QxBranch

Quantum Machine Learning for Election Modeling

April 4, 2018

Max Henderson, Ph.D.

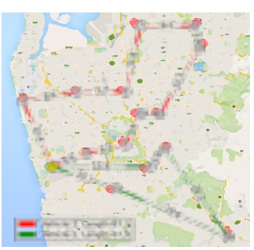
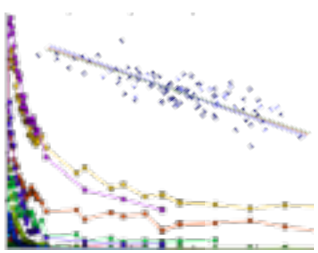
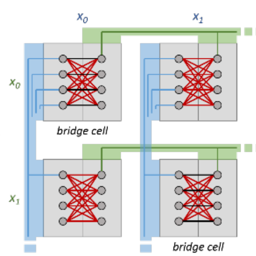
QxBranch Overview

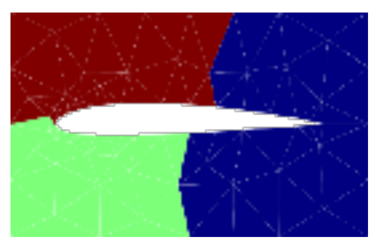
QxBranch delivers revolutionary data analytics software enabled by classical and emerging quantum computing capabilities that drive business value

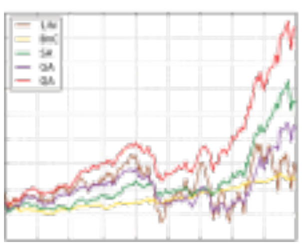
Data Analytics | Quantum Computing | Systems Engineering

- Established 2014 in Washington D.C. / London / Adelaide
- Team of ~20 software and systems engineers, data scientists
- Clients:
  - Global Investment Banks
  - Asset Management Firms
  - Technology Companies
  - Government
  - Energy
  - Pharmaceutical

- Apply data analytics expertise and software capabilities to manage complex data and provide actionable insights across multiple verticals
- Business domain expertise in finance, aerospace, defence, and technology domains
- Research & Development partnerships with clients and academia to identify business challenges that can be solved through cutting-edge applications of quantum computing (universal and adiabatic) and advanced data analytics






Election 2016: Case study in the difficultly of sampling

Survey finds Hillary Clinton has 'more than 99% chance' of winning election over Donald Trump

The Princeton Election Consortium found Ms Clinton has a projected 312 electoral votes across the country and only 270 are needed to win

Rachael Revesz New York | @RachaelRevesz | Saturday 5 November 2016 16:44 GMT | 106 comments

Where did the models go wrong?

ELECTION2016

FORECAST

PRESIDENT SENATE

By Natalie Jackson and Adam Hooper
Additional design by Alissa Scheller

PUBLISHED MONDAY, OCT. 3, 2016 12:56 P.M. EDT
UPDATED TUESDAY, NOV. 8, 2016, 12:43 A.M. EST

Photos: Getty

In the event of a tie, the newly elected House of Representatives will elect the president, and the newly elected Senate will elect the vice president.

CNN Money U.S. + Business Markets Tech Media Personal Finance Sma

A model that has correctly predicted the winner of every U.S. presidential race since Ronald Reagan in 1980 is forecasting a big victory for Hillary Clinton.

Clinton is expected to get 332 electoral votes, while Trump is predicted to get just 206, according to the Moody's Analytics model, which is based on three economic and three political factors.

QxBranch

State-by-state correlations

- Major issue: failure to model **correlations**[1-3] between states
- Most models assumed independence between results of each state
- An accurate correlation matrix can capture higher-level, richer structure in the data and account for systemic errors in polls

First, there are **errors of analysis**. As an example, if you had a model of last year's election that concluded that Clinton had a 95 or 99 percent chance of winning, you committed an analytical error. [4] Models that expressed that much confidence in her chances had a host of technical flaws, such as ignoring the correlations in outcomes between states. [5]

Similar states usually have similar outcomes
Correlation matrix after 20,000 simulations, polls-only model, June 27, 2016

	Ala.	Calif.	Fla.	Minn.	N.C.	N.M.	R.I.	Wis.
Alabama		.60	.61	.53	.72	.54	.41	.55
California	.60		.73	.67	.69	.80	.61	.68
Florida	.61	.73		.67	.75	.70	.63	.76
Minnesota	.53	.67	.67		.68	.58	.64	.84
N. Carolina	.72	.69	.75	.68		.60	.53	.67
New Mexico	.54	.80	.70	.58	.60		.54	.64
Rhode Island	.41	.61	.63	.64	.53	.54		.69
Wisconsin	.55	.68	.76	.84	.67	.64	.69	

1. http://www.independent.co.uk/news/world/americas/sam-wang-princeton-election-consortium-poll-hillary-clinton-donald-trump-victory-a7399671.html
2. http://elections.huffingtonpost.com/2016/forecast/president
3. http://money.cnn.com/2016/11/01/news/economy/hillary-clinton-win-forecast-moodys-analytics/index.html
4. http://fivethirtyeight.com/

Difficulty of sampling from correlated graphs

- Even with perfect data on correlations between states, using the correlation matrix is difficult due to the computational cost of sampling from fully-connected graphs

- Sampling from fully-connected graphs is analogous to sampling from a properly trained Boltzmann machine
    - Training coefficients of Boltzmann machines requires performing calculations on all possible states of the model
    - As this is intractable on large problem sizes, heuristics or other models are typically implemented instead

