

Prediction in Projection Using Google Search Trends

Chaotic Dynamics Final Project - April 27

Allie Morgan

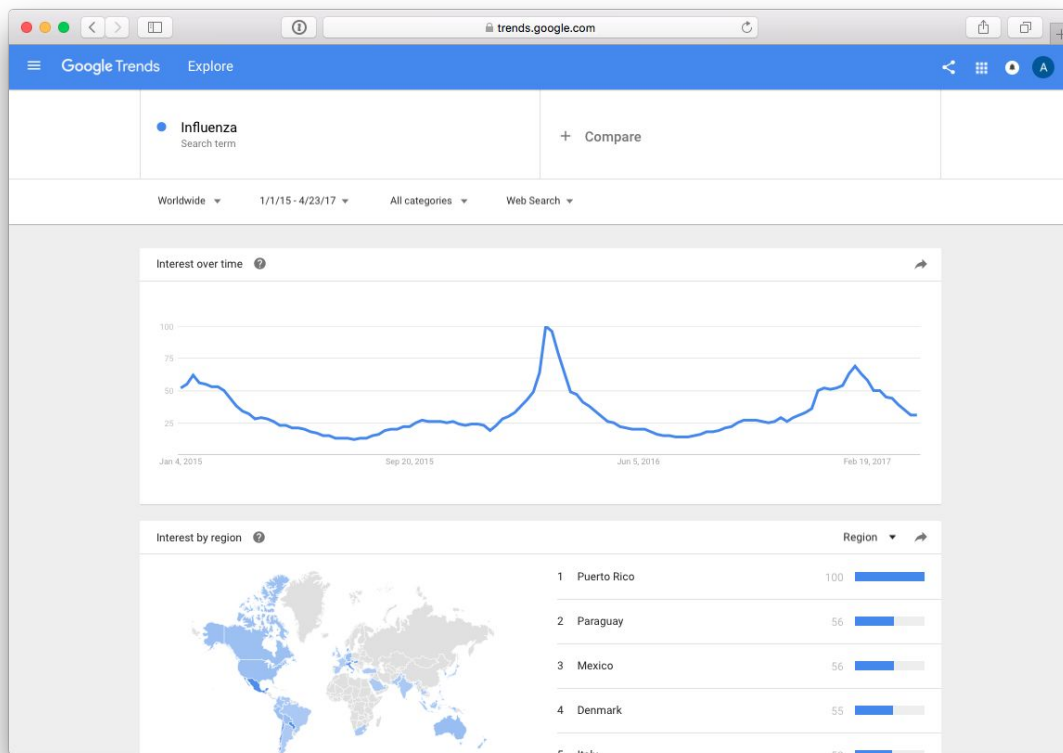
Research Questions

- Are Google search trends predictable?
- Are some trends more predictable than others?

Roadmap

- Google search trends
- Predictability: Weighted permutation entropy
- Review of delay coordinate embedding
- Prediction method: Lorenz method of analogues
- Results
- Future work

Google Search Trends

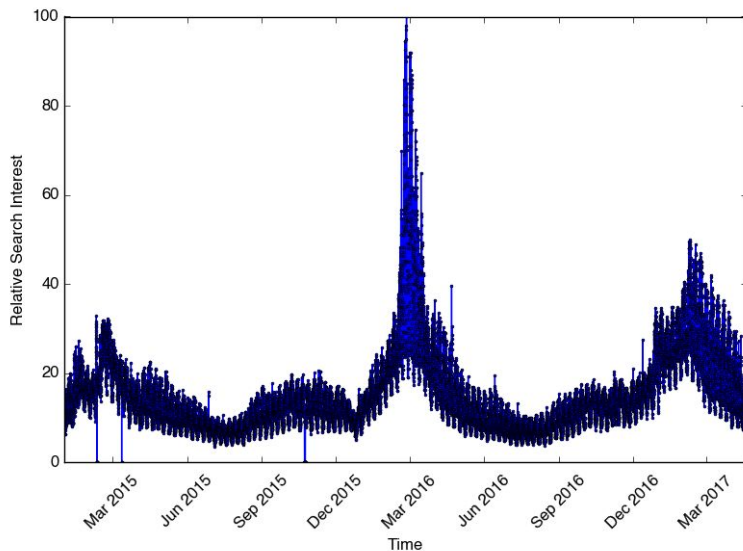


<https://trends.google.com>

Data

Three data sets:

- (1) “Influenza”
- (2) “Baseball”
- (3) “Full moon”



Hourly data from **January 1st, 2015**
to April 20th, 2017.

Resulting in **20,280 points** for each
trend.

