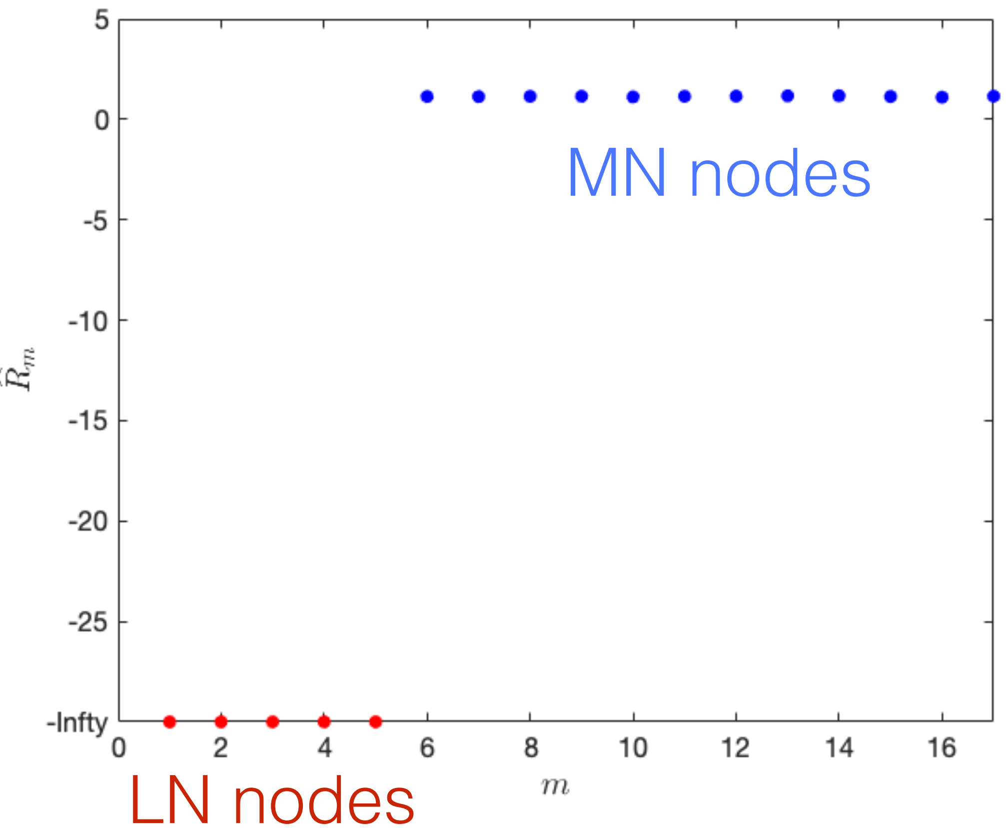
Testing Procedure

Log likelihood-ratio for node m:

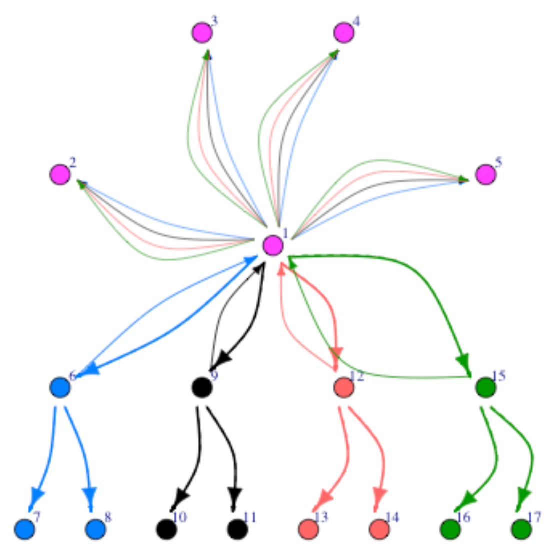
$$R_m = \frac{1}{T}\sum_{t=0}^{T-1} \ell^{LN}(A_m, \nu_m, \Sigma; X^t, \tilde{X}_m^{t+1}) - \ell^{MN}(A_m, \nu_m, a, \sigma^2; X^t, \tilde{X}_m^{t+1}),$$

