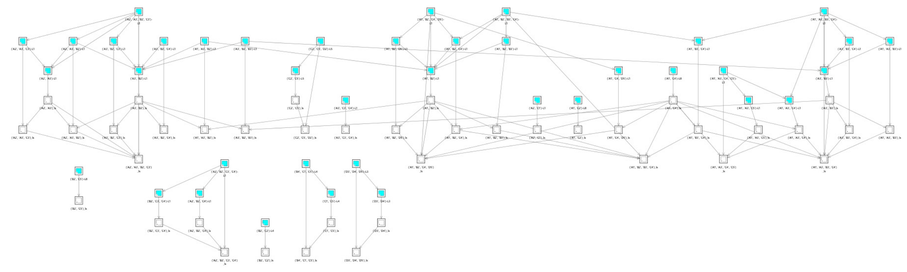
SAND2023-00188C

NETWORK EVENT FORECASTING

M. Hoffman, A. Sorensen, B. Arguello, S. Simpson, B. Yang, C. Doyle, J. Mott, N. Brown | mjhoffm@sandia.gov

Project Overview

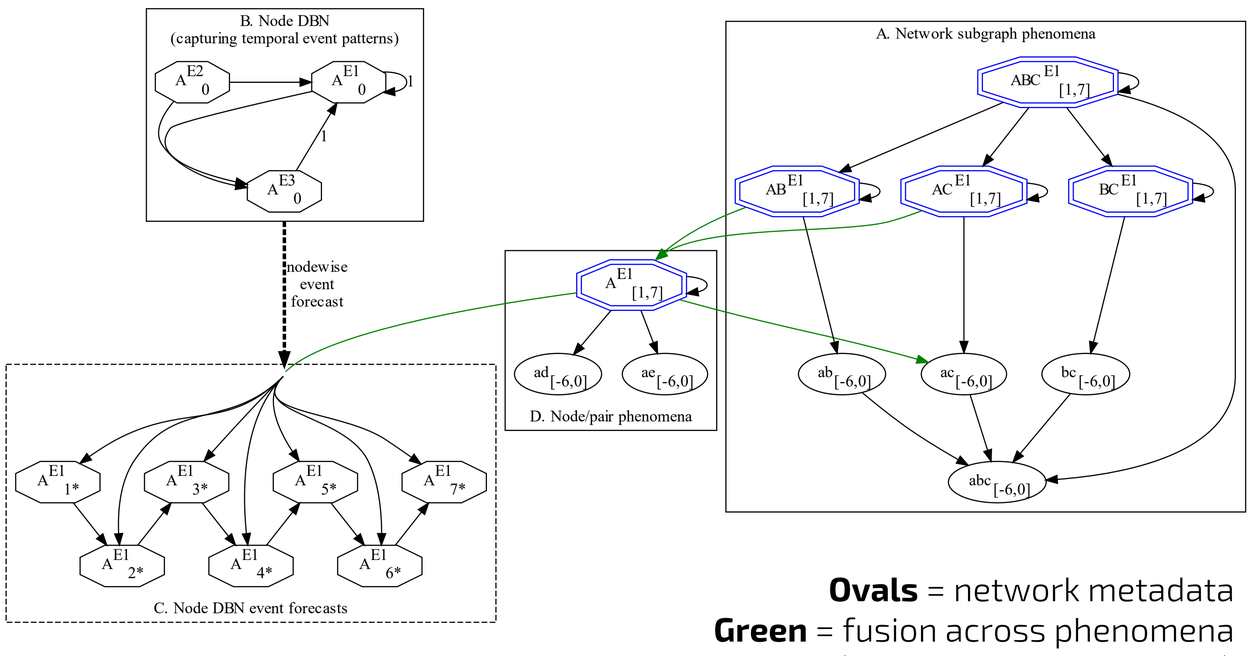
We pursue forecasting of discrete events that involve small subgraphs within a network (for example, raiding parties by a clique of players in an online game).
We search for hypothesized relationships between data elements (e.g., prior player behavior and future raiding party occurrences) and then build multilabel Bayesian Network (BN) time-shifted classifiers from the results.