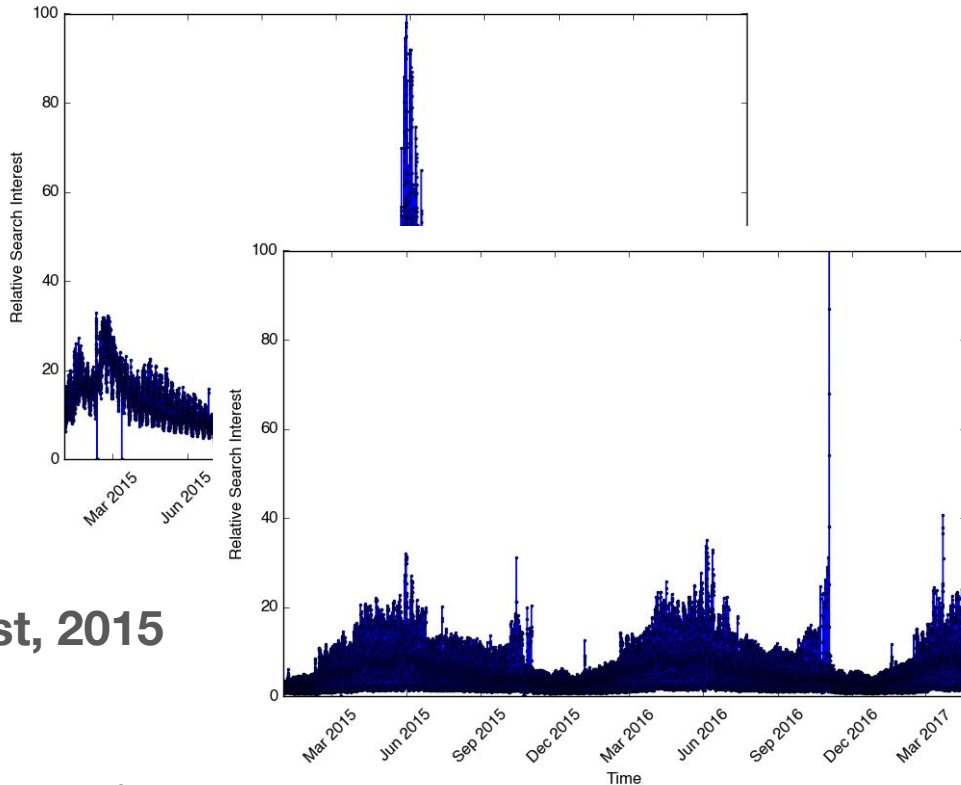
Data

Three data sets:

- (1) “Influenza”
- (2) “Baseball”
- (3) “Full moon”

Hourly data from **January 1st, 2015**
to **April 20th, 2017**.

Resulting in **20,280 points** for each
trend.



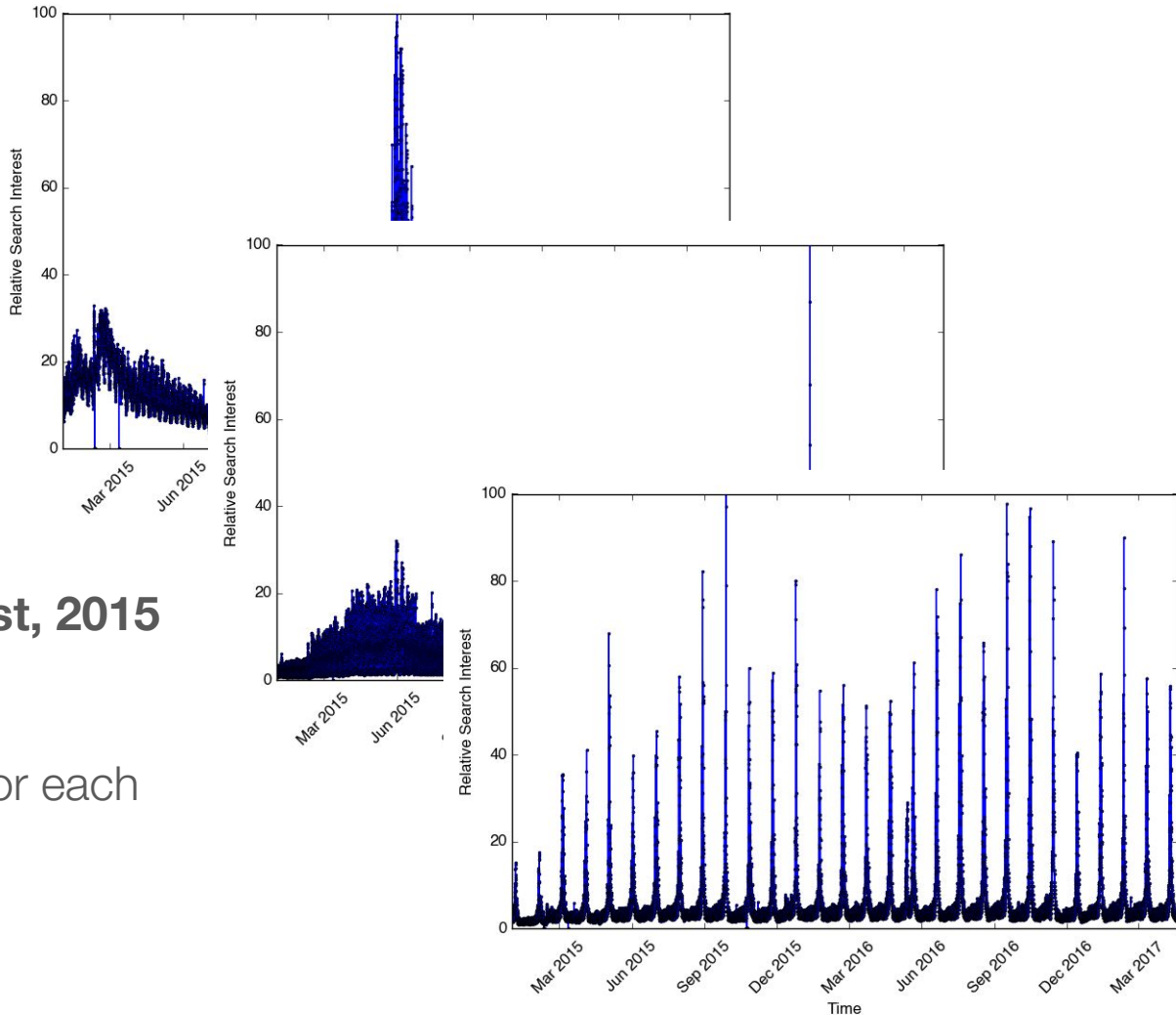
Data

Three data sets:

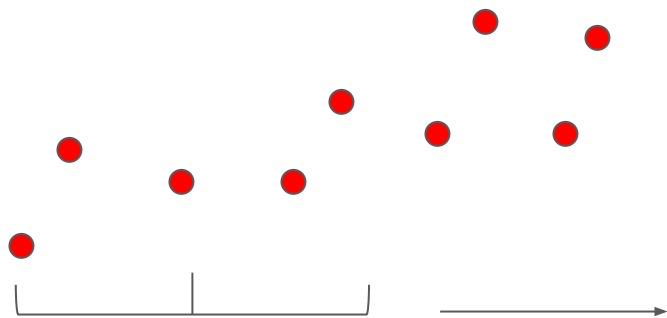
- (1) “Influenza”
- (2) “Baseball”
- (3) “Full moon”

Hourly data from **January 1st, 2015**
to **April 20th, 2017**.

Resulting in **20,280 points** for each
trend.



