Modeling Dynamic Behavior in Large Evolving Graphs

Ryan A. Rossi
Purdue University
rrossi@purdue.edu

Brian Gallagher
Lawrence Livermore Nat. Lab
bgallagher@llnl.gov

Jennifer Neville
Purdue University
neville@purdue.edu

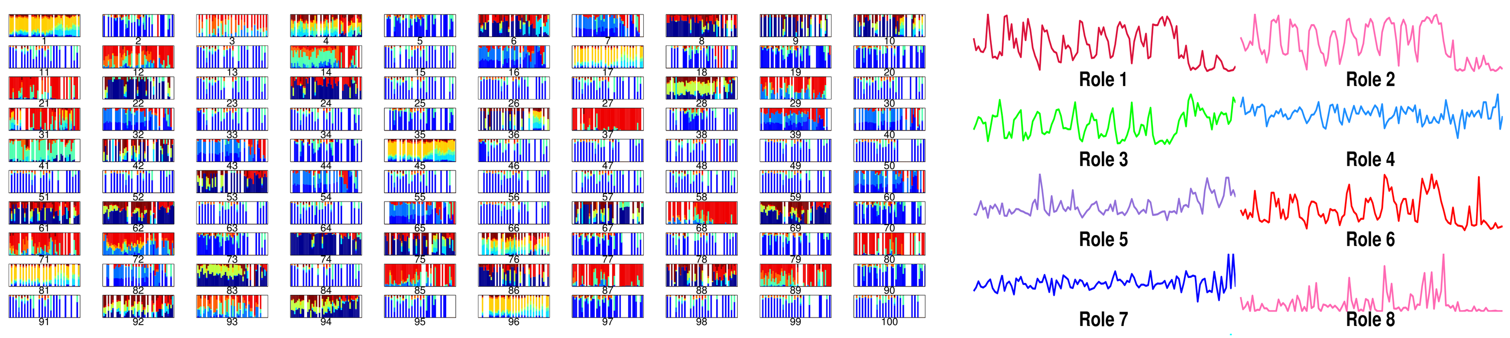
Keith Henderson
Lawrence Livermore Nat. Lab
keith@llnl.gov

Problem Formulation

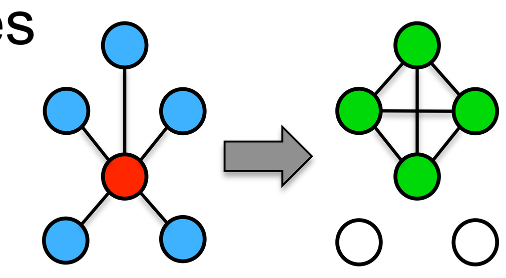
Given a large graph evolving over time, how can we model the roles and their evolution over time?

We proposed DBMM model for three main tasks:

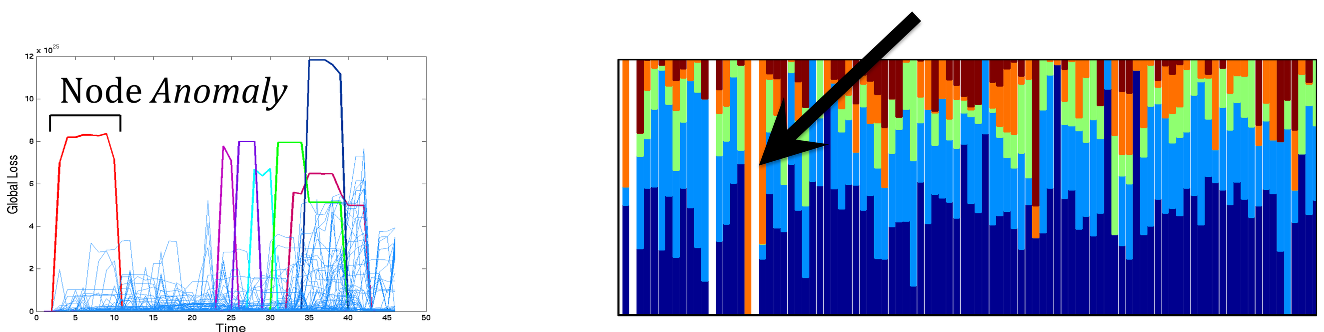
1. **Identify** dynamic patterns in node/global behavior



2. **Predict** future structural changes



3. **Detect** unusual transitions in behavior



Dynamic Behavioral Mixed-Role Model

We develop a role-based DBMM model to explore graph changes over time.