Estimate parameters under each model $\rightarrow$ $\hat{R}_m$

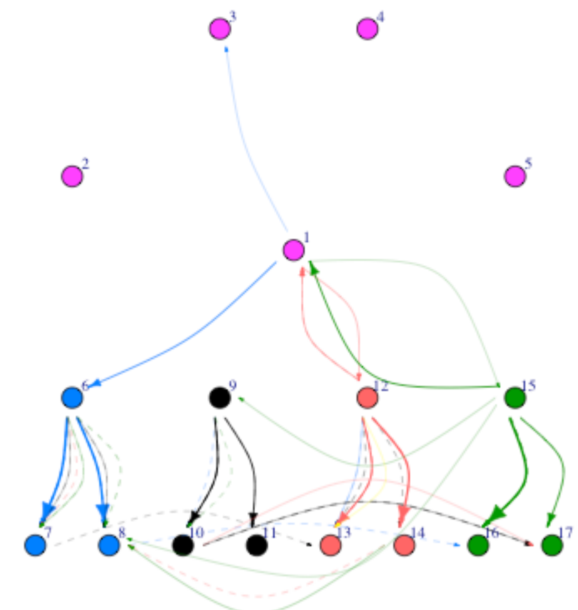
Synthetic example



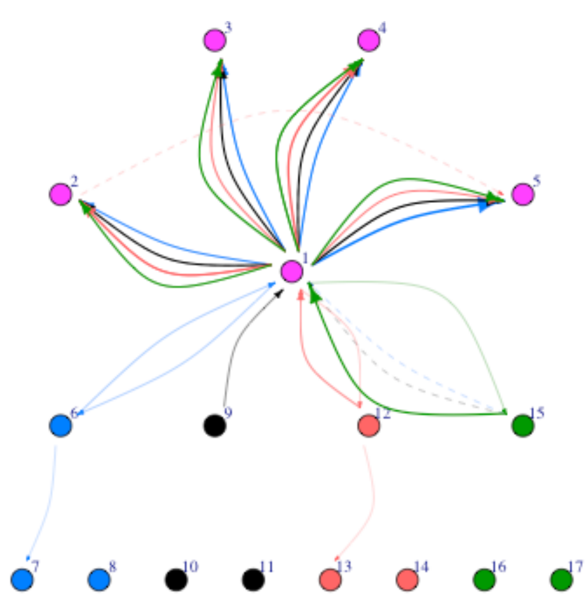
Synthetical Example (Estimates Comparison)



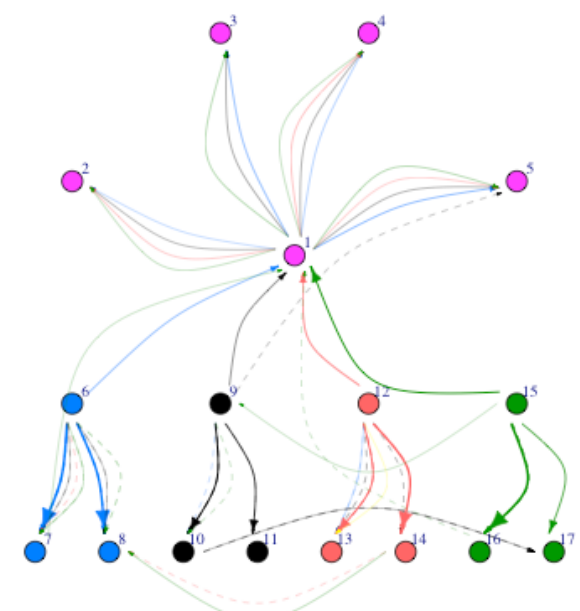
True network



Multinomial network estimate



Logistic-normal network estimate



Mixture network estimate

Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

MemeTracker Data

How does the influence among online media sources vary as a function of topics of news articles?

- Data set: the MemeTracker data set
  - News stories and posts of over 1 million online media sources.
  - Phrases that have variants occur frequently online are recorded for each post (August 2008 to April 2009)
  - Filter out media sources based on frequency and language

- Run topic modeling (LDA) with 5 topics ($K = 5$): Sports, International affairs, Lifestyle, Finance, Health.

http://www.memetracker.org/data.html

MemeTracker Data (Topic keywords)

Topics	Keywords
Sports	time, people, lot, thing, game, way, team, work, player, year
International Affairs	people, country, government, time, united_states, state, law, issue, case, work
Lifestyle	life, people, man, family, love, water, woman, world, story, music
Finance	market, company, business, economy, customer, time, service, industry, bank, product
Health	child, patient, food, health, people, drug, hospital, information, research, risk