Predictability - Weighted Permutation Entropy (WPE)



ℓ = # of teeth in your comb

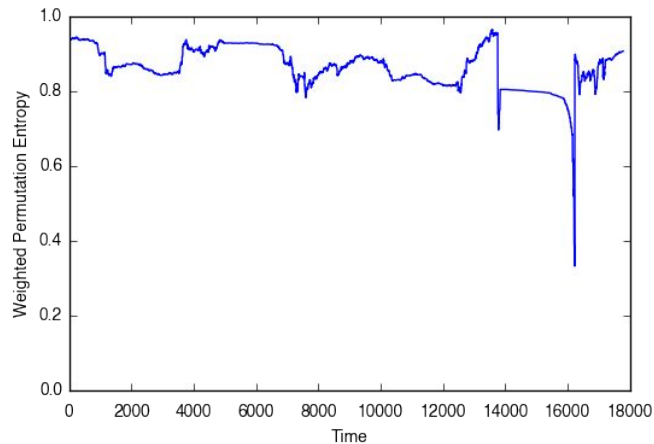
τ = # of points between teeth

\mathbf{N} = # of points averaged across

Basic overview:

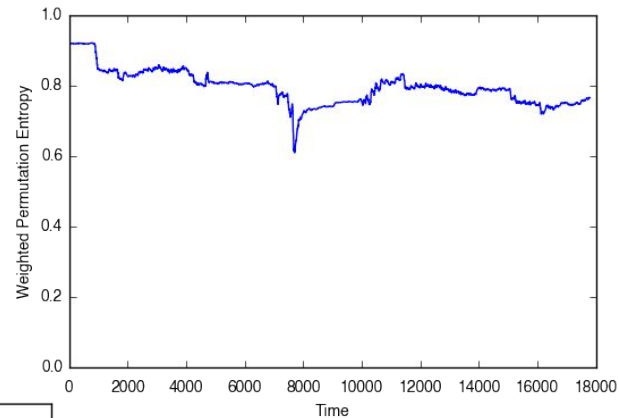
- (1) Pick ℓ points, τ apart
- (2) Rank those points
- (3) Do the same across \mathbf{N}
- (4) Return the probability of finding the first ordinal ranking within the window,
- (5) ... **weighted** by how far away from the average trend

WPE Results

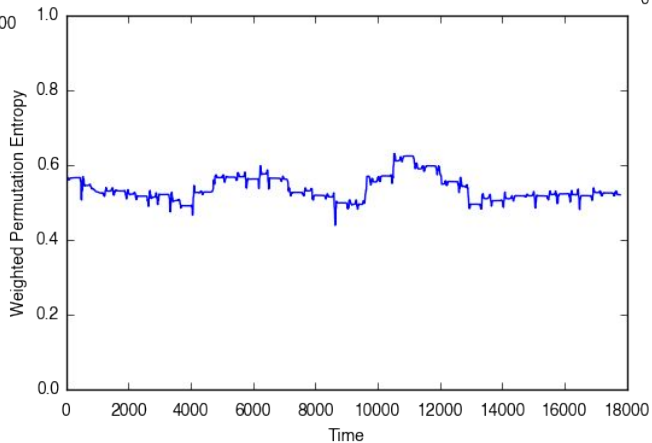


“Influenza”

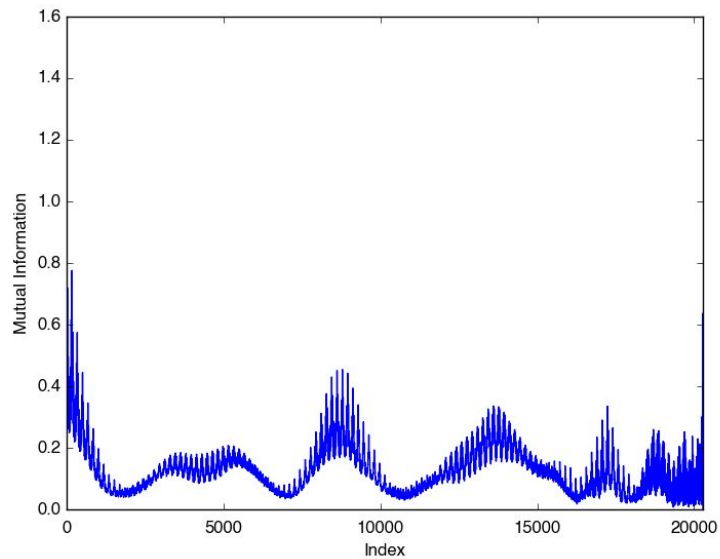
“Full moon”



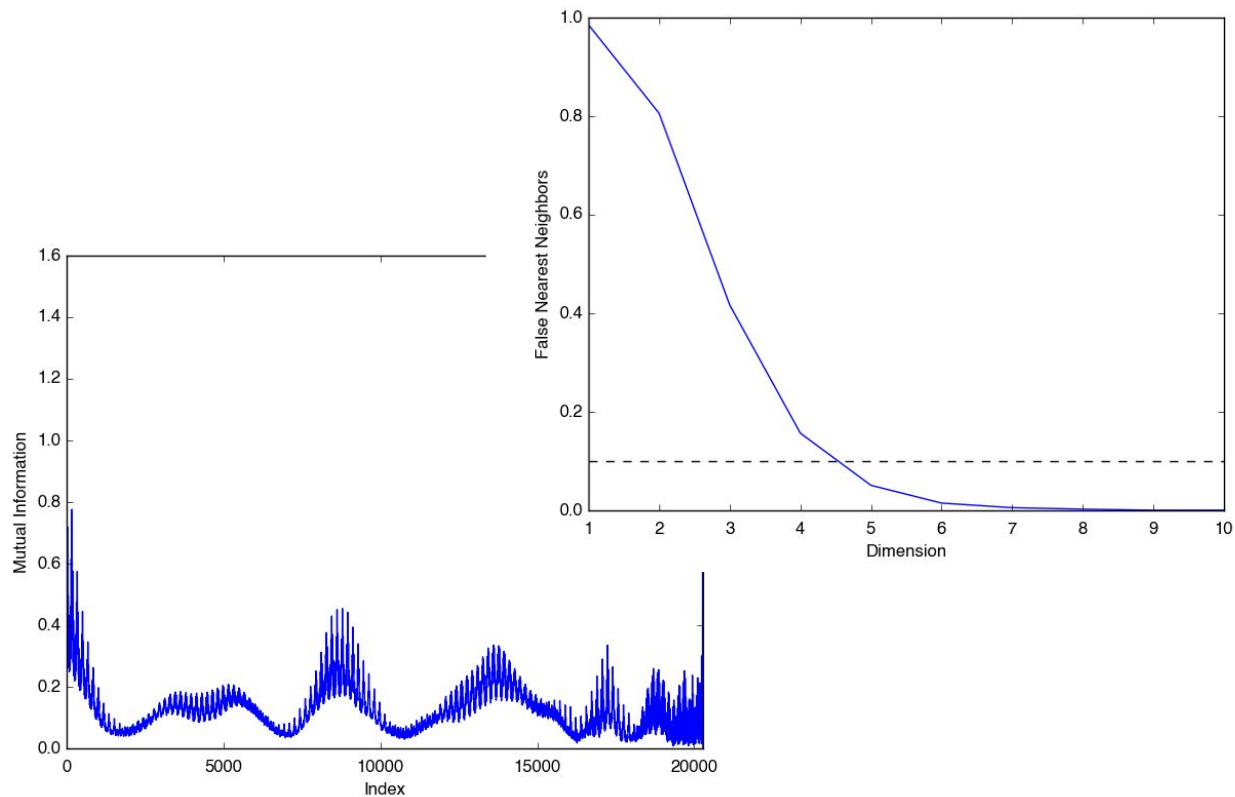
“Baseball”