Dynamic Roles. Given a sequence of evolving graphs S_0, S_1, ..., S_t, ... where both edges and vertices may become active or inactive:

1. Find set X of representative features for S_0
2. V = $\{V_t$: Extract the features X for each $S_t\}$
3. Given V_t and a positive integer r, use NMF to find $G_t \in \mathbb{R}^{n \times r}$ and $F \in \mathbb{R}^{r \times f}$ that minimizes:

$$f(\mathbf{G}_t, \mathbf{F}) = \frac{1}{2}||\mathbf{V}_t - \mathbf{G}_t\mathbf{F}||_F^2$$

Note r is selected w/ MDL: **min**(*number of bits + errors*)

4. Next, iteratively estimate $\boldsymbol{G} = \{G_t : t \in \mathrm{T}\}$ given F and $\boldsymbol{V} = \{V_t : t \in \mathrm{T}\}$ using NMF.

Behavioral Transition Models. Given G_{t-1} and G_t find a transition model T that minimizes the functional:

$$f(\mathbf{G}_t, \mathbf{G}_{t-1}) = \frac{1}{2}||\mathbf{G}_t - \mathbf{G}_{t-1}\mathbf{T}||_F^2$$

$$\begin{bmatrix} \mathbf{G}_{t-1} \\ \mathbf{G}_{t-2} \\ \vdots \\ \mathbf{G}_{k-1} \end{bmatrix} \mathbf{T} \approx \begin{bmatrix} \mathbf{G}_{t} \\ \mathbf{G}_{t-1} \\ \vdots \\ \mathbf{G}_{k} \end{bmatrix}$$

All models predict G_{t+1} using G_t.

<u>Snapshot</u>: Uses only the immediate past
<u>Stacked</u>: Uses training examples from k previous timesteps
<u>Summary</u>: Weight training examples from k previous timesteps

$$\mathbf{G}_{S(t)} = \alpha_1 \mathbf{G}_k + ... + \alpha_{w-1}\mathbf{G}_{t-1} + \alpha_w \mathbf{G}_t = \sum_{i=k}^{t} K(\mathbf{G}_i; t, \theta)$$

Baseline Models: Predict future role based on (1) previous role and (2) average role distribution.

DBMM has the following properties:

1. Automatic: No user-defined parameters
2. Scalable for **BIG** graphs: O(m) time to compute
3. Non-parametric & data-driven: # features / # roles
4. Interpretable: explainable trends / dynamics
5. Flexible: notion of role behavior is customizable

