Sandia National Laboratories

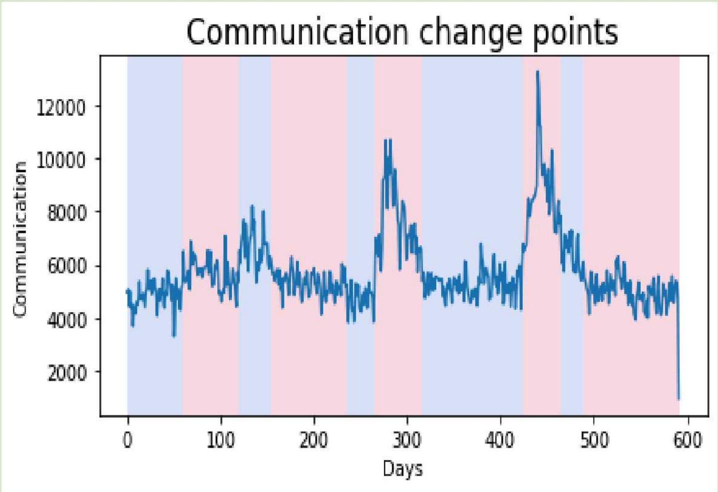
SAND2019-8113D

Identifying Anomalous Structure in Game Data

Can communication data be used to detect major events?

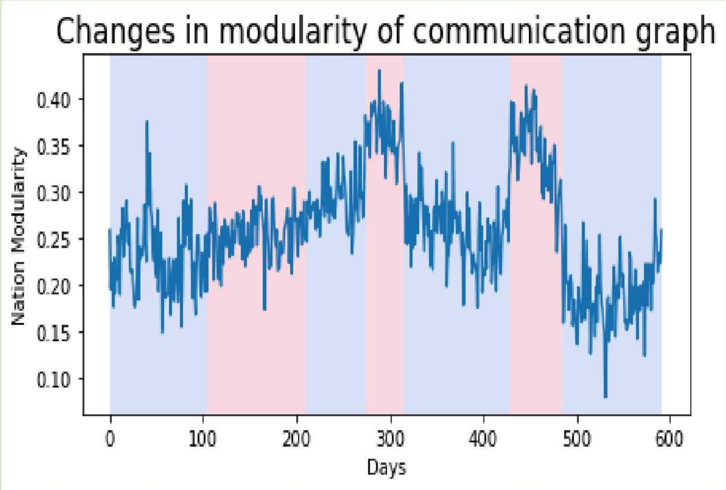
Given historical communication data between players each day in a video game, can we predict major historical events (wars) which occurred in that game?

- There are peaks in total communications during wartime.
- A standard change-point algorithm (PELT), can detect the peaks in communication data.
- **The comm. peaks can be detected online by an ARIMA model forecasting one day in advance by looking for errors four standard deviations away from the mean error.**
- **Nations experience a stronger sense of community during wartime.** That is, the modularity of the partition of the communication graph into nations peaks during wartime.
- **Guilds experience a slightly stronger sense of community during wartime.** Note that the first war isn't captured by the PELT algorithm.

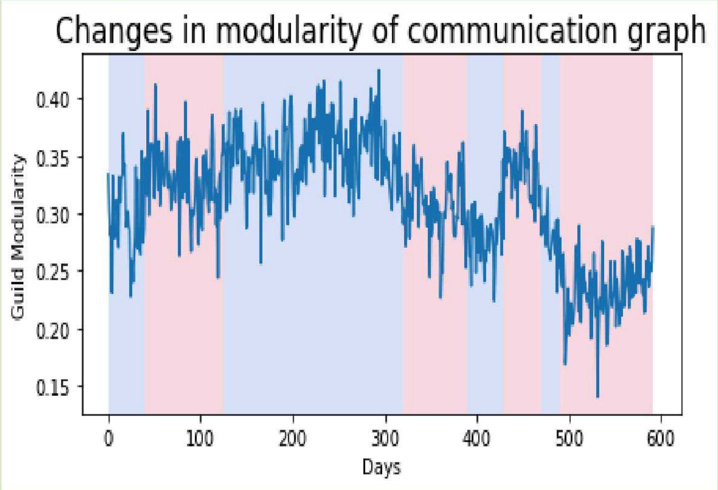


PELT detections of change-points in communication data.