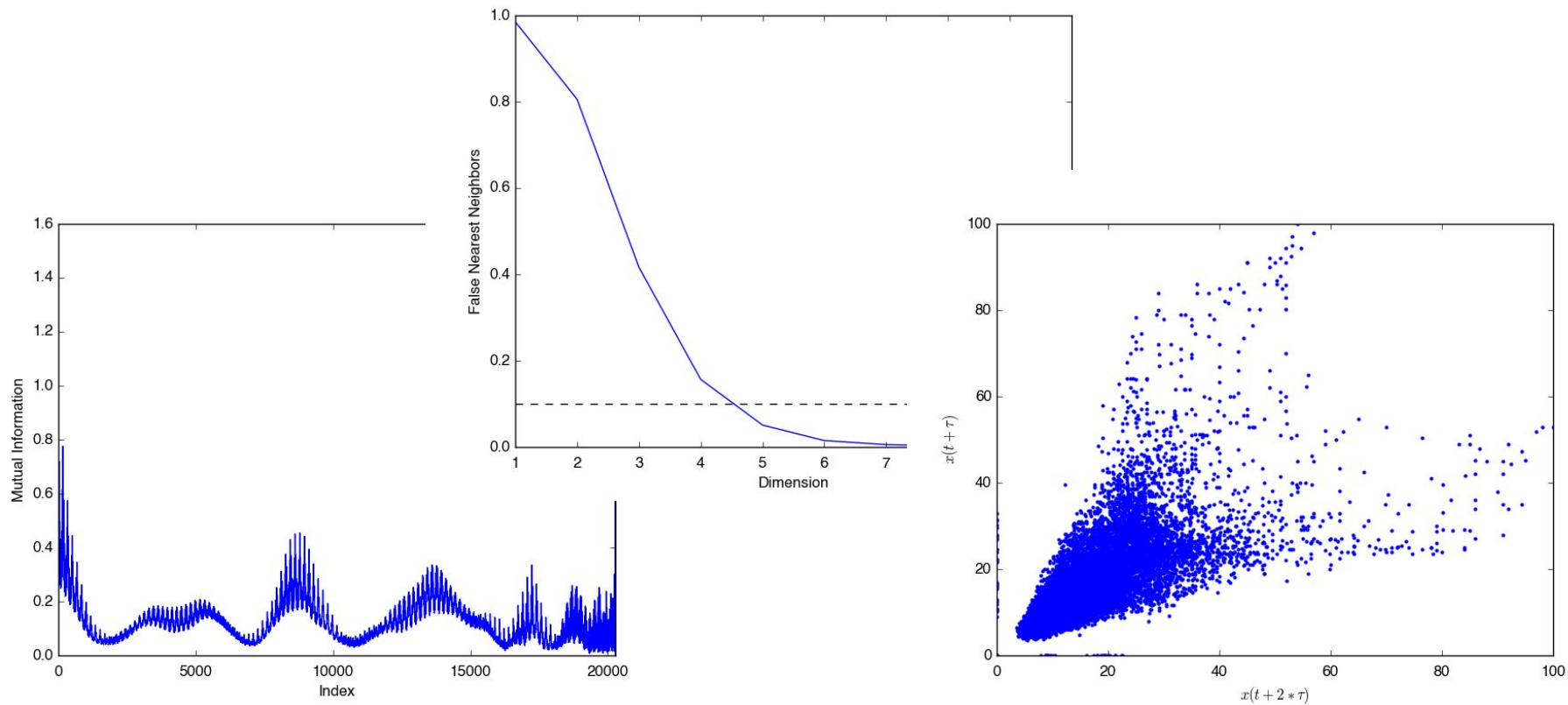
Review of Projection



Review of Projection

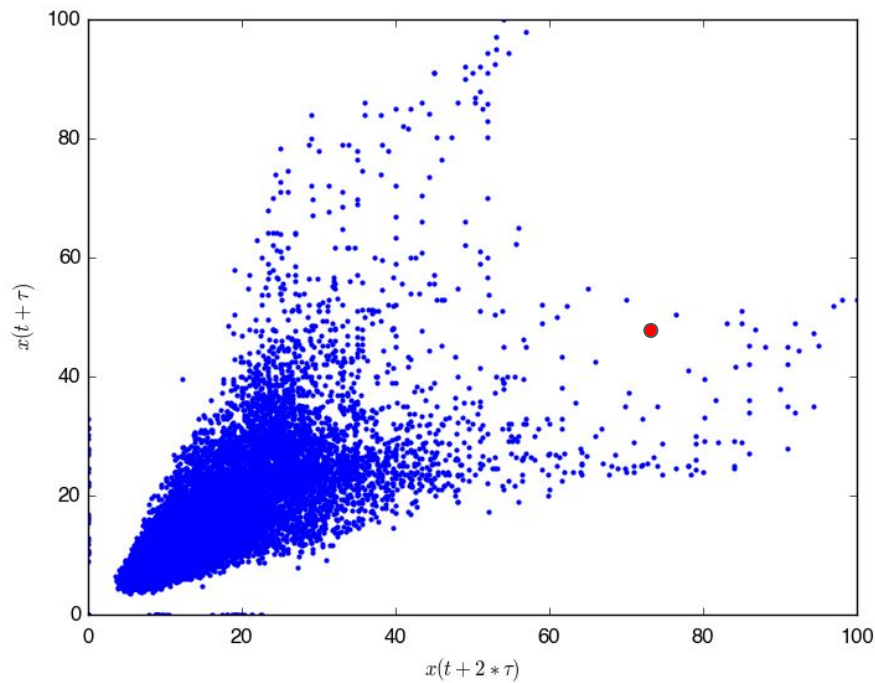


Review of Projection



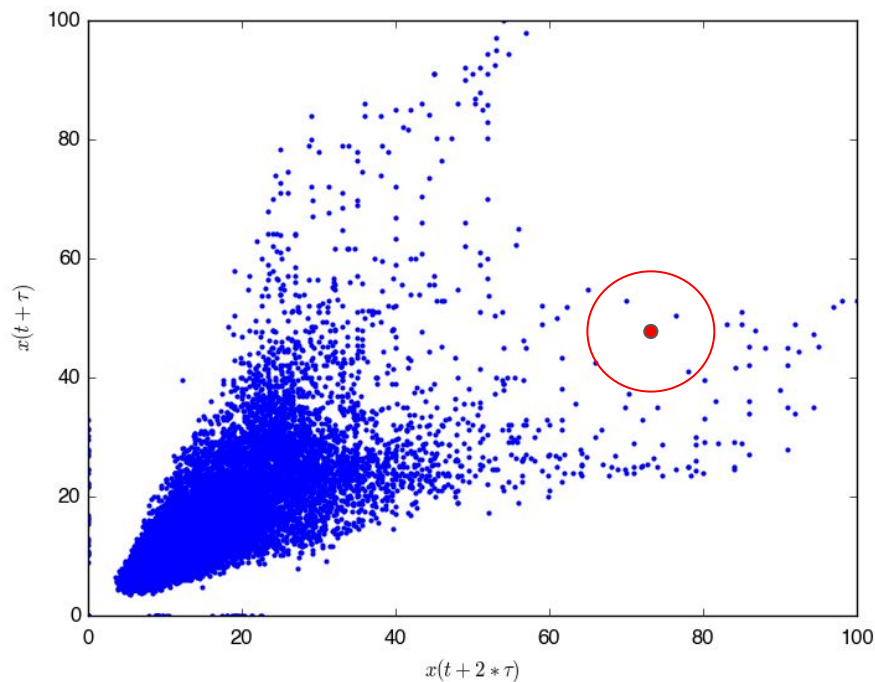
