SAND2019-7921D

Modeling Temporal Graph Data: Breaking the Stationary Assumption

Tegan Wilson (9365)
Cornell University, PhD Student, Computer Science
Manager: Michael J. Haass (9365)
Mentors: Jeremy Wendt (5854), William E. Hart (1464)
July 18, 2019

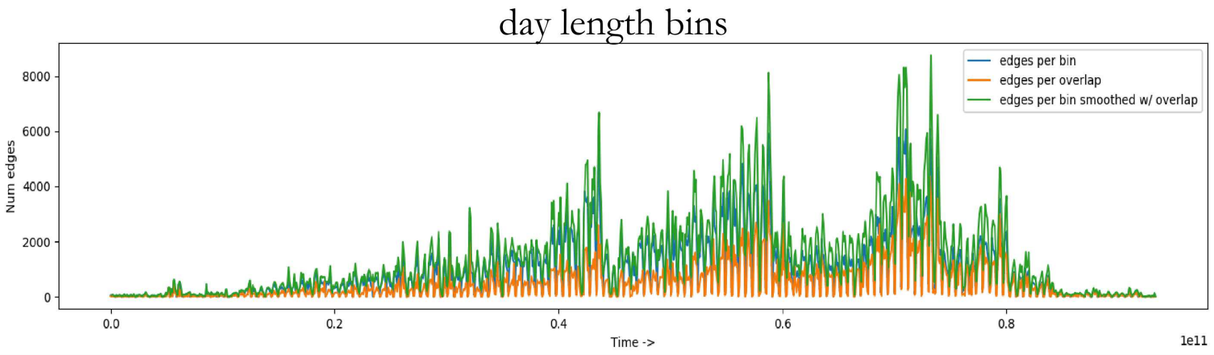
Introduction:

Analysis on temporal network data has become increasingly common to better understand key problems, such as identifying key players or cliques within a communication network. Most models assume some stationary finite underlying truth graph, and assume that at any point in time we see edge advertisements with some fixed probability. We provide examples which break this assumption, and an analytic that can determine if the underlying graph is shifting.

Goals:

- Measure how graphs shift over time
- Apply measure to different datasets
- Incorporate into existing modeling techniques

Enron data number of unique edges per bin
day length bins



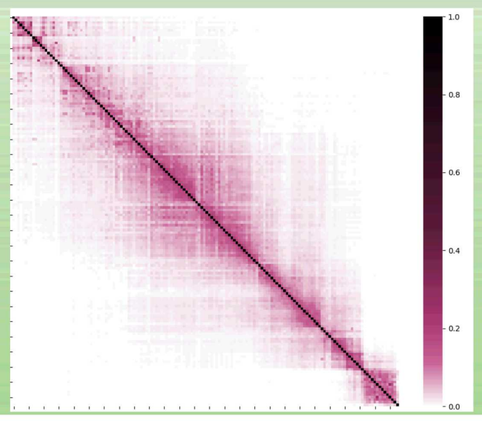
Datasets:

- Enron email
- Reddit responses in discussion threads
- PCAP computer network packet transfer

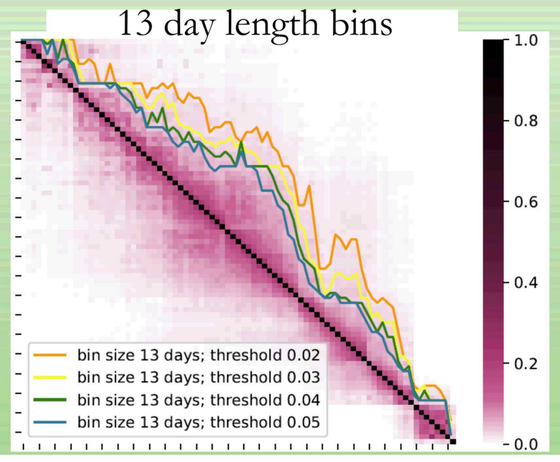
Methods:

- Separate (or bin) edge advertisements using nonoverlapping time intervals
- Jaccard similarity compared to past intervals
- Combine into one measure to judge how stationary a graph is
  - Closest (or furthest) distance into the past to have Jaccard similarity below (or above) certain thresholds
  - L1 (or L2) norm of Jaccard similarity of the last k bins for some parameter, k

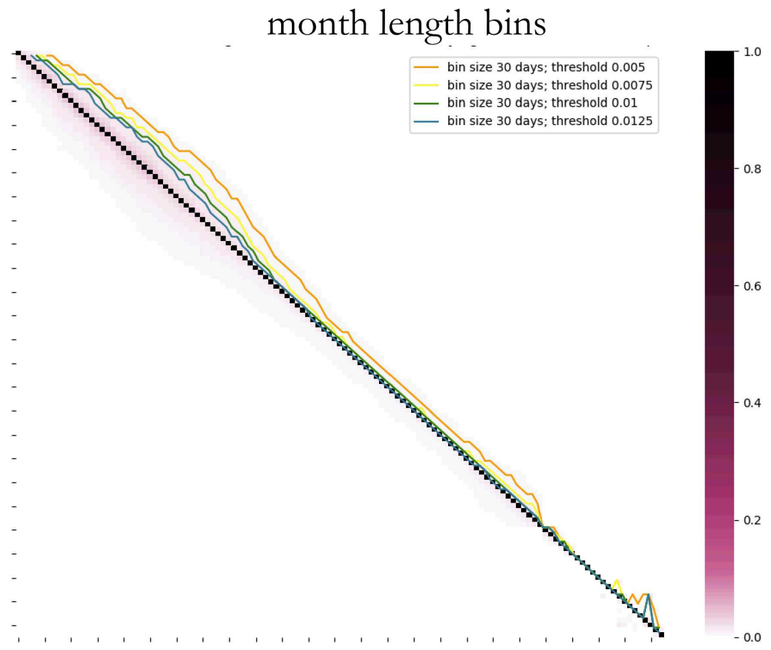
Enron data Jaccard similarity heatmap
week length bins



