



A Unique Similarity Metric for Anomaly Detection in Temporal Networks

Nathanael Brown (SNL, njbrown@sandia.gov), Katherine Cauthen (SNL),
Justin Durfee (SNL), Chris Frazier (SNL), Dr. Linda Nozick (Cornell)

CoDA: Conference on Data Analysis 2020

Core Idea

The ability to detect structural changes in a network over time can be viewed as an anomaly detection problem. We propose using a control chart on a network similarity metric to detect these changes. The similarity metric ranges from zero (totally dissimilar networks) to one (identical networks) and is defined as follows:

$$S = 0.25 * (1 - L_S) + 0.25 * L_M + 0.25 * N_M + 0.25 * L_C$$

- **Link Strength Difference (L_S)** provides a measure of the link strength similarity between networks and is the sum of the absolute strength differences normalized by the sum of the strengths across both networks.

$$L_S = \frac{\sum_{i=1}^N |S_i^A - S_i^B|}{\sum_{i=1}^N [S_i^A + S_i^B]}$$

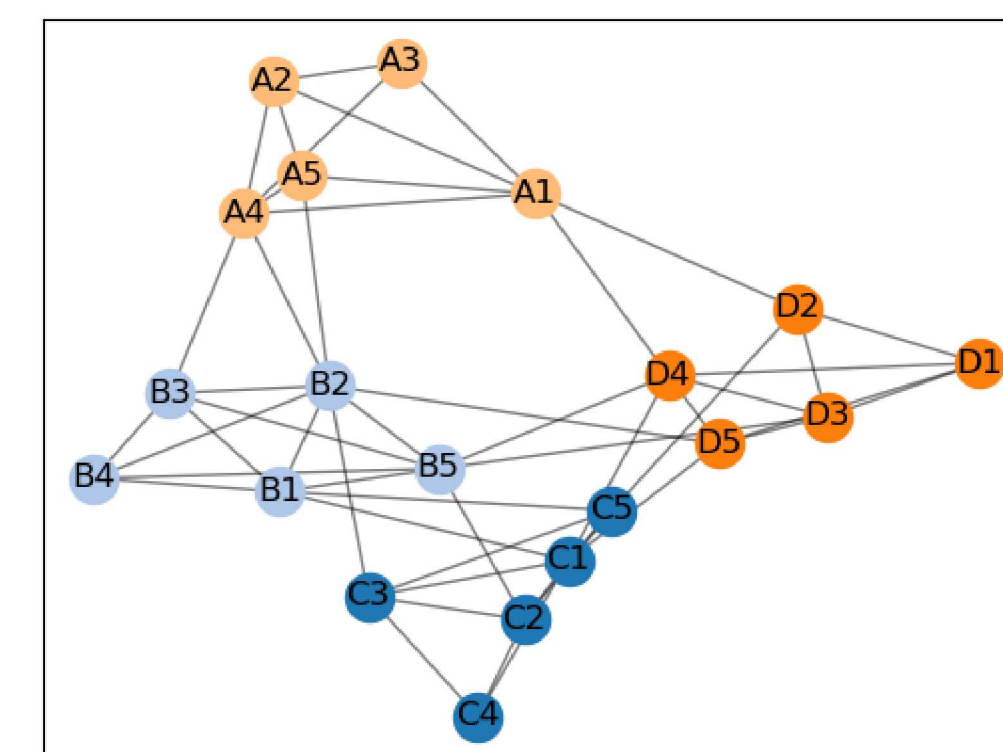
- S_i^A – strength of link i from network A
- S_i^B – strength of link i from network B
- N – max number of links between both networks

- **Matching Link Ratio (L_M)** is the ratio of the number of matching links between networks to the total number of unique links across both networks.
- **Matching Node Ratio (N_M)** is the ratio of the number of matching nodes (nodes with equivalent IDs) between networks to the total number of unique nodes in both networks.
- **In-Cluster Link Difference (L_C)** provides a measure of the within-cluster link strength similarity between networks.

$$L_C = \left| \frac{\sum_{i=1}^M \lambda_i S_i^A}{\sum_{i=1}^M S_i^A} - \frac{\sum_{j=1}^N \lambda_j S_j^B}{\sum_{j=1}^N S_j^B} \right|$$