Prediction - Lorenz Method of Analogues (LMA)

K-Nearest Neighbors (KNN) in the embedding of the training data



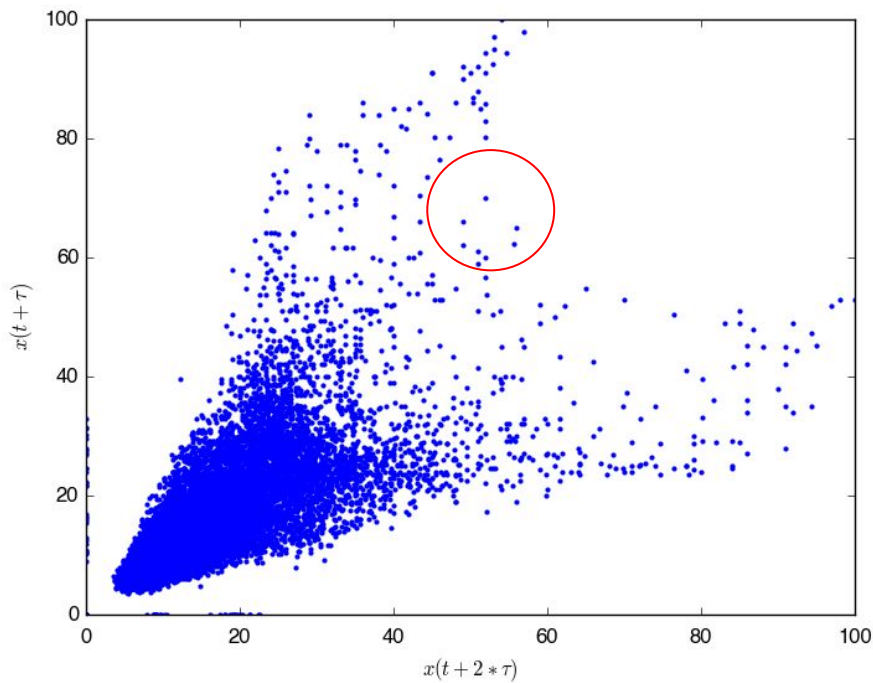
Prediction - Lorenz Method of Analogues (LMA)

