Dynamic Bipartite Networks: Detecting and visualizing structural change

Honor Thesis Defense

Yuhao(Eric) Zhao
Mathematics Department, Bates College

Advisor: Dr. Carrie Diaz Eaton

April 20, 2024

Outline

Today my presentation will cover...

- Introduction
- Stochastic Simulation Model
- Visualizations
- Anomaly Detection on Key Indicators
- Takeaways and Directions

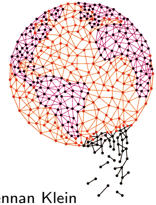


Image Courtesy: Brennan Klein

Bipartite Network

- In order to capture the heterogeneity in the node's nature and complexity in relational data, more and more studies have pivoted toward networks which involve two or more parties
- Bipartite network is a good representation of interaction between two groups of individuals/agents
- But it is often hard to intuitively show the structure change on bipartite networks

	Level 1	Level 2
E 1	1	4
E 2	1	5
E 3	2	5
E 4	3	4
E 5	3	5

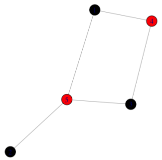


Figure 1: *Simple Bipartite network representation*

Anomaly Detection

- When analyzing large and complex datasets, knowing what stands out in the data is often at least, or even more important and interesting than learning about its general structure.
- Detection of anomalies or any critical transition points at certain timestamps serves well in finding emerging patterns in systems including network-based ones
- Anomaly detection has numerous high-impact applications in security, finance, health care, law enforcement, and many others.

Highlights

This thesis primarily focuses on analyzing evolutionary dynamics of an in-silico bipartite network.

- On visualization part, we extend dynamic visualization method on unimodal networks to the synthetic interaction networks for two parties over time
- On anomaly detection, we first-time adapt an unsupervised model-free method based on time delay embedding in finding unique-events and compare it with traditional methods based on statistical bounds and testing

Model Description

- Plant population i and pollinator population j are tracked via a connectivity matrix, A. J is the set of all plant species that interact with pollinator species j and is of size m_J . Likewise, I is the set of all pollinator species that interact with plant species i and is of size n_I .
- Each entry in A is either 0 for not connected or 1 for connected.
- Probabilistic function for updating evolving interaction matrix:

$$P(A'(i,j)=1) = cA(i,j) + (1-c)m(i,j)$$

and

$$m(i,j) = \frac{w_j(i|j)}{\bar{w}_j}$$

is defined as a fitness probability function that describe how likely species i and j will interact in the absence of previous association for matching.

Static Representation

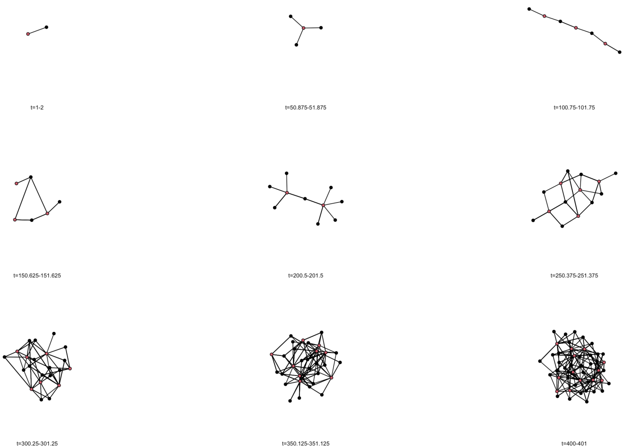


Figure 2: *Snapshot summary of the animation*

Static Representation

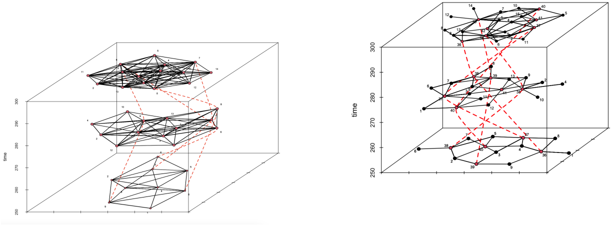


Figure 3: *Layered snapshots of bipartite network (right) and its projection (left) in the prism form at t=250, 275, 300 for sample one run*