Overlaid with closest distance in past
below certain thresholds
13 day length bins

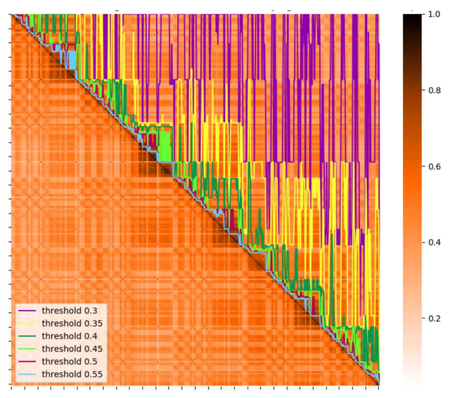


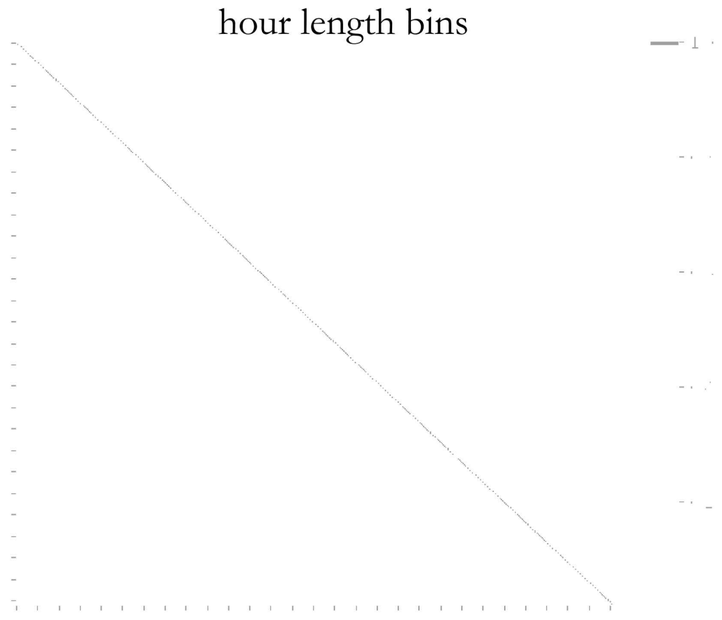
Reddit data Jaccard similarity heatmap
overlaid with closest distance in past below certain thresholds
month length bins



PCAP data Jaccard similarity heatmap
hour length bins

Overlaid with closest distance
from past below certain
thresholds



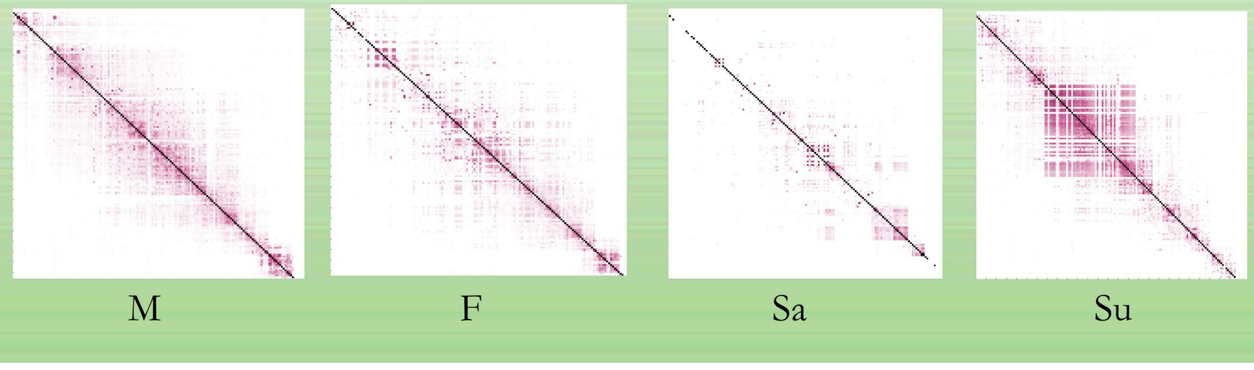


Results:

- Identifies shifting consistently well on each dataset
- Adaptable: each dataset displays different shifting patterns:
  - Slow shift over time, abrupt shift, consistent patterned shifts, etc.
- Multiple parameters allow for use in other applications

Conclusion and Next Steps:

- Incorporate into existing models: gives a way to identify stationary parts of a graph
- Jaccard similarity of vertices over time
- Analytical models that account for shifting graphs?
- Model shifting graph topology?



U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. _____

Sandia National Laboratories