K-Nearest Neighbors (KNN) in the embedding of the training data



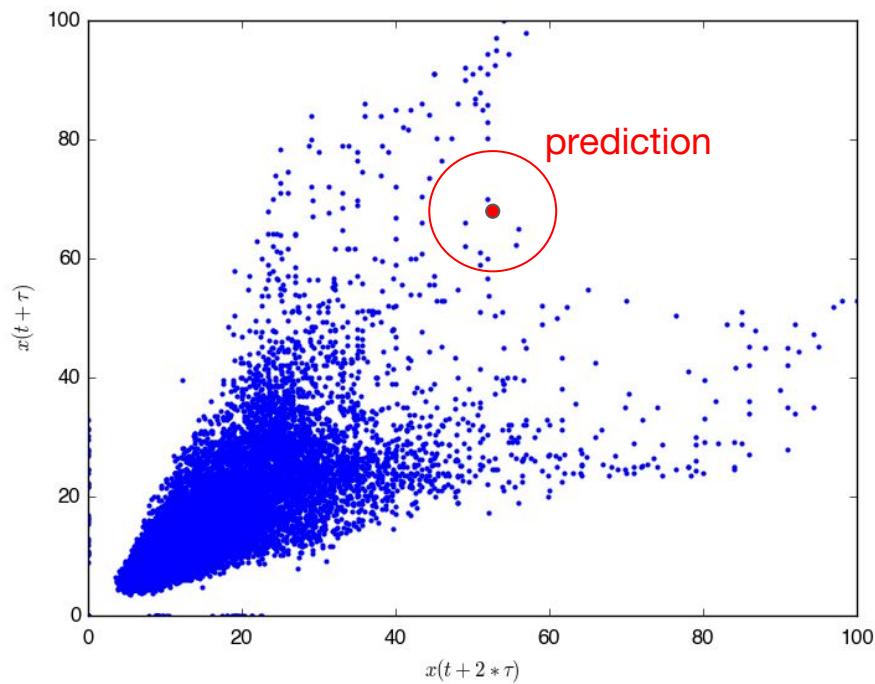
Prediction - Lorenz Method of Analogues (LMA)

K-Nearest Neighbors (KNN) in the embedding of the training data



Prediction - Lorenz Method of Analogues (LMA)

K-Nearest Neighbors (KNN) in the embedding of the training data



Evaluating Prediction Accuracy

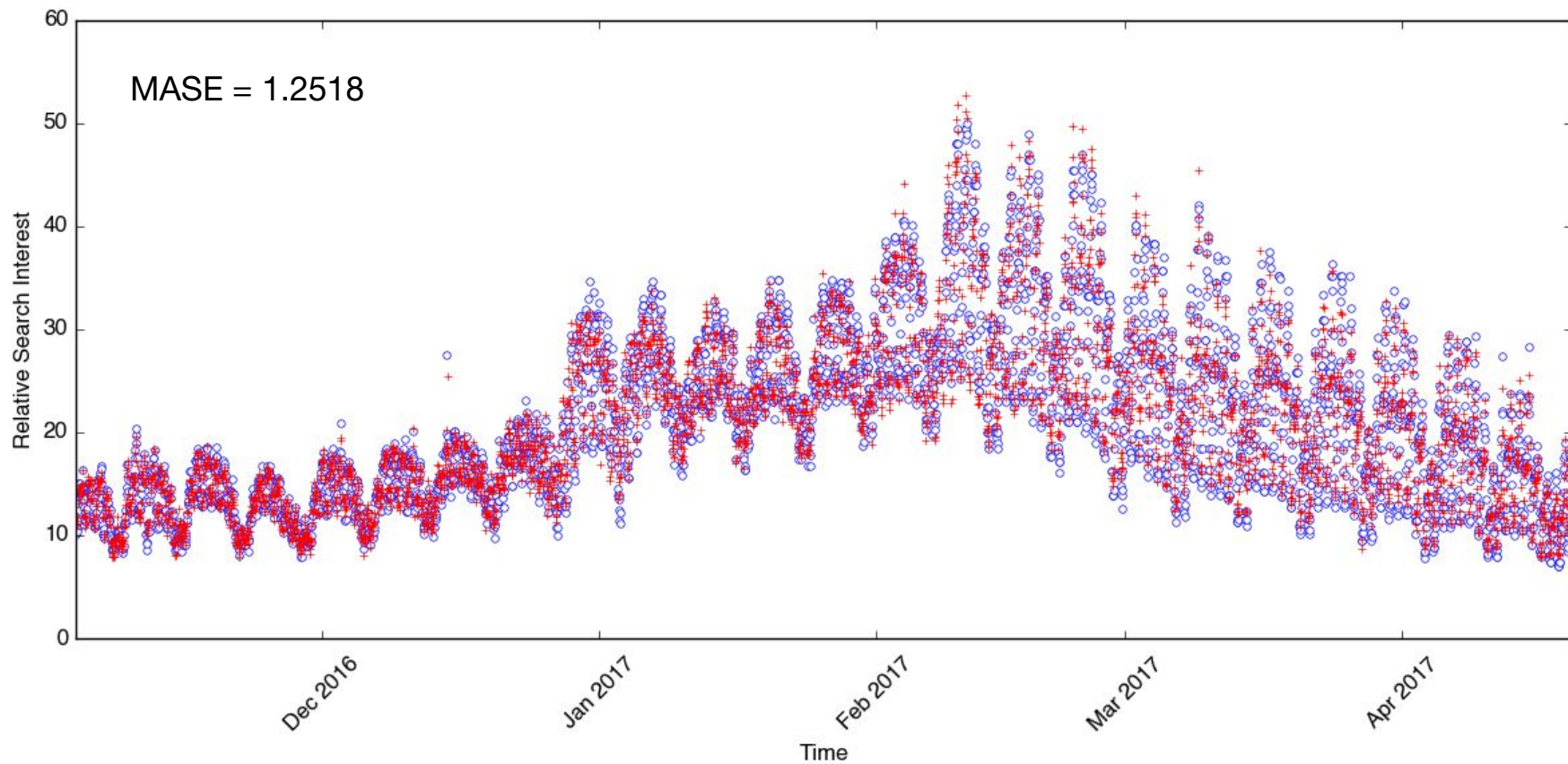
Used **mean absolute scaled error** (MASE), to evaluate our forecasts:

$$MASE = \frac{\sum_{j=n+1}^{k+n+1} |p_j - c_j|}{\frac{k}{n-1} \sum_{i=2}^n |x_i - x_{i-1}|}$$