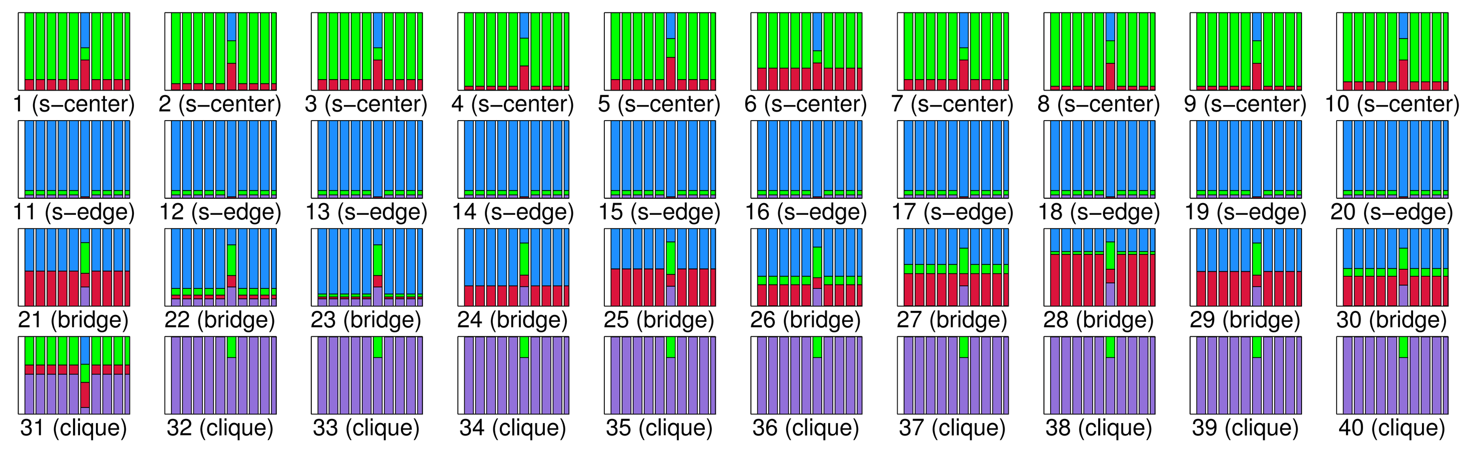
Role Statistics of Networks

Dataset	Feat.	Roles	|V|	|E|	|T|	length
Twitter	1325	12	310K	4M	41	1 day
Twitter-Cop	150	5	8.5K	27.8K	112	3 hours
Facebook	161	9	46.9K	183K	18	1 day
Email-Univ	652	10	116K	1.2M	50	60 min
Network-Tra	268	11	183K	1.6M	49	15 min
Internet AS	30	2	37.6K	505K	28	3 months
Enron	173	6	151	50.5K	82	2 weeks
IMDB	45	3	21.2K	296K	28	1 year
Reality	99	5	97	31.6K	46	1 month

Networks with a greater number of roles are highly adaptive and exhibit more complex dynamics

Dynamic Exploratory Analysis

Synthetic Graph Experiments. Generate sequences of graphs consisting of four main patterns, assess accuracy of DBMM model role discovery