Dynamic Representation

t=0-1

Figure 4: *Animation of sample 1 bipartite network evolution*

Network Structure Indicators

Indicator	Definition	Formula
Size	Number of species/nodes at different scales	$n(A_t) = \lvert U \rvert + \lvert V \rvert$
Average Degree	mean of degree of every node	$\overline{k(v)} = \frac{\sum k(v)}{n(A_t)}$
Web Asymmetry	balance between number of species in the two levels	$W(A_t) = \frac{k_h - k_l}{k_h + k_l}$
Connectance	edge density (realization of all possible edges)	$C(A_t) = \frac{L}{k_h \times k_l}$
# of Compartment	Number of blocks along Jordan matrix: $J = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$	
Modularity	Barber's modularity score for bipartite graph	$Q_B = \frac{1}{L} \sum_{u=1}^{k_l} \sum_{v=1}^{k_h} (A_{u,v} - P_{u,v}) \delta(g_u, h_v) = \frac{1}{L} \sum_{u=1}^{k_l} \sum_{v=1}^{k_h} (A_{u,v} - \frac{k_u d_v}{L}) \delta(g_u, h_v)$
Clustering Coefficient	Average proportion of closed triangles divided by total number of open and closed triangles through node v_i	$CC = \frac{1}{n} \sum_{i=1}^{n} \frac{\gamma_G(v_i)}{\tau_G(v_i)}$

Our Approach

- Selecting 10 key indicators that give general information about networks
- Keeping track of selected indicators over time
- Independently run 10 times of the simulated evolutions for sampling of time-series stats
- Decomposing the time-series stats and removing the seasonality and trend component
- Statistical Detection and Unsupervised-learning Detection on stationary remainders

Statistical Detection Results of Sample1

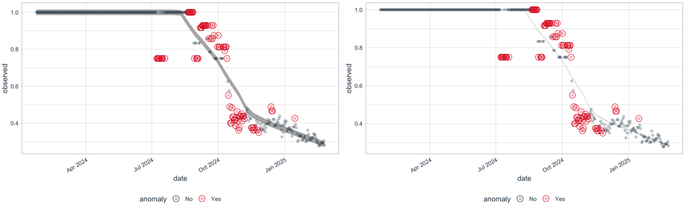


Figure 5: *Statistical Detection of Connectance of Sample 1 using IQR (left) and GESD (right)*

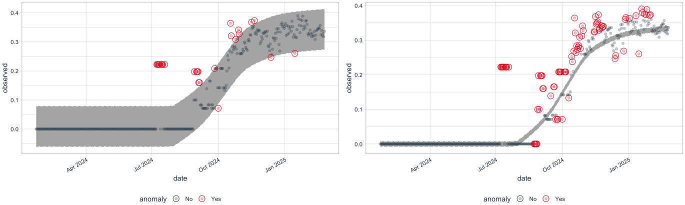
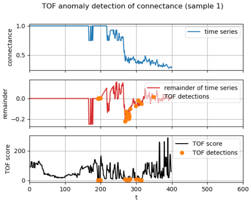


Figure 6: *Statistical Detection of Modularity of Sample 1 using IQR (left) and GESD (right)*

TOF Detection Results of Sample 1



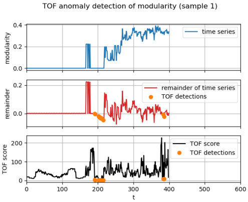
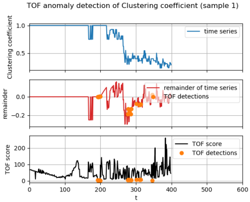


Figure 7: *t=192-199; 269-283; 299, 305, 307, 314*

Figure 8: *t=191-192, 202-215; 384, 386*

Figure 9: *t=192-199, 277-287, 310-313, 346*



Detection Comparison

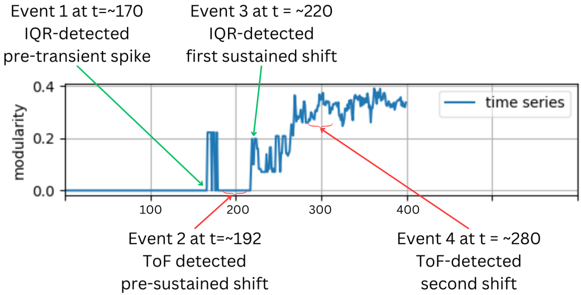


Figure 10: *Comparing detection across IQR and ToF methods on Modularity indicator.*

Takeaways

- Dynamic visualization and even static representation really help people intuitively understand the structural change over time
- Visualization provide a good starting point for event detection
- Choice of detection method heavily depends on the definition of "unique" events
- TOF detection method based off reconstructed state-space perform better
- different indicators might return in different event detection results but common signals among different anomaly detection methods collectively reflecting system change

Future Direction

- Test on parameter change in network generation model and increase sampling size
- Software tool development that supporting better network visualization controls
- Beyond uni-variate detection: Extending TOF detection to Graph-based Detection or Multivariate Detection that better "summarize" network features?
- Considering different scales: community detection techniques in exploration of structural changes