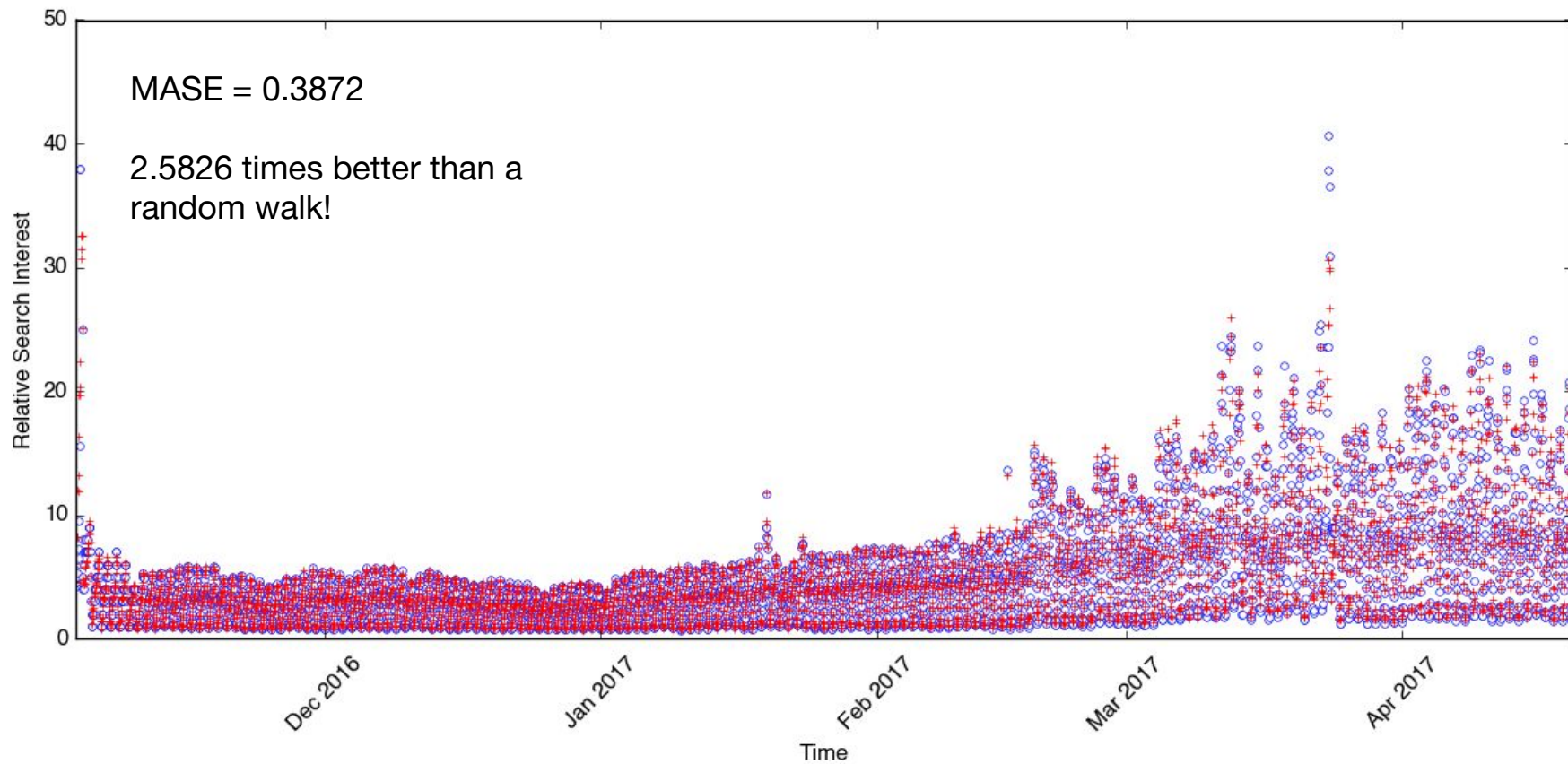
MASE < 1 - prediction error did better, on the average, than the in-sample error of a random-walk forecast