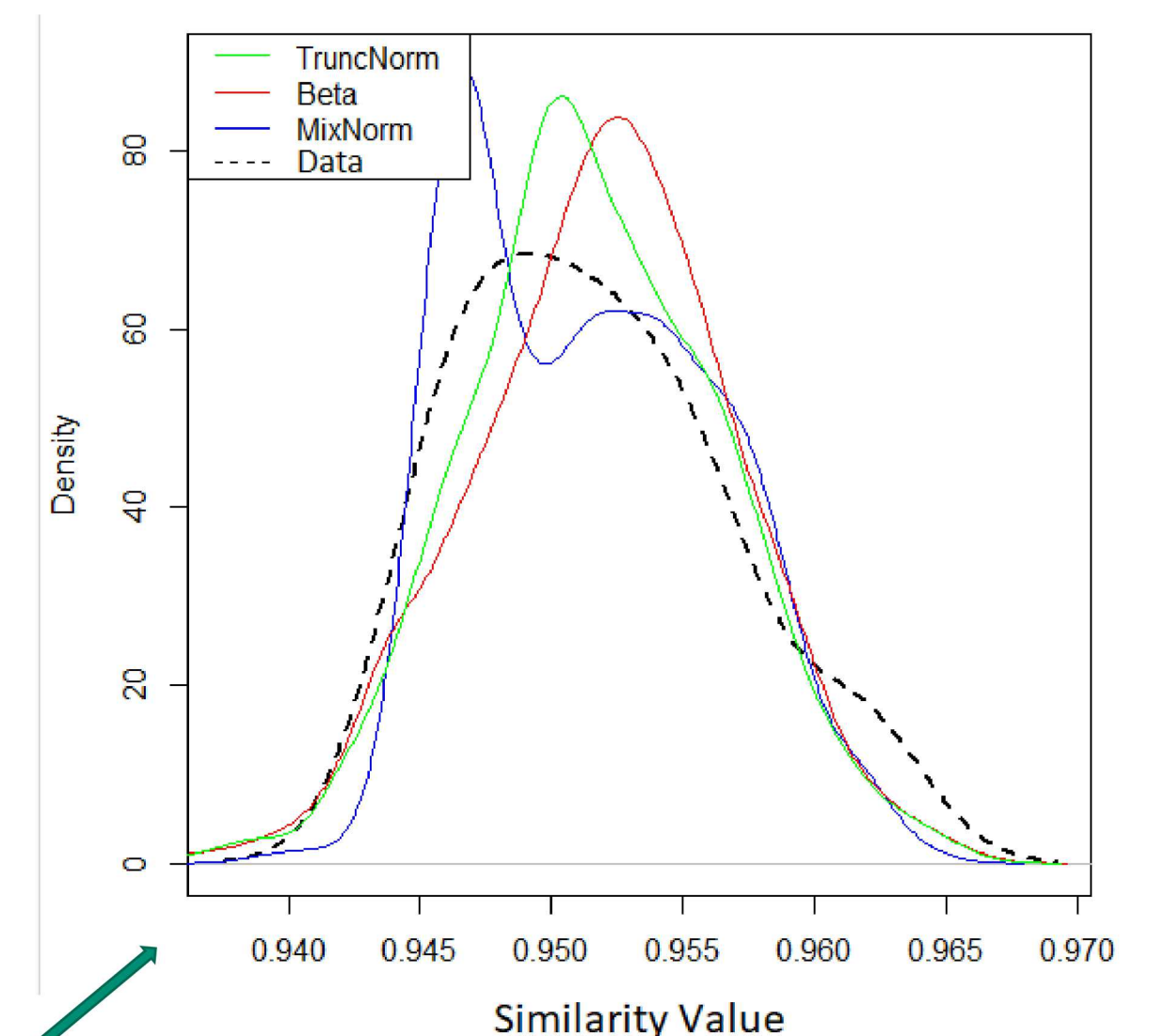
- S_i^A – strength of link i from network A
- S_j^B – strength of link j from network B
- λ_i – binary variable which is one if link i is an in-cluster link and zero otherwise
- M – maximum number of links in network A
- N – maximum number of links in network B