Difference between predicted and actual data in a (5,1,1) ARIMA model. Red lines show wars, black bars show dates with significant errors.



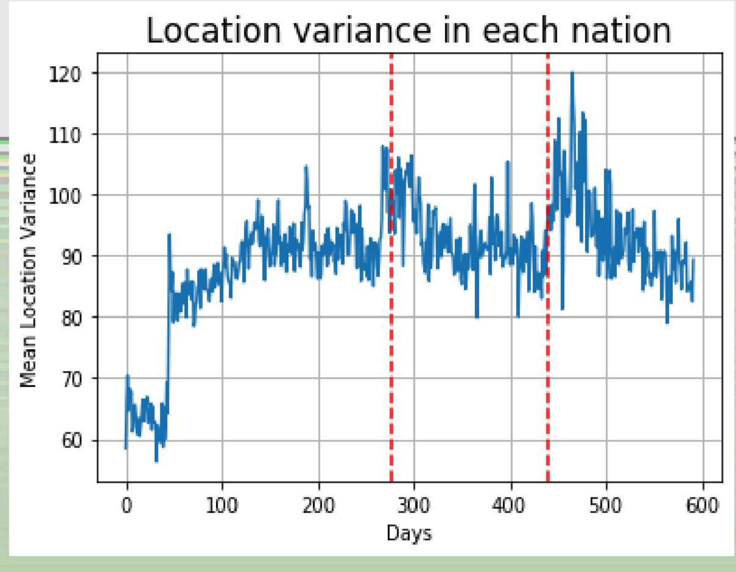
PELT detections of change-points in modularity of partition of communication graph into nations.



PELT detections of change-points in modularity of partition of communication graph into guilds.

Can location data be used to detect major events?

- **No clear pattern linking location variance with wartime.**
- The large spike early on in location variance is due to expansion of the game universe.



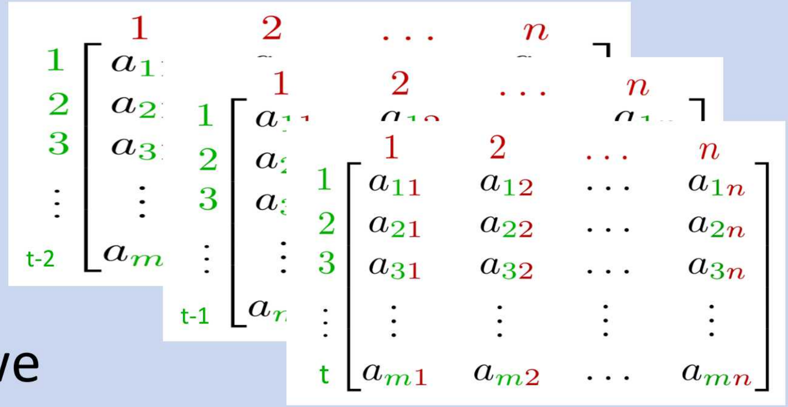
Variance of the locations of players in a given nation (mean over all nations). Red lines are dates of wars.

Can changes in the shape of the data predict major events?

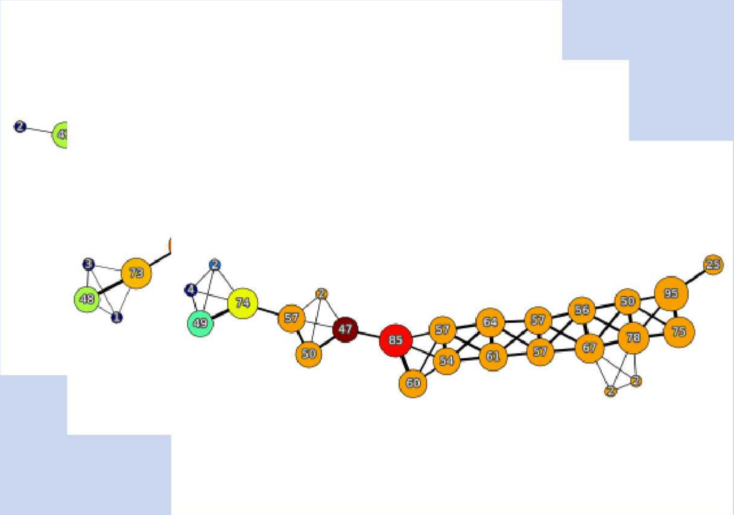
Using the Mapper algorithm, can we predict the peaks in the communication data before they happen? The steps to performing the algorithm are outlined below.