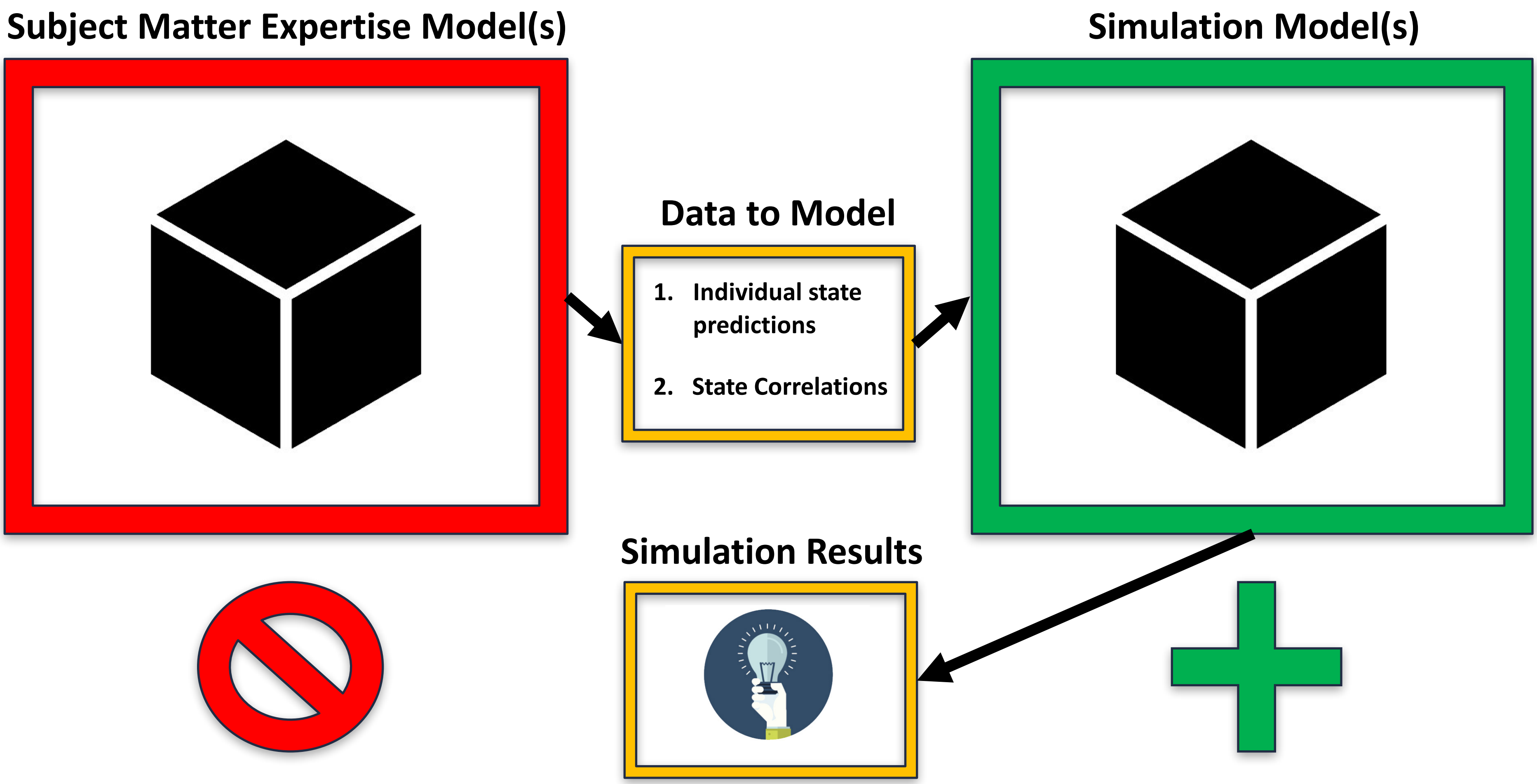
Forecasting elections on a quantum computer

- Quantum computing (QC) research has shown potential speedups in training deep neural networks[1-3]

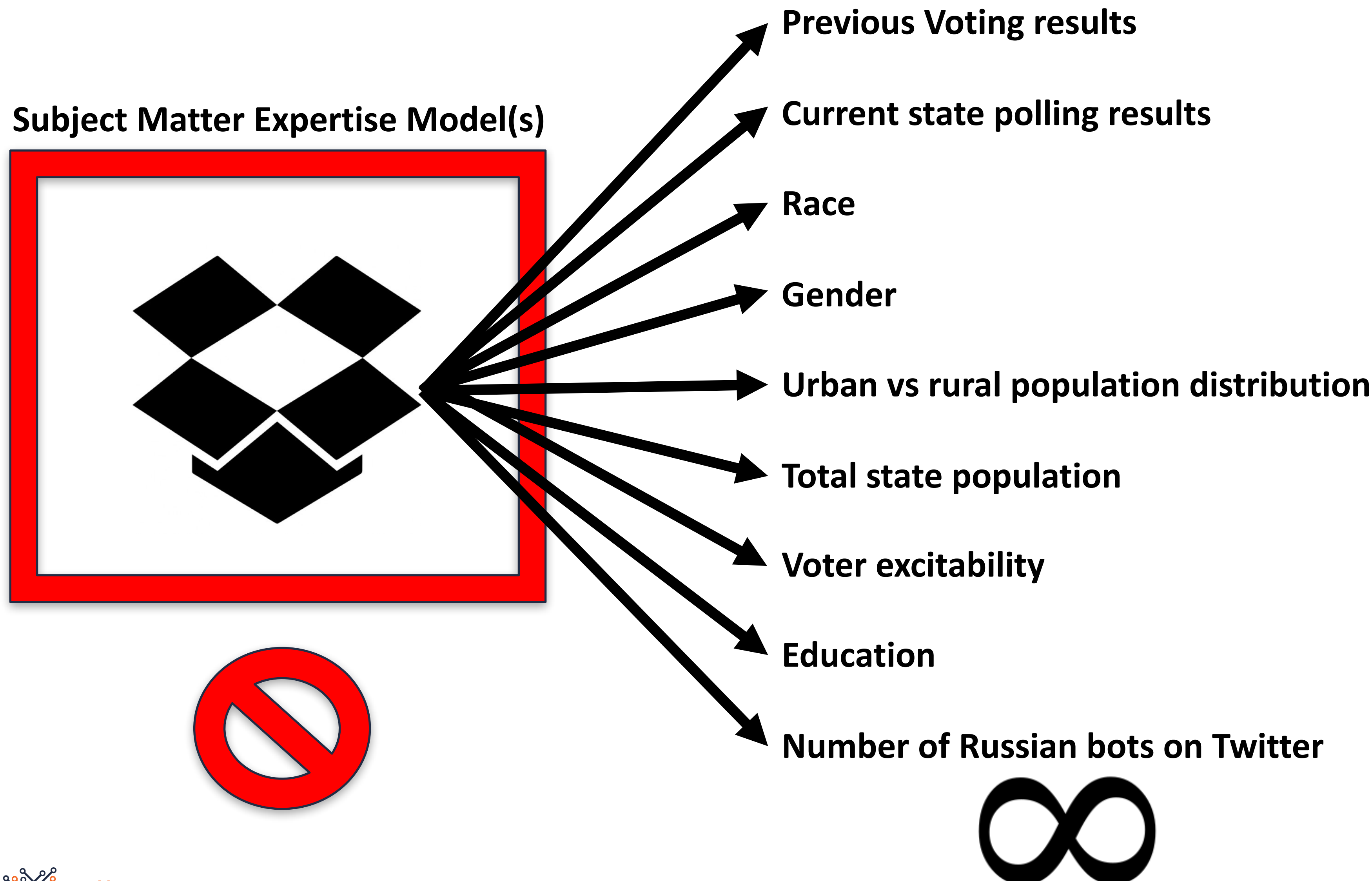
- Core idea: By using QC-trained models to simulate election results we can achieve:
  - More efficient sampling / training
  - Intrinsic, tuneable state correlations
  - Inclusion of additional error models

1. Adachi, Steven H., and Maxwell P. Henderson. "Application of quantum annealing to training of deep neural networks." *arXiv preprint arXiv:1510.06356* (2015).
2. Benedetti, Marcello, et al. "Estimation of effective temperatures in quantum annealers for sampling applications: A case study with possible applications in deep learning." *Physical Review A* 94.2 (2016): 022308.
3. Benedetti, Marcello, et al. "Quantum-assisted learning of graphical models with arbitrary pairwise connectivity." *arXiv preprint arXiv:1609.02542* (2016).