MASE > 1 - prediction method did worse, on average, than a random-walk

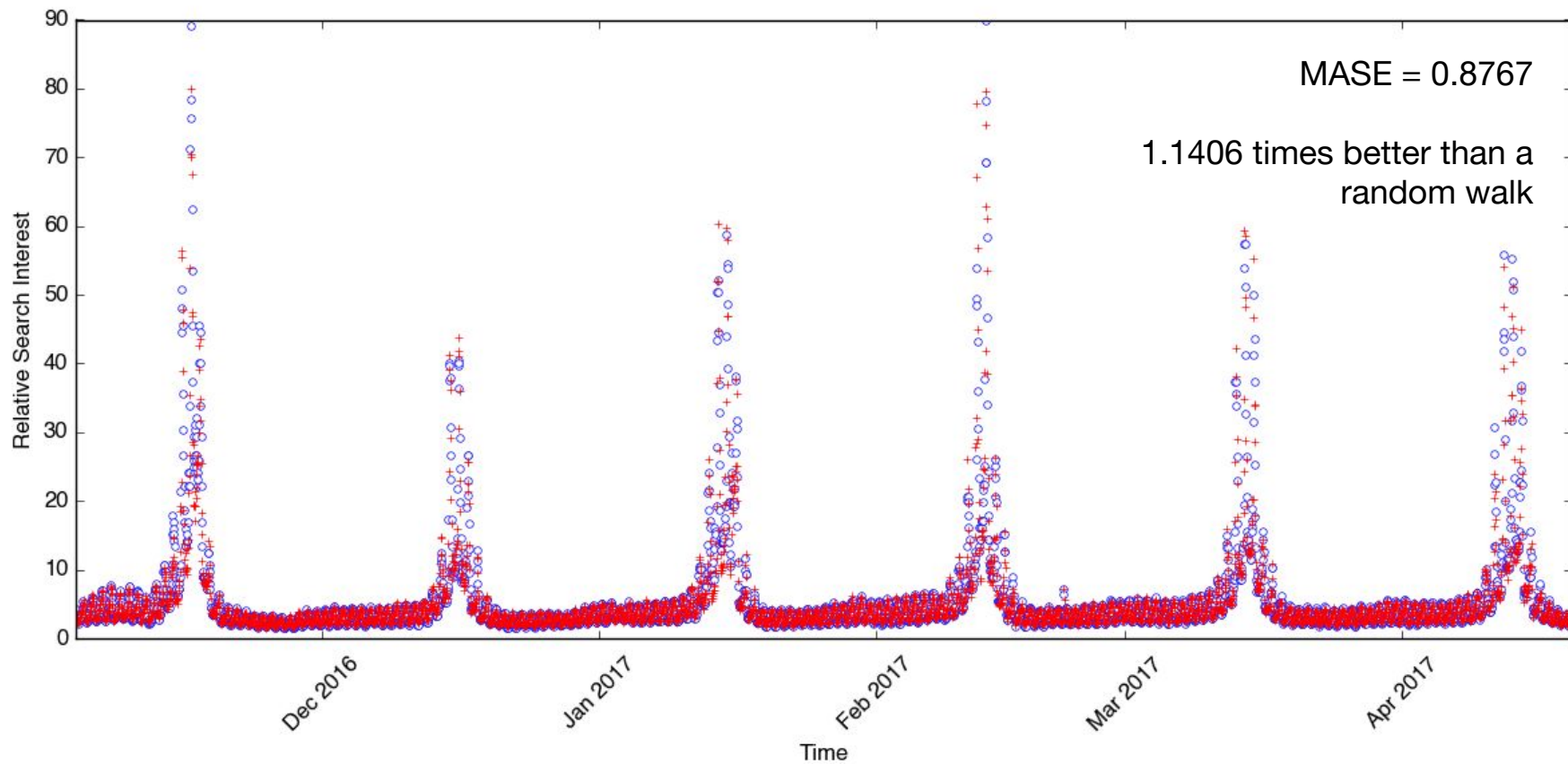
LMA Results - Influenza



LMA Results - Baseball



LMA Results - Full Moon



Conclusion

- Are Google search trends predictable?
 - We can predict “baseball” and “full moon” better than a random walk
- Are some trends more predictable than others?
 - Lower prediction error for search interest in “baseball” than “full moon” or “influenza”

Question: Why is WPE lowest for “full moon”, yet we do a better job at predicting “baseball”? Why are some search interest trends easier to predict?

Future Work: Other search query terms, multi-step forecasting

Works Cited

Weighted Permutation Entropy:

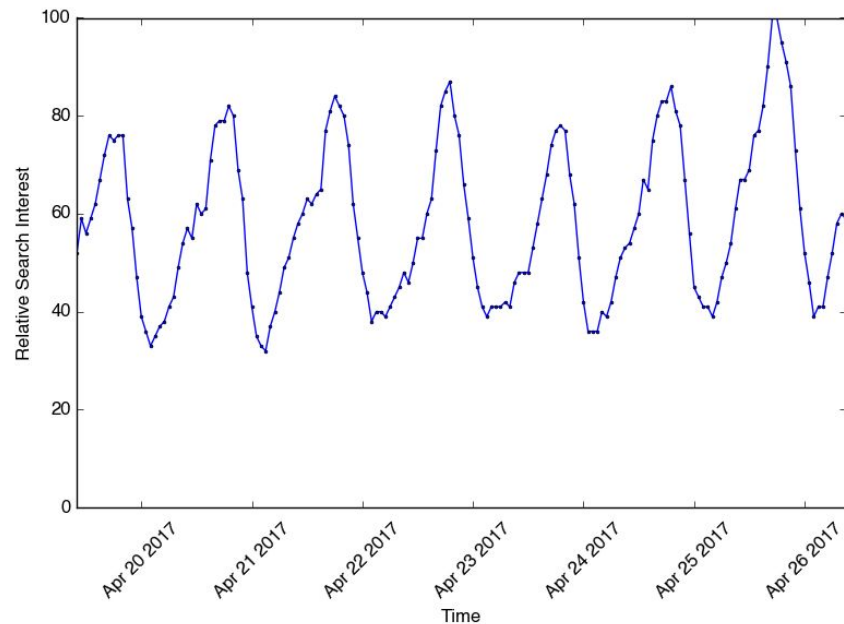
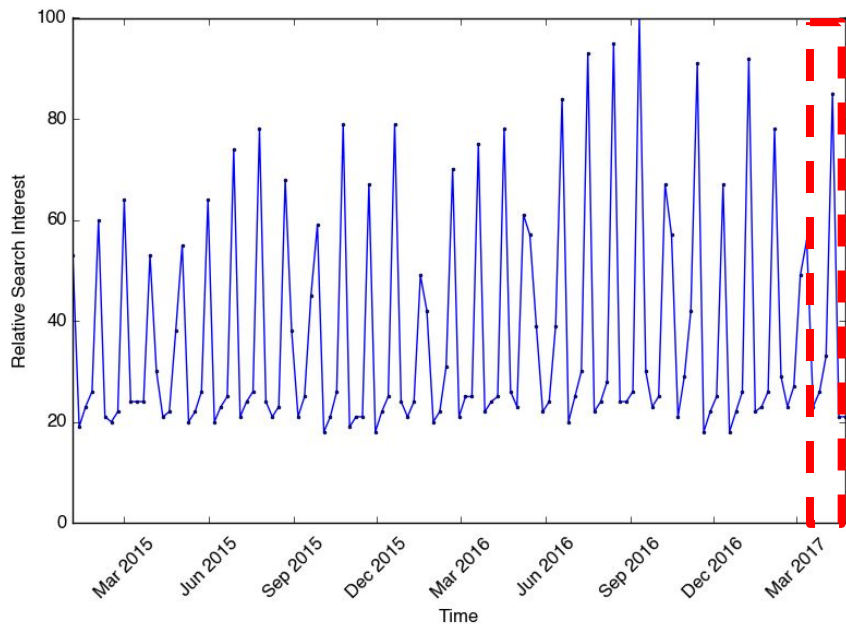
Joshua Garland. *Prediction in Projection: A new paradigm in delay-coordinate reconstruction*. PhD thesis, University of Colorado Boulder, 2016.

Bilal Fadallah, et. al. Weighted Permutation Entropy: A complexity measure for time series incorporating amplitude information. *Phys. Rev. E*. 2013.

Lorenz Method of Analogues:

Joshua Garland and Liz Bradley. Prediction in Projection. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2015.

Aside: Data Engineering



Aside: Data Engineering

Requested search interest on a weekly basis in order to get hourly resolution.

Stitched together windows of time by requesting the same point in both windows to find the scaling factor.