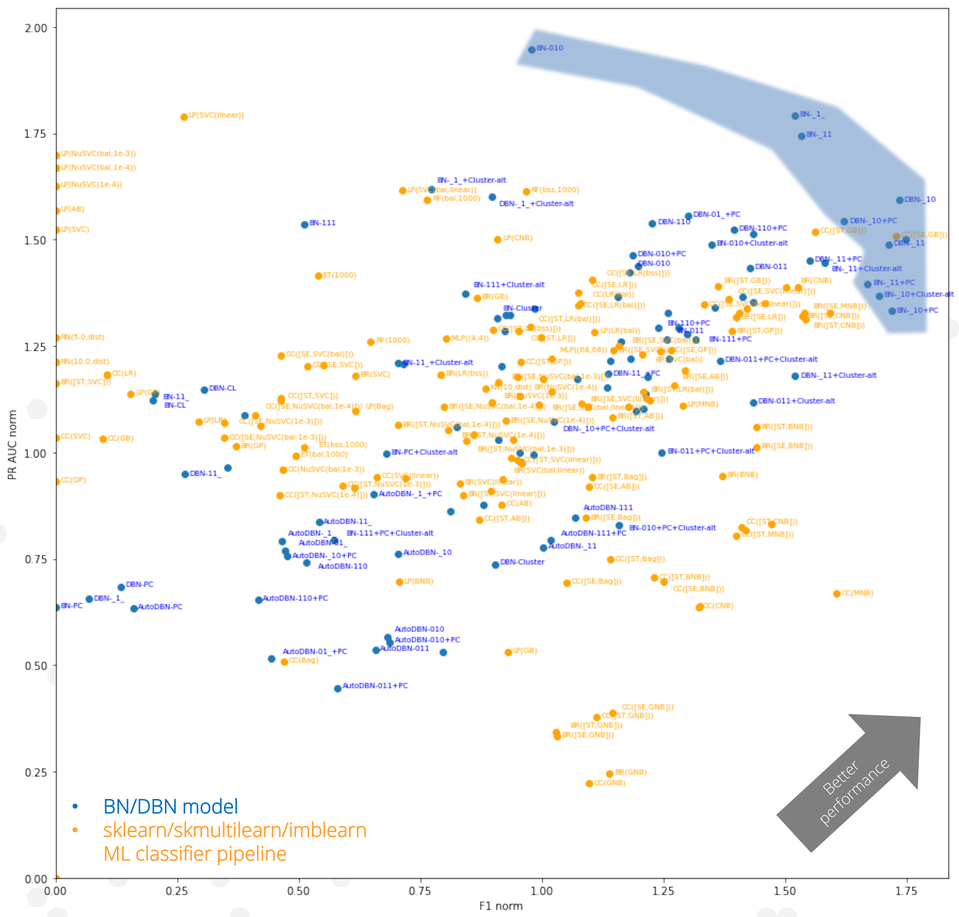
Multidimensional Bayesian Networks (BNs) based on network subgraph phenomena

- act as multilabel classifier of future events
- provide probabilistic forecast from prior data
- can perform better than traditional ML pipelines as well as algorithmically built BNs (see blue highlighted region below)

Predictive Phenomena/Models



Ovals = network metadata
Green = fusion across phenomena
Octagons = events (blue are forecast targets)



Research in Progress

Exploring improvements to forecasting method performance, generality, and scalability via:

- new BN structure learning genetic algorithm
- supervised feature discretization research
- exploratory data analysis of massively multiplayer online game
- academic partnerships exploring alternate encoding/forecasting (e.g., LSTM+GNN)
- more sophisticated problem and feature identification methods, e.g.,
  - frequent subgraph mining/ neural subgraph learning
  - two-mode generalized blockmodeling
  - **persistent homological data encodings**

Algorithm	Coffee	Trace	BirdChicken	Worms	Herring
HIVE-COTE2	99.88	100	95.17	75.80	62.60
InceptionTime	99.63	100	95.17	78.10	62.50
TopRocket	100	100	90.33	75.24	61.04
MultiRocket	100	100	88.83	72.99	61.3
Arsenal	100	100	88.61	71.47	62.40
LowerstarNN	98.93	99.03	90.0	71.21	**63.91**
TakenBettiRF	95.95	99.83	**96.83**	71.04	**64.84**
Catch22	97.98	98.3	89.33	72.5	55.57
BettiNN	91.66	99.6	88.0	68.07	**68.07**
NN	98.81	87.03	88.66	58.83	66.77
RISE	98.45	99.2	86.83	68.66	59.84
DTW-F	97.29	99.74	86.50	67.34	60.92
TSF	98.69	99.2	81.50	62.21	60.42
DTW-Baseline	100.0	100.0	75.0	58.44	53.12
Lower*BettiRF	85.97	94.03	72.67	65.37	55.52
ED-Baseline	100.0	76.0	55.0	45.45	51.56

Simple classifiers using persistent homological features can outperform state-of-the-art classification algorithms on time series classification benchmarks

Sorensen, A. Simpson, S., Krofcheck, D., Hoffman, M., Topological Feature Face-off for Time Series Classification, IJCNN 2023 (in review)

Key Points

- Enables forecasting on small, wide data via fusing weak indicators
- Results in relatively interpretable, self-explainable models
- Enhancing Python library for BN ML (built on COTS API)

LDRD LABORATORY DIRECTED RESEARCH & DEVELOPMENT
U.S. DEPARTMENT OF ENERGY
NNSA 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.
SAND2023-00188C

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.