1. Create a data matrix where rows correspond to days $0{:}\,t$, where t denotes the current day. Columns correspond to nodes in the daily communication graph.
2. The feature vector for the communication graph on day i is the vector whose jth entry is the closeness centrality of node j.
3. Use the Mapper algorithm to turn this matrix into a graph. The nodes in the graph correspond to clusters of days; an edge connects two nodes if the days comprising the two nodes have similar data.
4. Calculate a simple graph metric like number of nodes, Euler characteristic, etc.
5. As t increases, create a time series where the vertical axis corresponds to the quantity computed in step 4.
6. Make the time series stationary by computing first differences. Plot the first differences over time to look for anomalies.

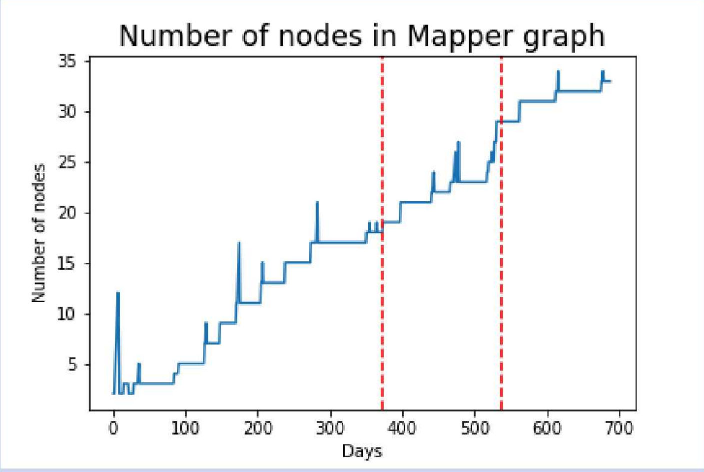
Anomalies are detected one week before the ARIMA model notices the spike in communications.



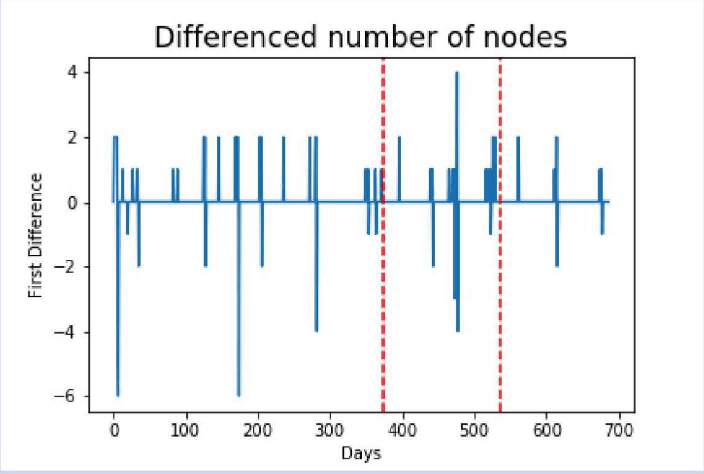
Data matrices. Rows correspond to days, columns correspond to features of the daily communication graph.



Three outputs of the Mapper algorithm.



Number of nodes in the Mapper graph for days 0:t as t ranges from 1 to 689.



First differences of the time series above. Anomalies are detected one week before the ARIMA model notices the spike in communications.

Future Directions

- Create a pipeline to ease interpretability of topologically detected anomalies.
- Remove reliance on a choice of specific feature vector from communication graph.

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