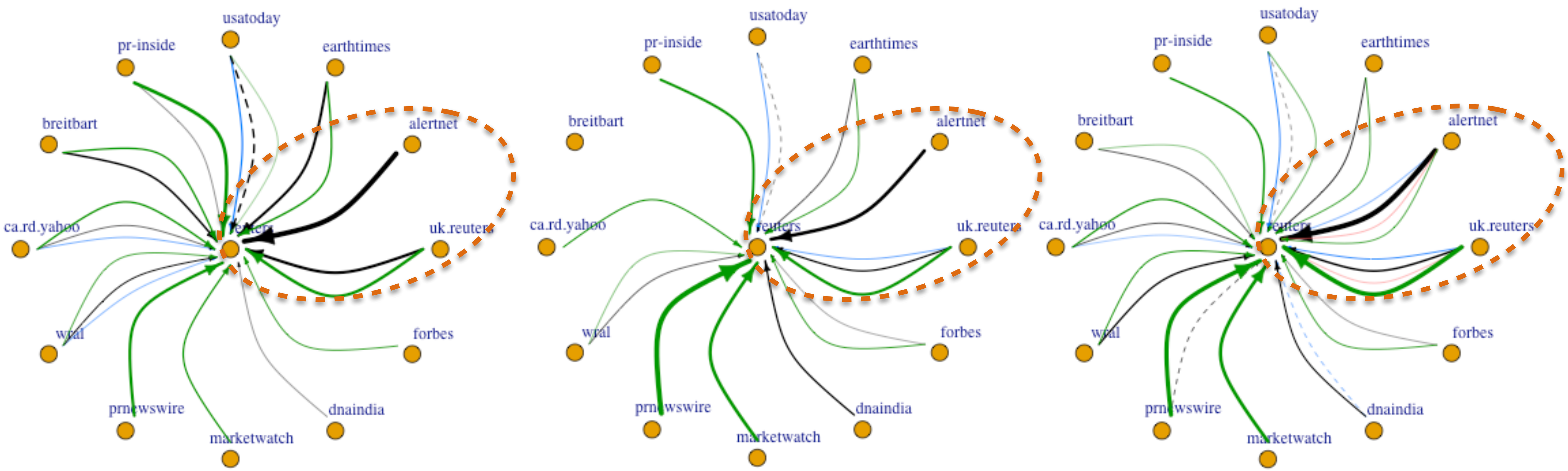
- Choose 58 media sources ($M = 58$) (15 media sources with highest topic weights for each of the first 4 topics)
- 1-hour discretization: 5807 time intervals ($T = 5807$)

MemeTracker Data (Supporting Evidence for Edges)

- Phrase cluster data
- Each phrase cluster forms a cascade

- Supporting evidence for the edge from $\mathbf{m}'$ to $\mathbf{m}$:
    - If media source $\mathbf{m}$ posts a phrase cluster within an hour after $\mathbf{m}'$: influence-involved phrase clusters from $\mathbf{m}'$ to $\mathbf{m}$
    - Number and topics

MemeTracker Data (Network Estimates)



Multinomial network Mixture network Logistic-normal network

blue edge: Sports; black edge: International affairs;
red edge: Lifestyle; green edge: Finance; yellow edge: Health

MemeTracker Data (Validate Edges)

Influence upon *reuters.com*

Media Source	Total Number of Posted Phrase Clusters	Influence-involved Phrase Clusters	Percent	Rank
alertnet	2552	758	29.70%	1
uk.reuters	4998	875	17.51%	2

alertnet ⟶ reuters

Sports	International Affairs	Lifestyle	Finance	Health
0.1380	**0.5647**	0.0458	0.1370	0.1146

Multinomial: International Affairs

Mixture: International Affairs

Logistic-normal: Sports, International Affairs, Lifestyle, Finance

MemeTracker Data (Validate Edges)

Influence upon *reuters.com*

Media Source	Total Number of Posted Phrase Clusters	Influence-involved Phrase Clusters	Percent	Rank
alertnet	2552	758	29.70%	1
uk.reuters	4998	875	17.51%	2