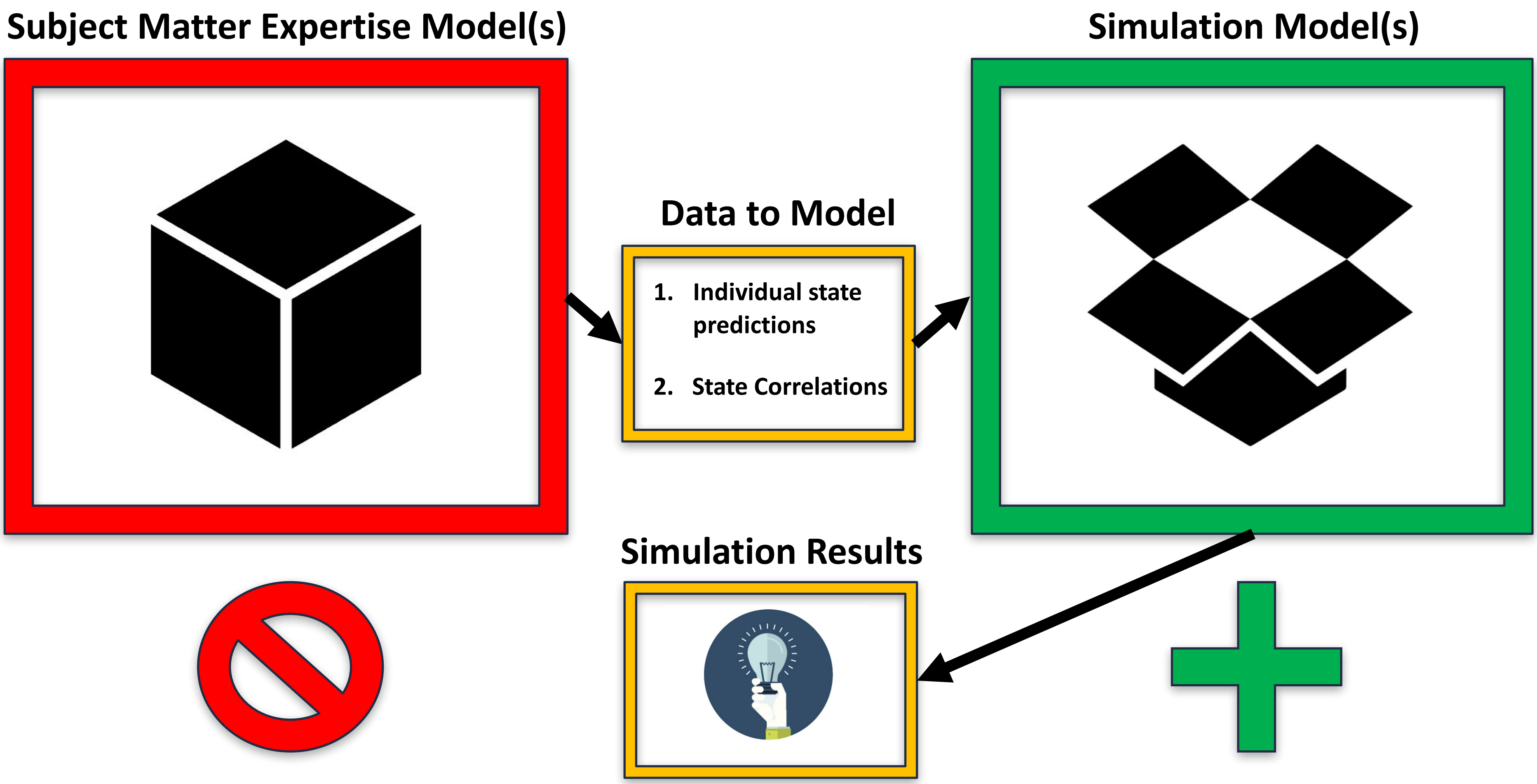
What we ARE doing vs. what we AREN'T



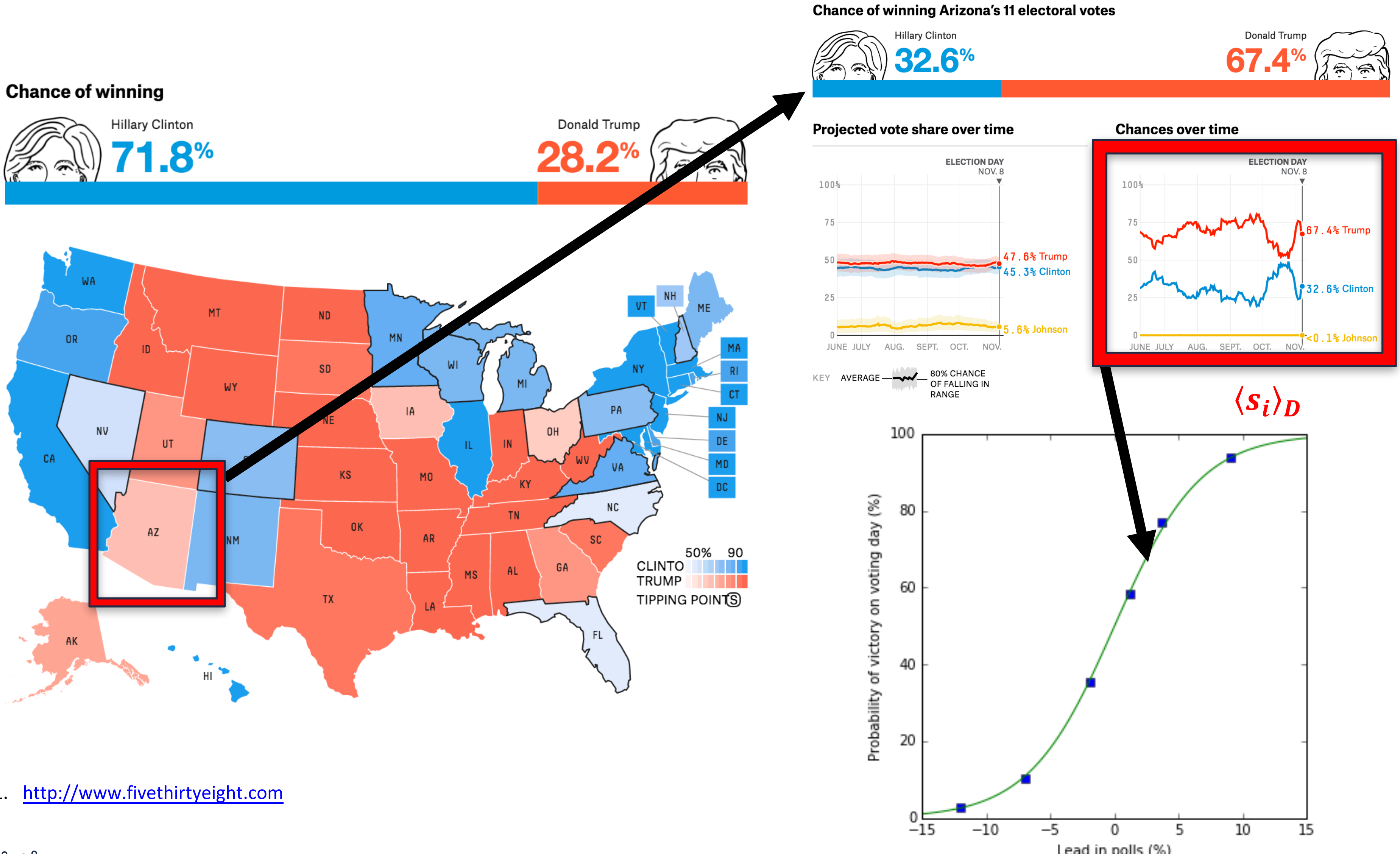
What we ARE doing vs. what we AREN'T



What we ARE doing vs. what we AREN'T



Step 1: Mapping an election to a Boltzmann machine



1. http://www.fivethirtyeight.com

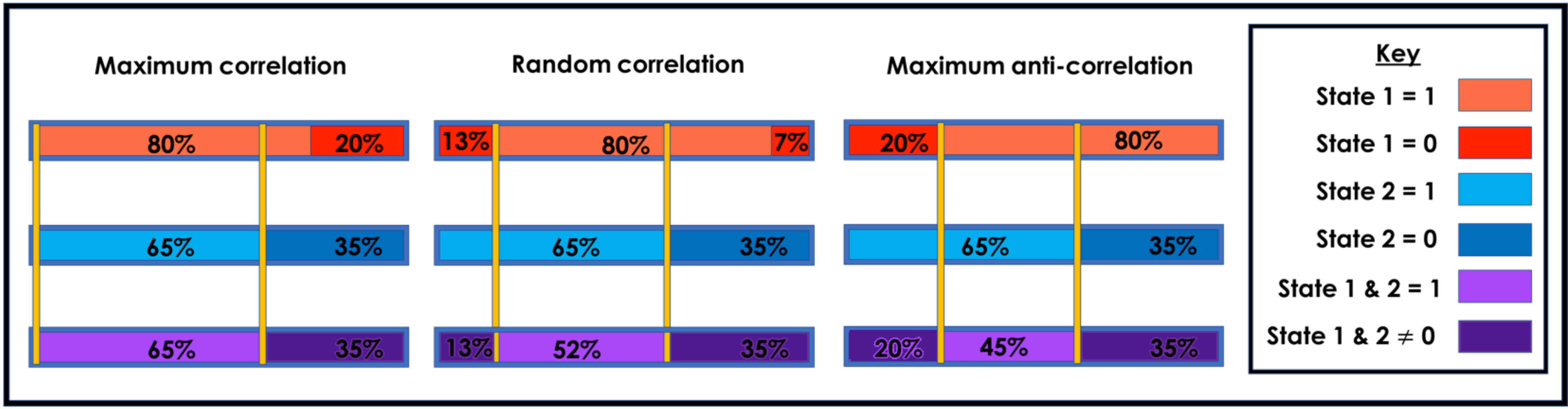
Available data is limited

- What we would like:
  - Detailed breakdowns of demographics
  - Meticulously curated biases and correlations
  - All of the data that 538 has spent years and thousands of dollars curating

- What we have:
  - Publicly available results of previous US elections
  - State probabilities, as told by polls
  - Publicly accessible data from 538

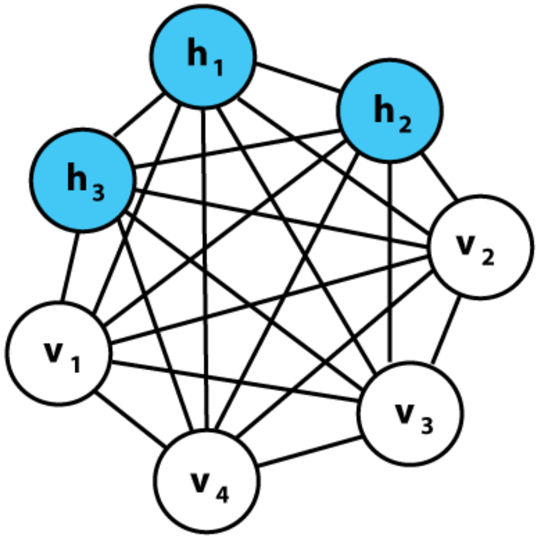
Calculating the missing second order moments

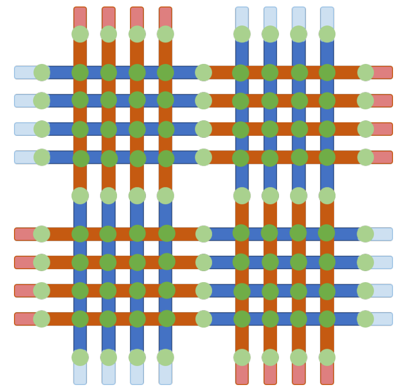