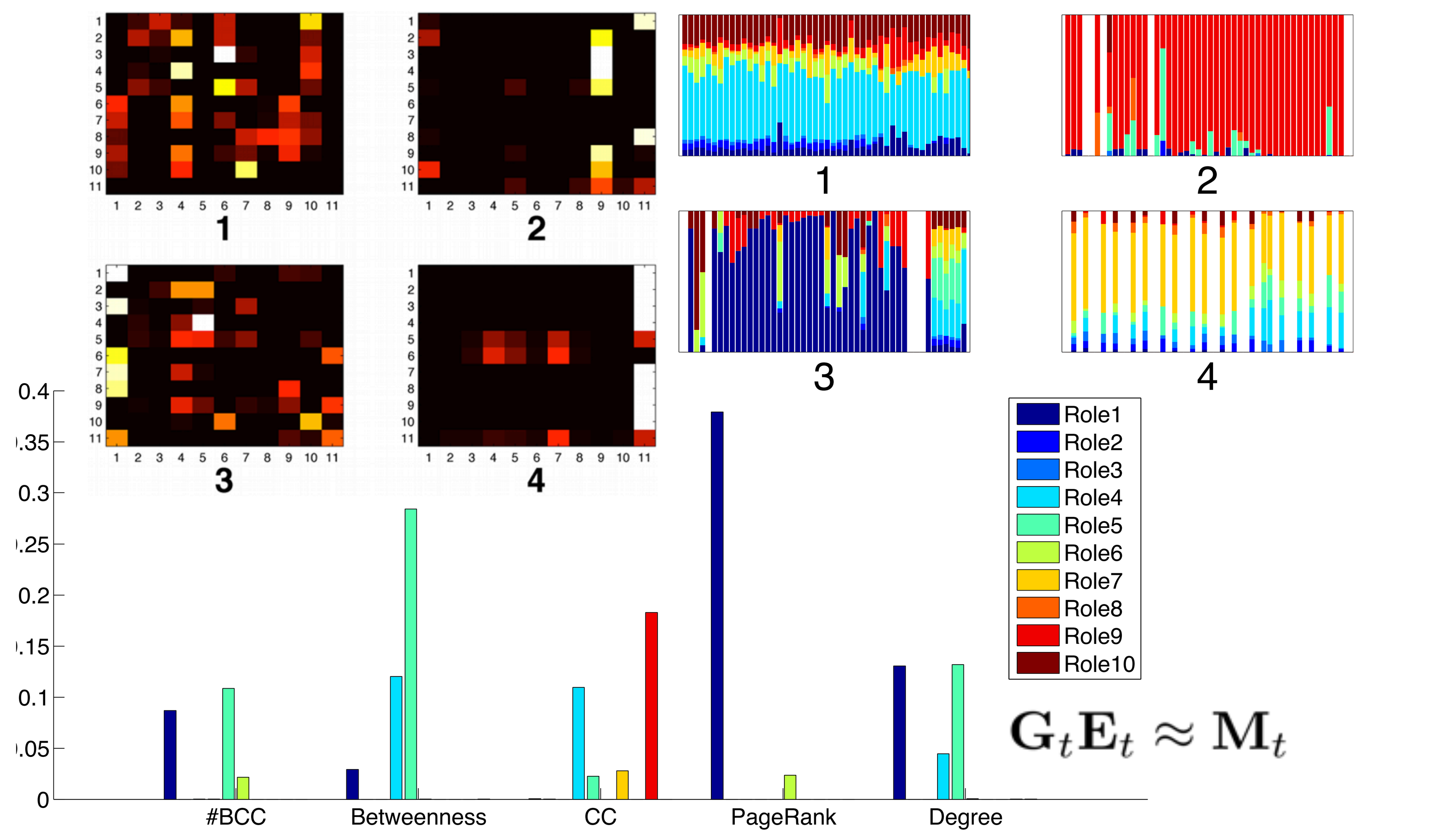


Roles	S-Center	S-Edge	Bridge	Clique
S-Center	**0.07**	0.25	0.33	0.35
S-Edge	0.28	**0.10**	0.22	0.40
Bridge	0.29	0.18	**0.16**	0.37
Clique	0.24	0.25	0.29	**0.22**

Features	S-Center	S-Edge	Bridge	Clique
S-Center	**0.08**	0.25	0.34	0.33
S-Edge	0.27	**0.11**	0.25	0.37
Bridge	0.29	0.20	**0.17**	0.34
Clique	0.24	0.24	0.29	**0.23**

DBMM successfully distinguishes between the ground-truth roles, accurately revealing the known dynamics

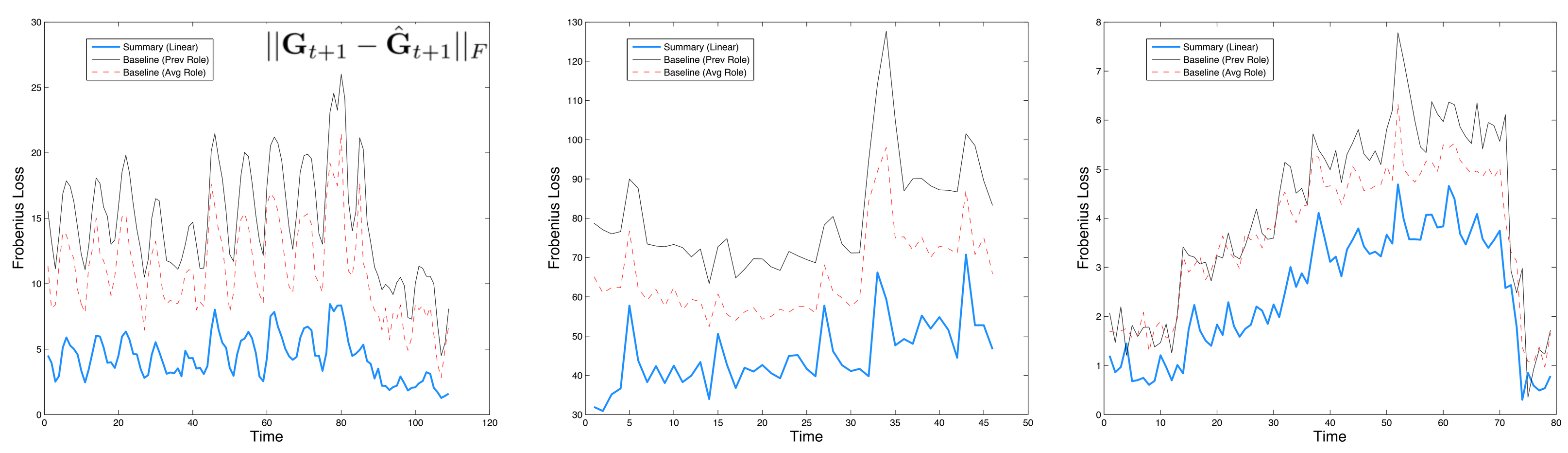
Interpretation and Analysis of Role Patterns.
Some roles are stationary while others are non-stationary and exhibit many of the traditional time-series patterns



Analysis above shows how DBMM reveals the diverse dynamics in a computer network. Roles can be interpreted by a few known statistics.