uk.reuters ⟶ reuters

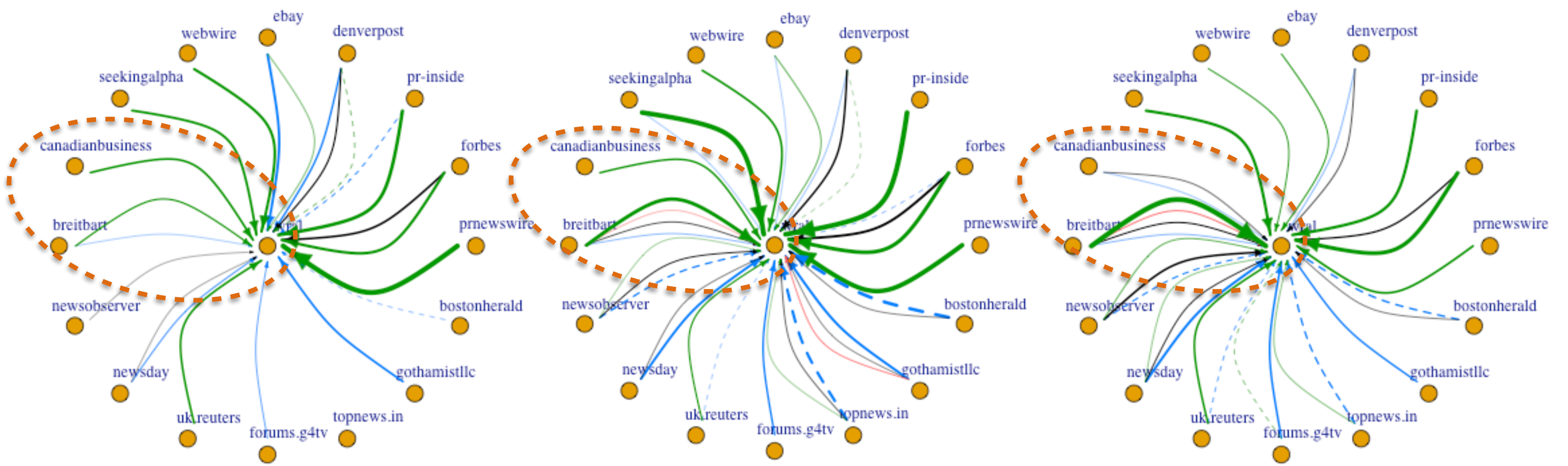
Sports	International Affairs	Lifestyle	Finance	Health
0.2359	**0.4634**	**0.1274**	**0.1384**	0.0348

Logistic-normal: Sports, International Affairs, Lifestyle, Finance

Mixture: Sports, International Affairs, Finance

Multinomial: International Affairs, Finance

MemeTracker Data (Network Estimates)



Multinomial network Mixture network Logistic-normal network

blue edge: Sports; black edge: International affairs;
red edge: Lifestyle; green edge: Finance; yellow edge: Health

MemeTracker Data (Validate Edges)

Influence upon *wral.com*

Media Sources	Total Number of Posted Phrase Clusters	Influence-involved Phrase Clusters	Percent	Rank
canadianbusiness	2339	252	10.77%	3
breitbart	19279	1408	7.30%	4

canadianbusiness → wral

Sports	International Affairs	Lifestyle	Finance	Health
0.1062	**0.2153**	0.0144	**0.6255**	0.0387

Multinomial: Finance

Mixture: Finance

Logistic-normal: Sports, International Affairs

MemeTracker Data (Validate Edges)

Influence upon *wral.com*

Media Sources	Total Number of Posted Phrase Clusters	Influence-involved Phrase Clusters	Percent	Rank
canadianbusiness	2339	252	10.77%	3
breitbart	19279	1408	7.30%	4

breitbart ⟶ wral

Sports	International Affairs	Lifestyle	Finance	Health
0.2412	**0.3392**	**0.1030**	**0.2744**	0.0423

Logistic-normal: Sports, International Affairs

Mixture: Sports, International Affairs, Lifestyle, Finance

Multinomial: Sports, Finance

MemeTracker Data (Validate Hypothesis)

The **logistic-normal** approach better for nodes whose events exhibit **mixed membership**;
The **multinomial** approach better for nodes whose events fall in **one category only**;
The mixture approach does well for both.

Media sources	uk.reuters	breitbart	alertnet	canadianbusiness
Top topic	0.4400	0.4061	0.5539	0.5694
Quantile	27.59%	8.62%	75.86%	84.48%
$\hat{R}_m$	$-\infty$	$-\infty$	-1.2601	-1.4343

Outline

- Networks in Auto-regressive Models
  - Context-dependent Network Estimation
    - Models and Methods
    - Theoretical Guarantees
    - Practical Considerations: Interpretation
    - Practical Considerations: Model Choice
    - Real Data Applications
    - Conclusion

Conclusion

- Directed graphs can be understood through autoregressive models
- Context-dependent network:

  explore single category v.s. mixed membership

  data-dependent **testing** procedure

- Models with predictions

  variable importance network

  underlying network structure

Thank you!