- In lieu of better curated data concerning second order moments, we calculated our own terms from previous US election results
- Our methodology should not "break" first order moments

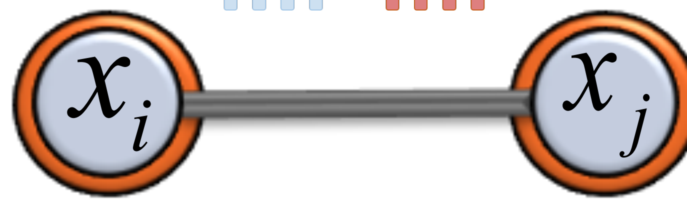


- Assumptions in this model:
  - In each previous election, if two states had the same election result, that increased their correlation
  - Elections that were more recent have a higher weight

Step 2: Mapping a Boltzmann machine to the QC







$$E = -\left(\sum_{i<j} w_{ij}\, s_i\, s_j + \sum_i \theta_i\, s_i\right)$$

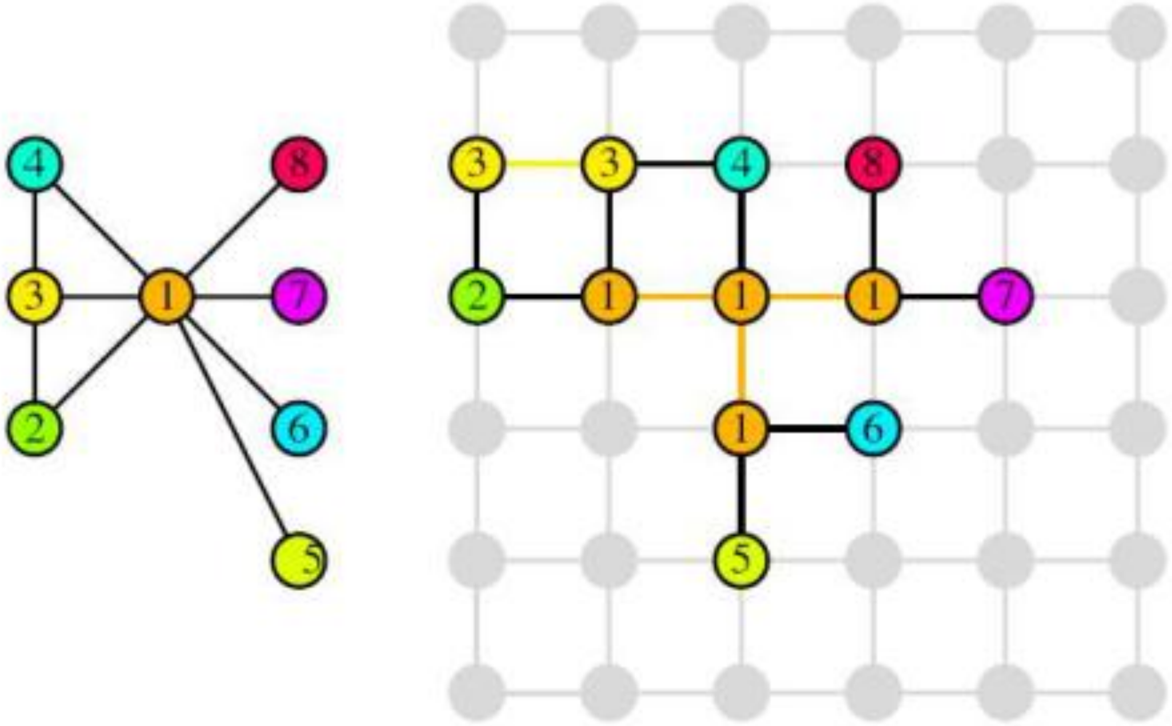
$$H = \sum h x_i + \sum J x_i x_j$$

The update equations for training the model: → **Potential quantum advantage**

$$\Delta w_{ij} = -\frac{1}{\eta}\left(\langle s_i s_j \rangle_D \boxed{- \langle s_i s_j \rangle_M}\right)$$

$$\Delta \theta_i = -\frac{1}{\eta}\left(\langle s_i \rangle_D - \langle s_i \rangle_M\right)$$