Predicting Future Behavior

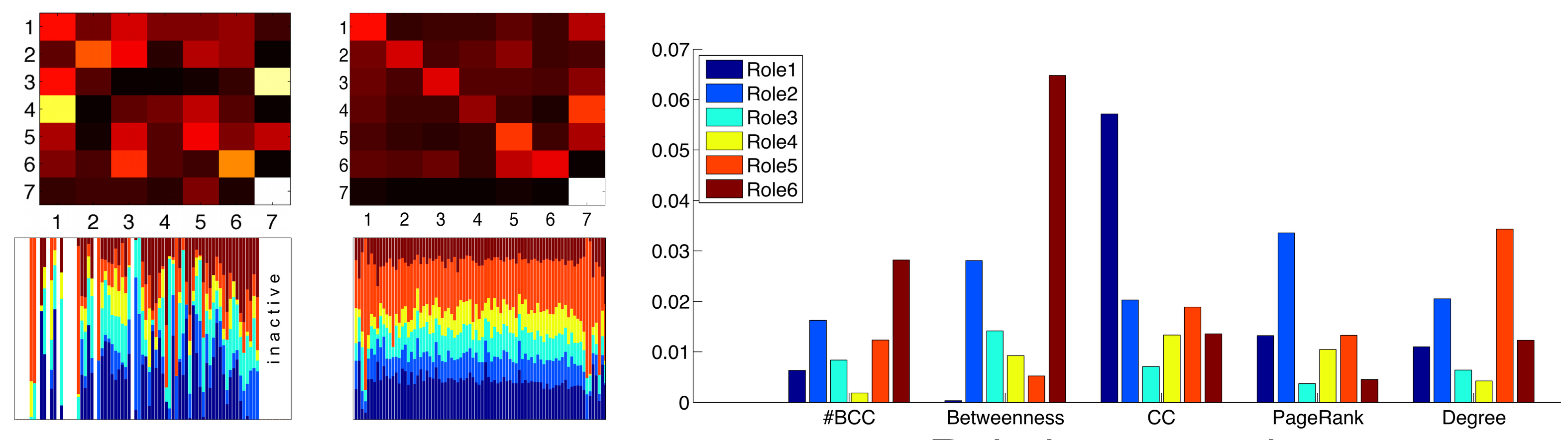
For each vertex, we predict their future role (one-step-ahead), using past role-memberships as training.

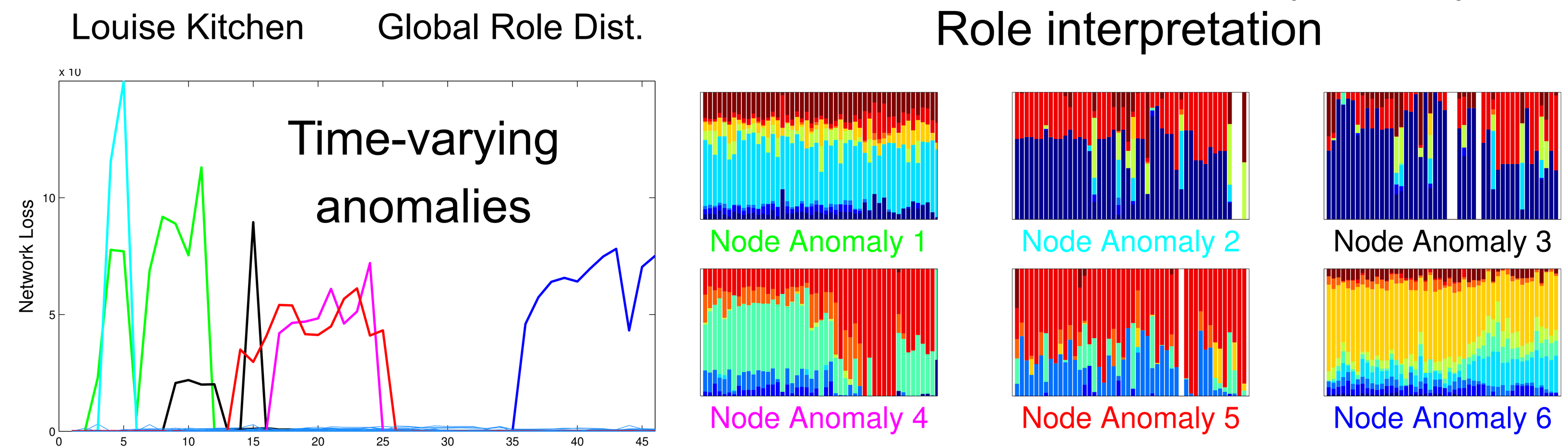


DBMM is more accurate at predicting future behavior than baselines.

Anomalous Structural Transitions

Detect anomalous vertices whose role transitions significantly deviate from the global role transitions.





DBMM finds vertices that are anomalous for only short periods of time and normal otherwise.