Experimental Validation

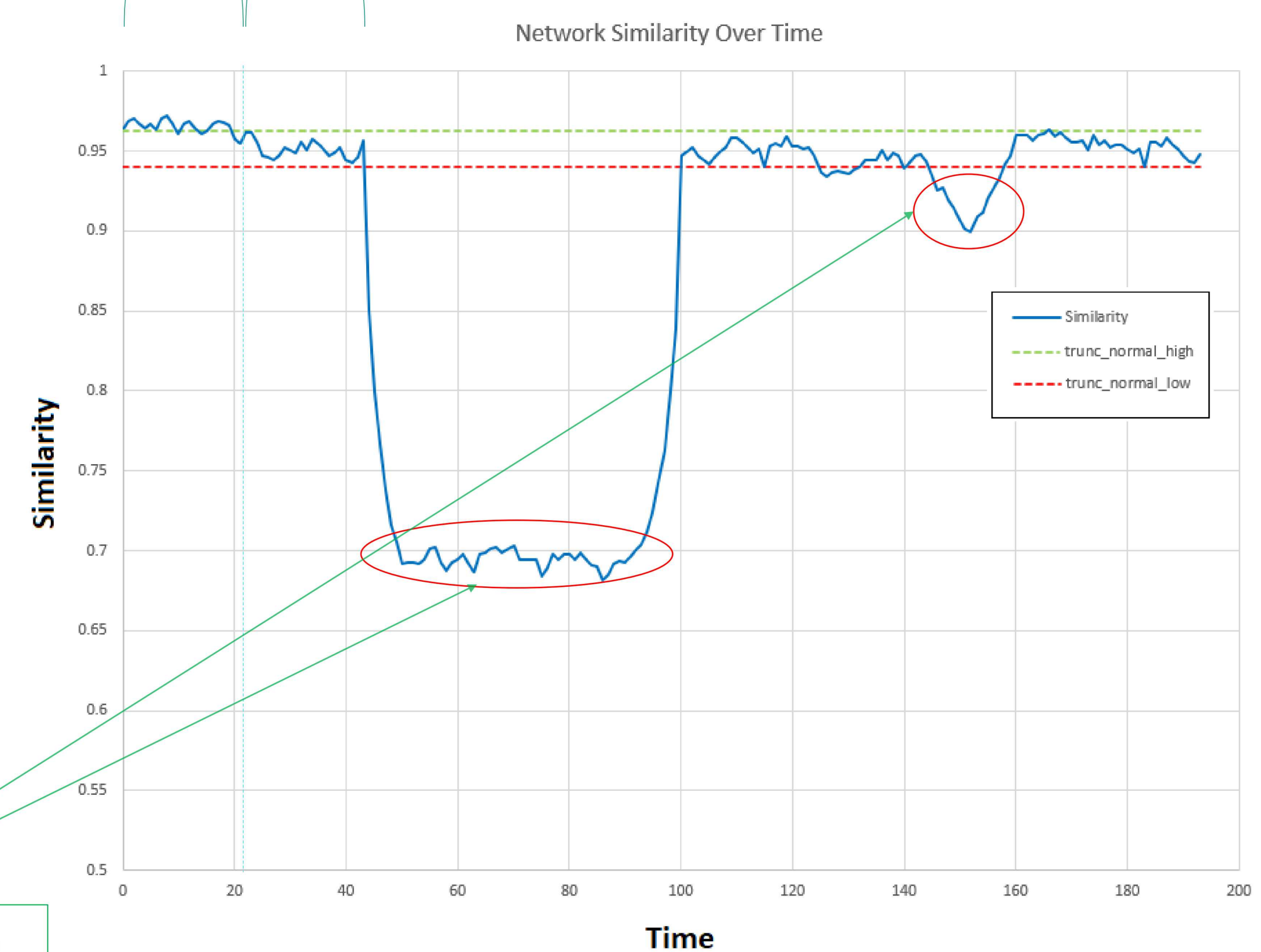


Louvain Clustering for Reference Network

Establish
reference
network



Determine control limits
based on distribution fit



Same anomalous
regions identified
by both similarity
metric and PCA
Rank-k leverage