Graph embedding – Qubit chains

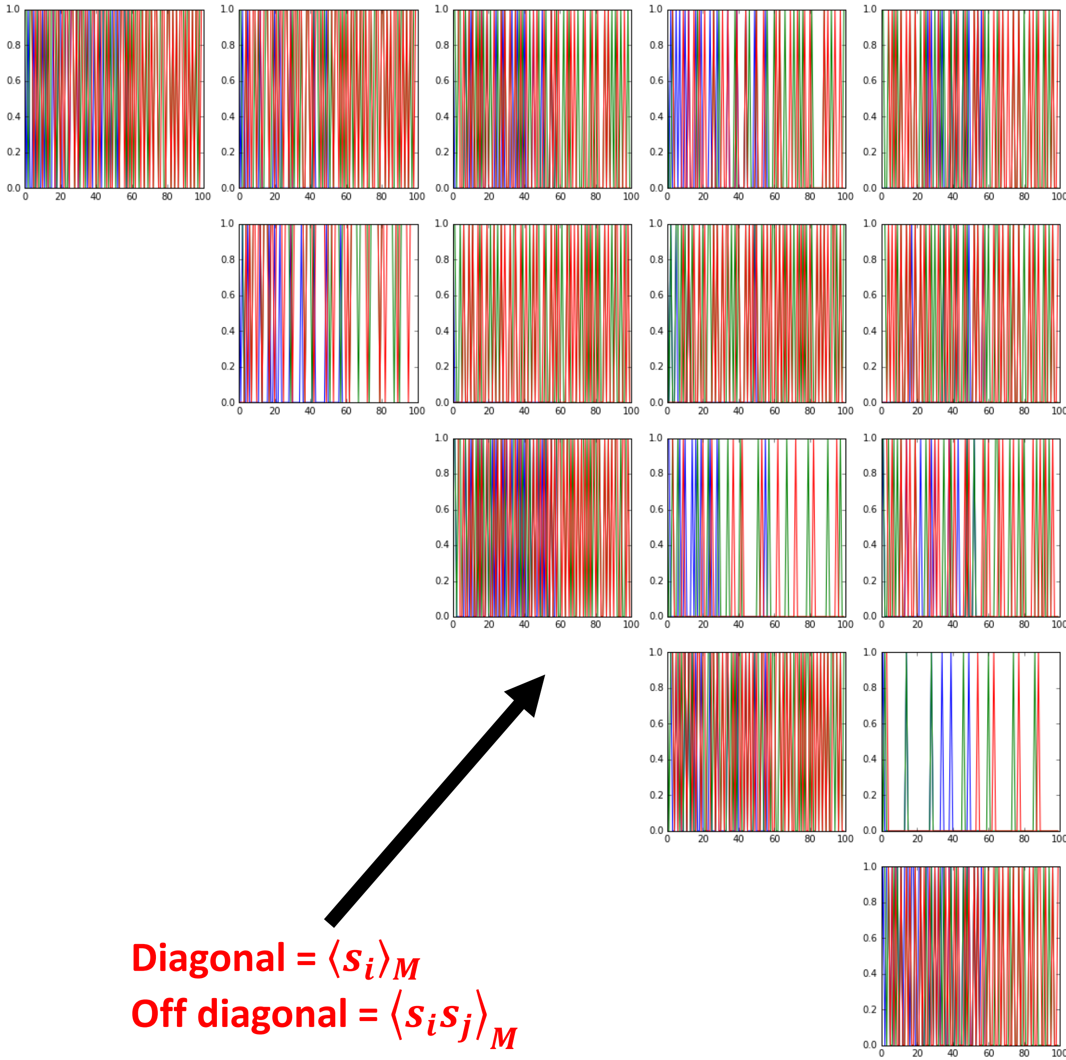


Example of embedding a problem (left) into a fixed graph structure (right)[1]

1. Choi, Vicky. "Minor-embedding in adiabatic quantum computation: II. Minor-universal graph design." *arXiv preprint https://arxiv.org/pdf/1001.3116v2.pdf* (2010).

Effect of embedding: Short qubit chains

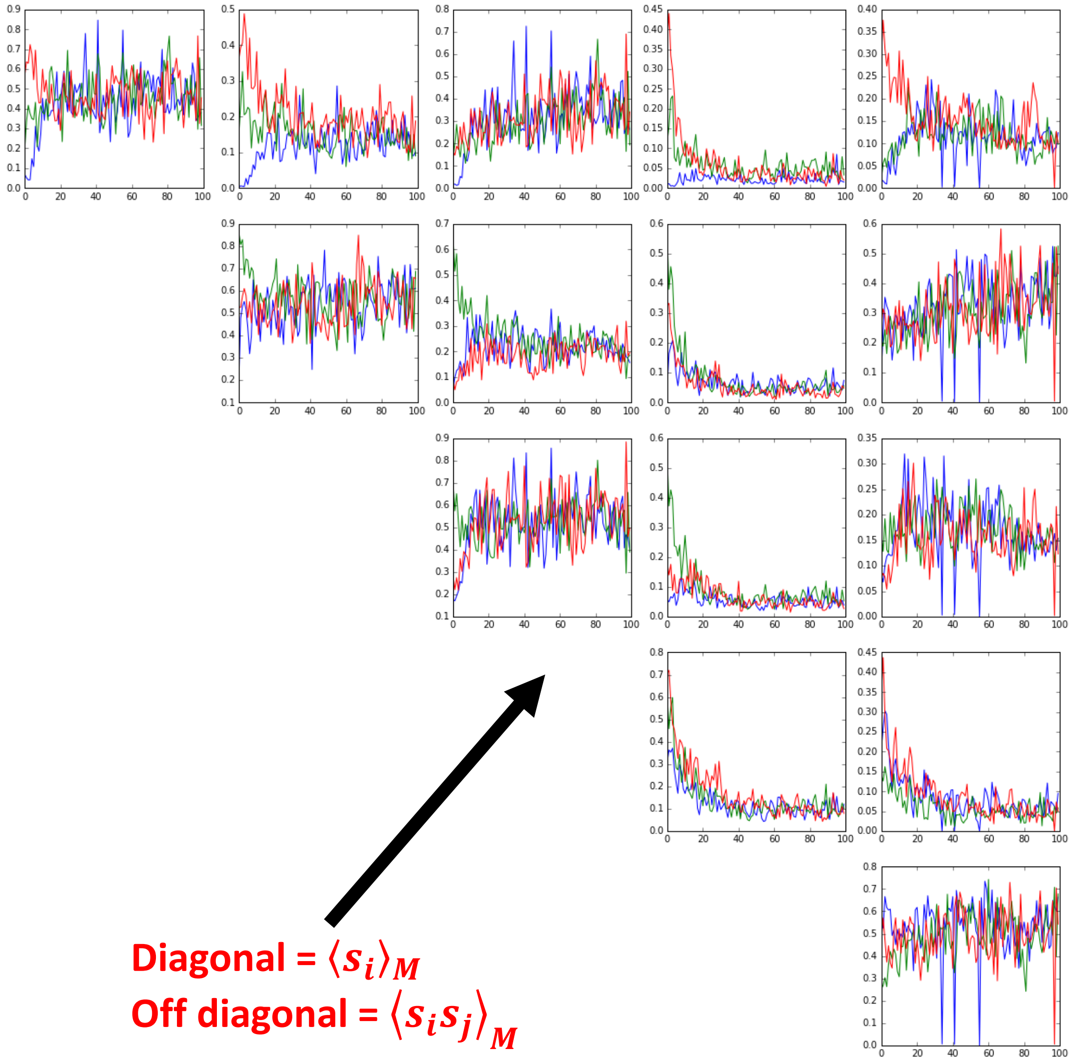
- To validate the approach, we randomly chose first and second order terms for a hypothetical 5-state nation
- Using the smallest embedding chains, this network was unable to properly train
  - “Hopfield” like results; optimal solutions rather than probabilistic results
  - Leads to huge changes in weights/biases, causing network instability



Diagonal = $\langle s_i \rangle_M$
Off diagonal = $\langle s_i s_j \rangle_M$

Effect of embedding: Long qubit chains

- For larger problem sizes, the embedding will necessarily have longer qubit chains
- To simulate this for our small network, we artificially increased the qubit chains
- With this approach, arbitrary first and second order moments were learned by the networks

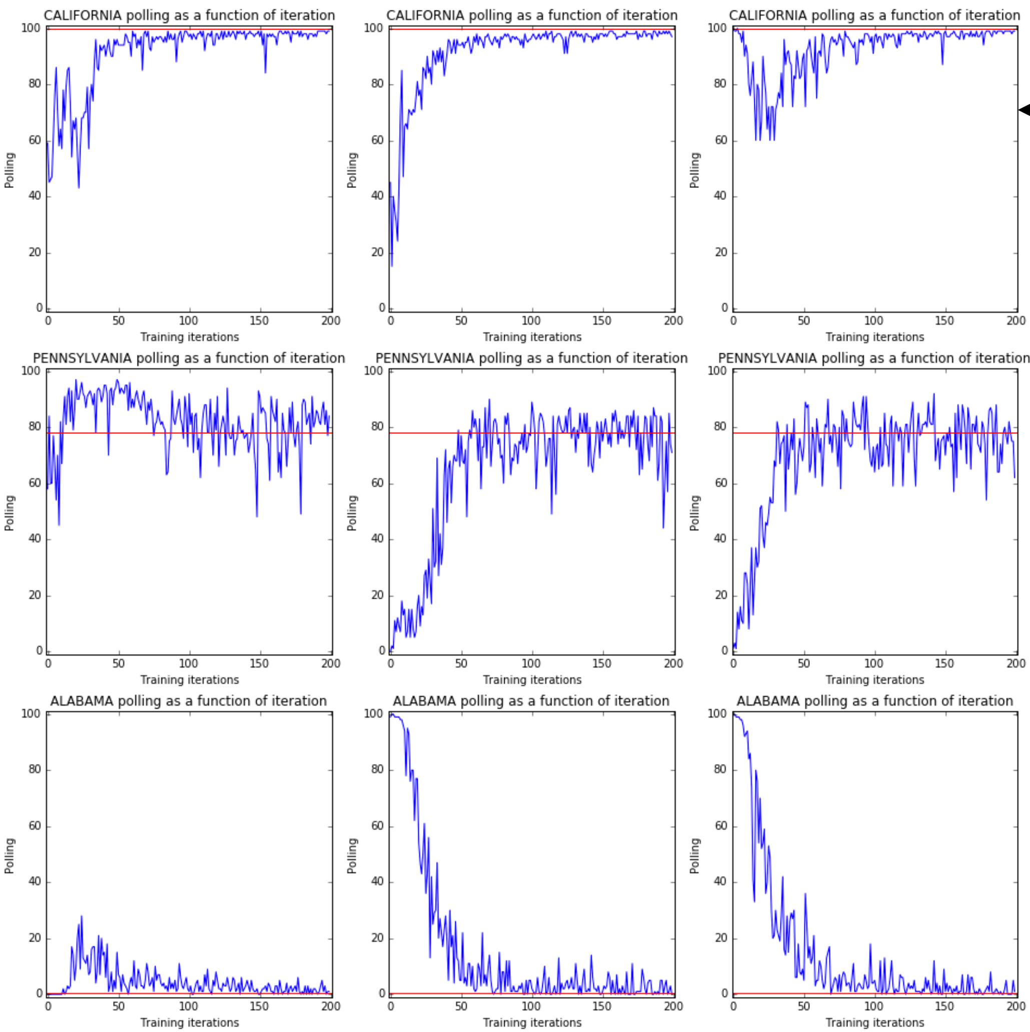


Diagonal = $\langle s_i \rangle_M$
Off diagonal = $\langle s_i s_j \rangle_M$

Primary experiment

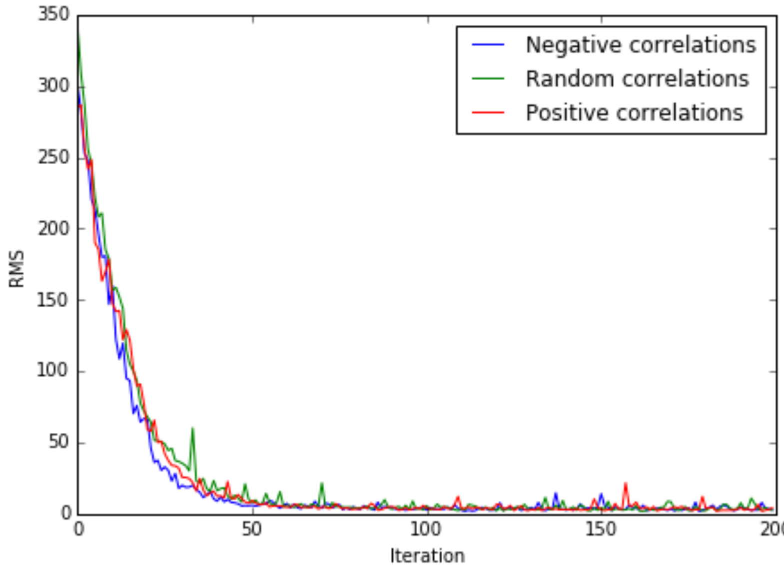
- Goal: Using historical data and the QC-training methodology presented here, reproduce election forecasts over time
- Some caveats:
  - Multiple models needed for modeling national error; 25 were used here
  - Limited time windows of D-Wave access, so results were generated every two weeks instead of daily
  - Limited hardware size made us omit 1 state and province (sorry Maryland and DC… you always vote D anyway)
  - For simplification, Maine and Nebraska were considered winner-take-all

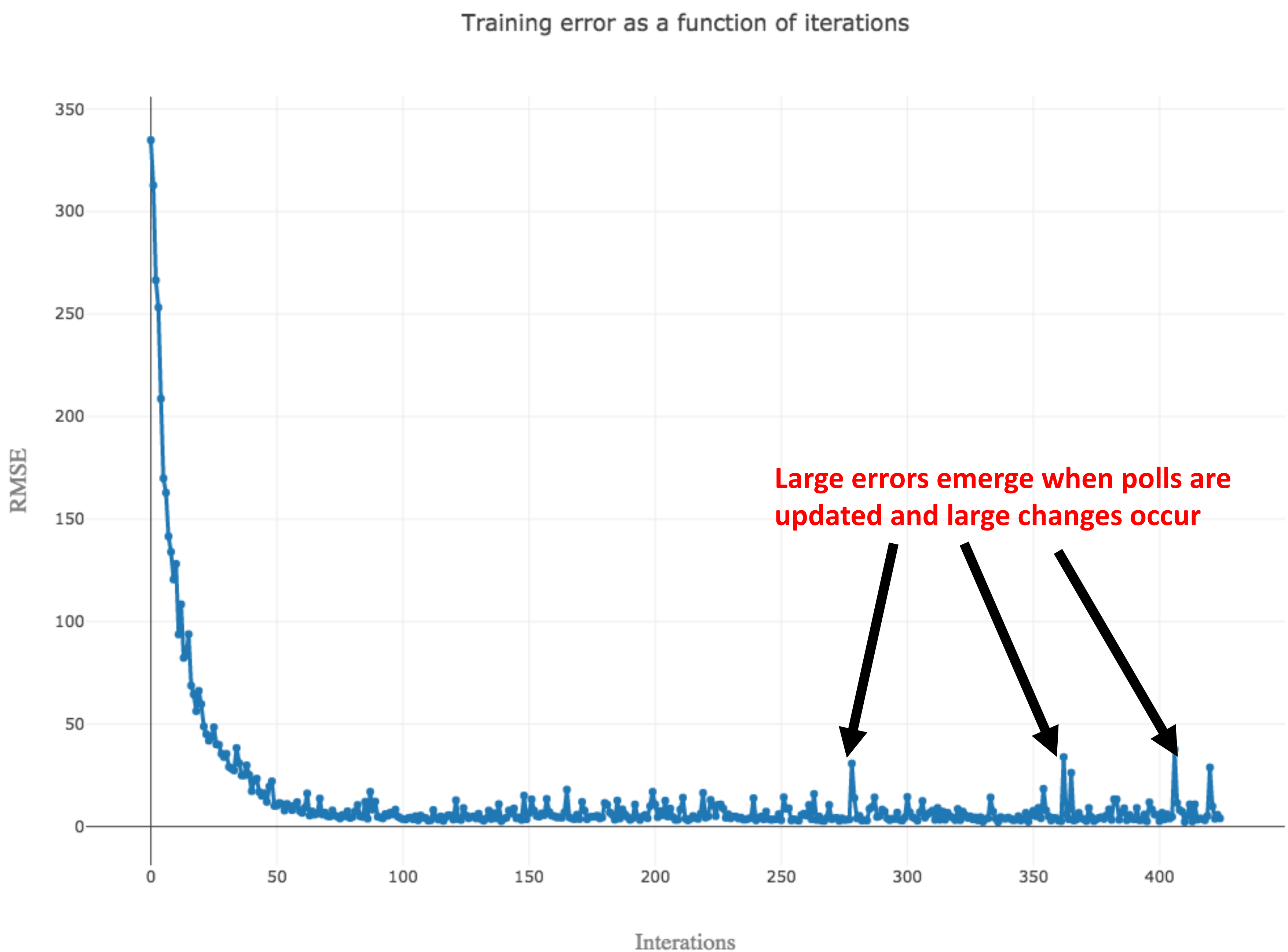
Results – Training errors



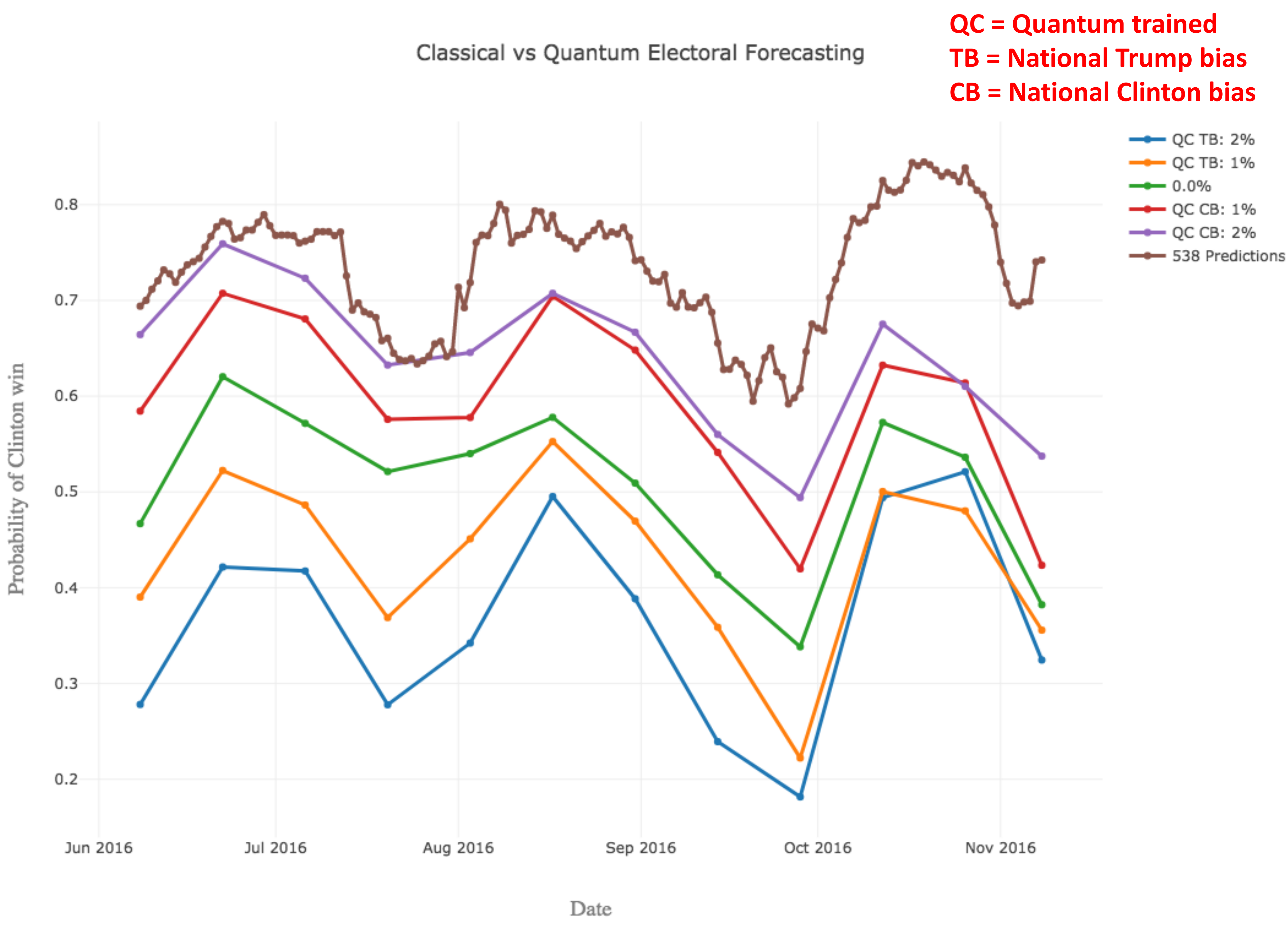
Red lines = $\langle s_i \rangle_D$
Blue lines = $\langle s_i \rangle_M$

Examples testing extremes of correlations: negative, random, & positive





Training error as a function of iterations
RMSE
Interations
Large errors emerge when polls are updated and large changes occur
0
50
100
150
200
250
300
350
400



QC = Quantum trained
TB = National Trump bias
CB = National Clinton bias
Classical vs Quantum Electoral Forecasting
QC TB: 2%
QC TB: 1%
0.0%
QC CB: 1%
QC CB: 2%
538 Predictions
Probability of Clinton win
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Jun 2016
Jul 2016
Aug 2016
Sep 2016
Oct 2016
Nov 2016
Date

The most "impactful" states

- Pearson correlation coefficients for the 10 states most (top) and least (bottom) correlated with the election forecasting results

Our models

State	Correlation coefficients
Ohio	0.204
Florida	0.163
Nevada	0.178
New Hampshire	0.167
Pennsylvania	0.155
Iowa	0.152
Michigan	0.145
North Carolina	0.137
Colorado	0.130
Arizona	0.127
Illinois	0.002
Nebraska	0.004
Alabama	0.005
Oklahoma	0.006
California	0.008
West Virginia	0.008
Delaware	0.008
Oregon	0.009
Idaho	0.015
Arkansas	0.016

538

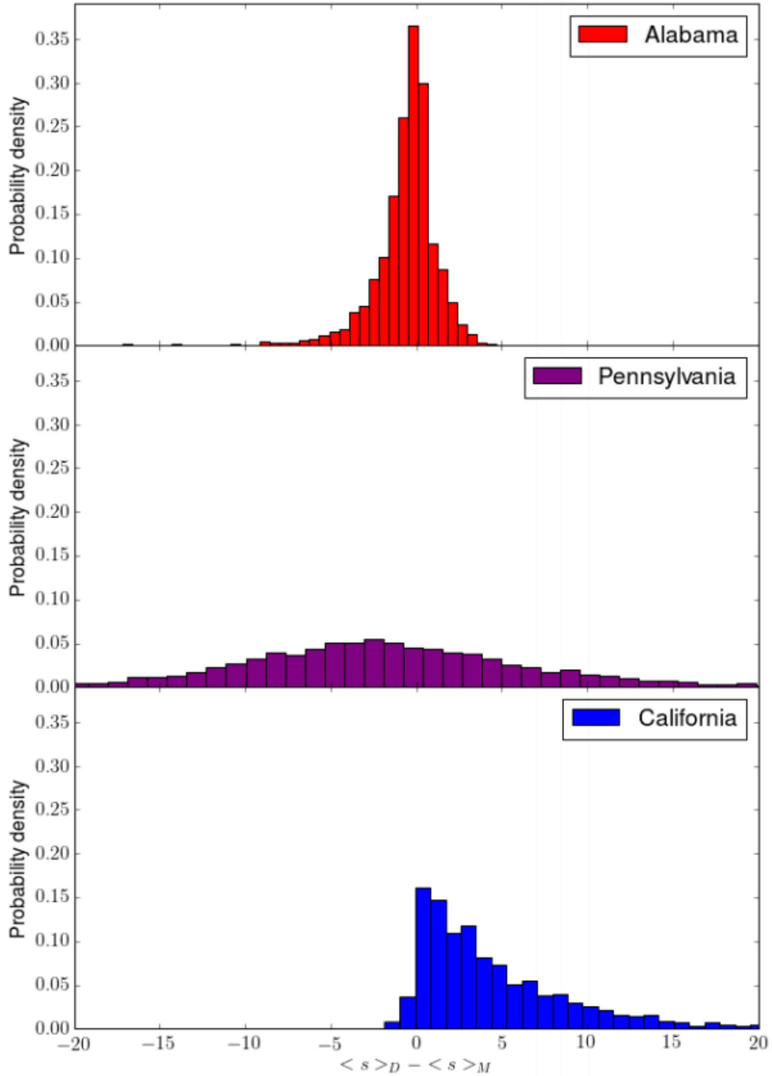
Florida	17.6%
Pennsylvania	12.3
Michigan	11.7
North Carolina	11.2
Virginia	6.0
Colorado	6.0
Ohio	5.2
Wisconsin	4.8
Minnesota	3.8
Nevada	3.7

Alabama	<0.1
California	<0.1
North Dakota	<0.1
Massachusetts	<0.1
Hawaii	<0.1
Maryland	<0.1
Oklahoma	<0.1
West Virginia	<0.1
Vermont	<0.1
Wyoming	<0.1
Nebraska 3rd District	<0.1
District of Columbia	<0.1

State errors

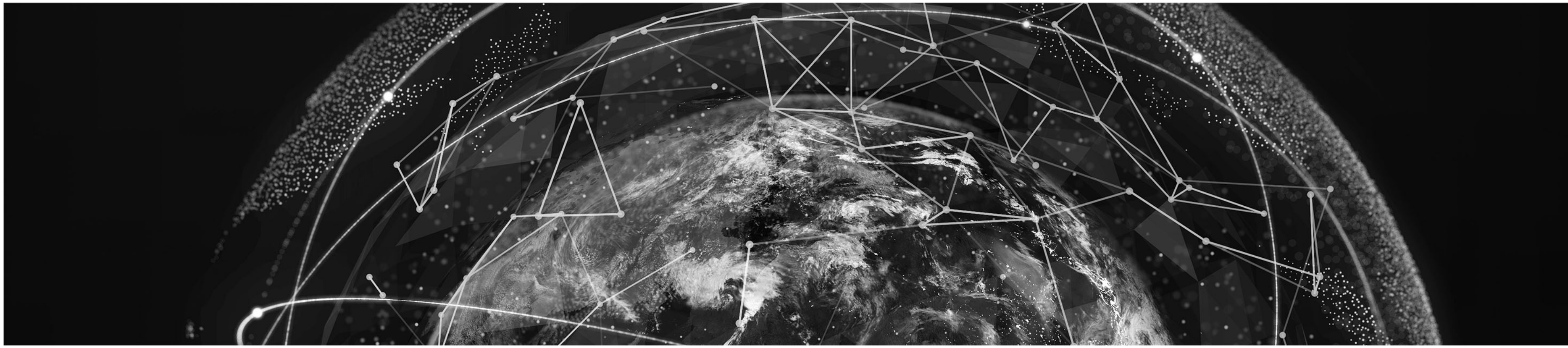
- Individual states error distributions was highly dependent on if the state was a hard red, blue, or purple state

- Different ways of dealing with errors of this form:
  - Shimming
  - Multiple gauges



Summary

- The QC-trained networks were able to learn structure in polling data to make election forecasts in line with the models of 538
- Trump was given a higher likelihood of victory (compared to other pollsters), even though the first order moments remained unchanged
    - Ideally in the future, we could rerun this method using correlations known with more detail in-house from 538
- Each iteration of the training model quickly produced 25,000 simulations (one for each national error model), which eclipses the 20,000 simulations 538 performs each time they rerun their models



QxBranch

www.qxbranch.com | @qxbranch