

ML Applications @ZeD Labs

ishanu chattopadhyay

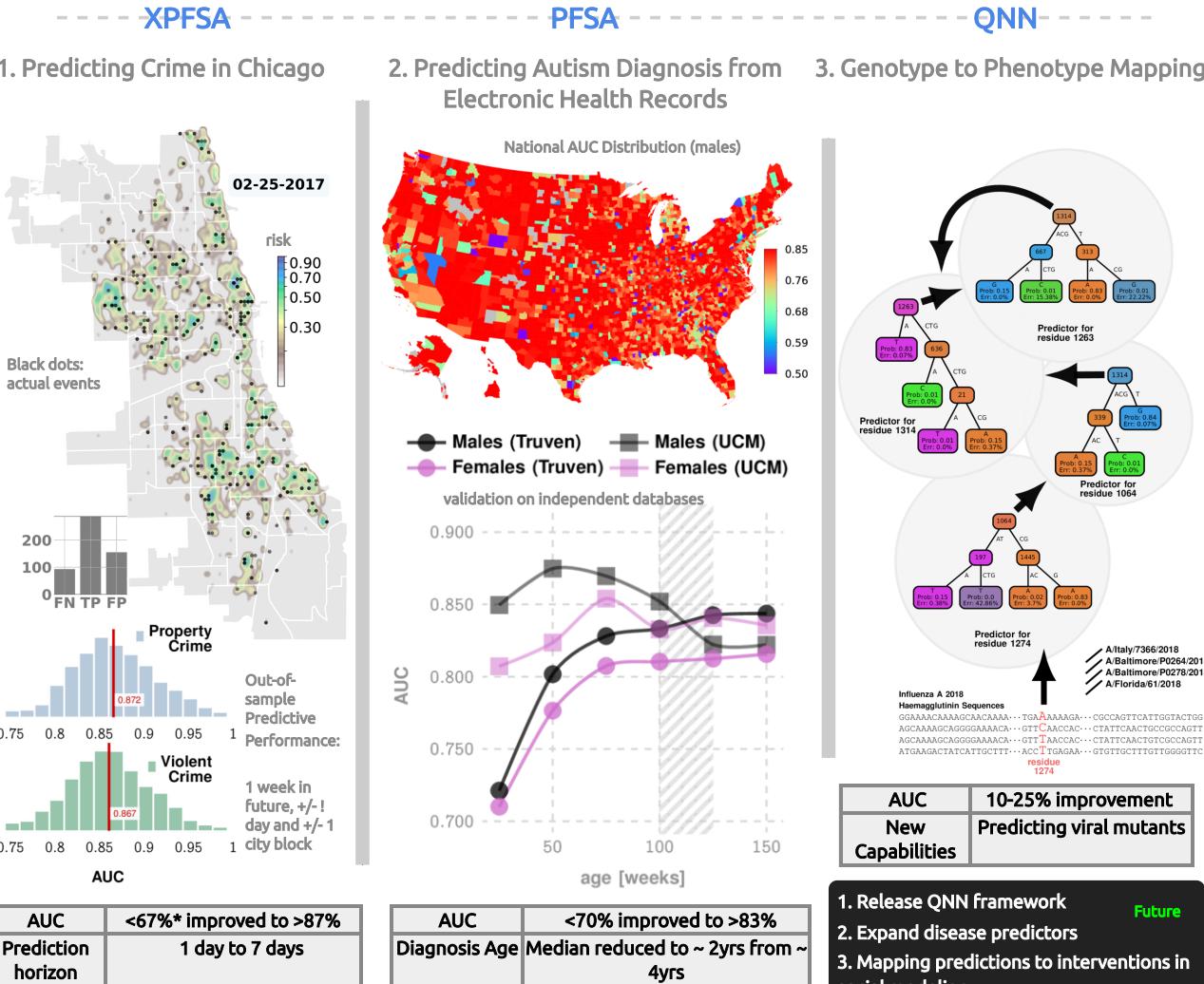
CCTS 40500

THE UNIVERSITY OF
CHICAGO



Hypothesis: Decomposing the learning problem, to first estimate unknown problem structure, followed by standard inference leads to faster, better, efficient learning

Key Insight: Notion of structure changes between problem domains, but basic idea remains valid



Publications

- "Beyond edit distance: A function aware adaptive distance metric for evolving populations", T. Li, P. Leon, A. Weinberger, and I. Chattopadhyay **In Preparation**
- "QNN: Using Quasi-species Sequence Variations to Inform Neural Net Modeling of The Genotype to Phenotype Problem", T. Li, I. Chattopadhyay **In preparation**
- "Zero-burden Digital Biomarkers For Autism: Exploiting Co-morbidity Patterns To Drive Early Intervention", D. Onischenko, Y. Huang, P. Smith, and I. Chattopadhyay **Under Review in Nature Medicine**
- "Distilling Event Logs For Actionable Prediction of Urban Crime", T. Li, Y. Huang, J. Evans, and I. Chattopadhyay **Submitted to Science**

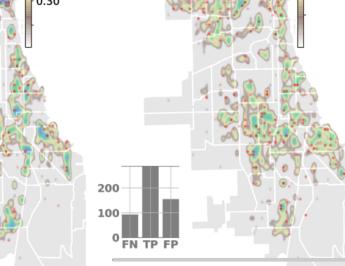
Patent:

Provisional patent application submitted for EHR based autism diagnosis

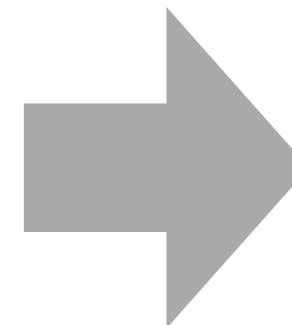
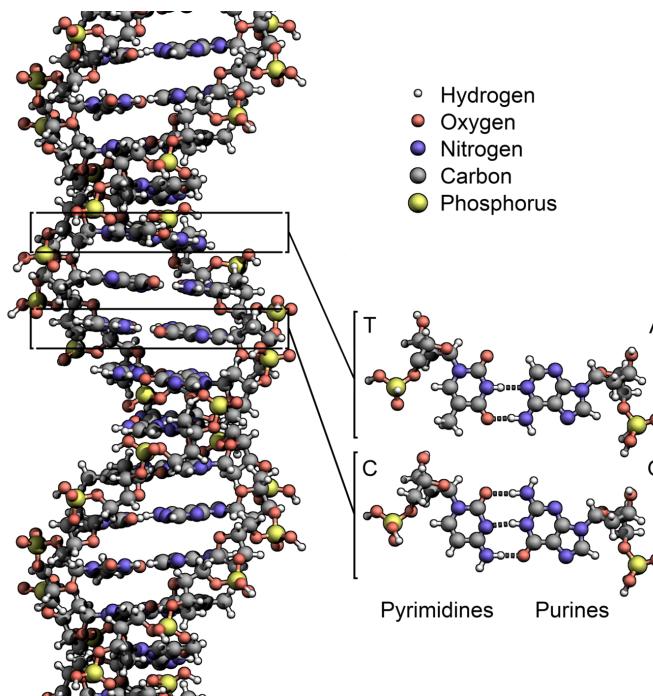
Software: quasinet ehrzero cynet

1. Release QNN framework
2. Expand disease predictors
3. Mapping predictions to interventions in social modeling

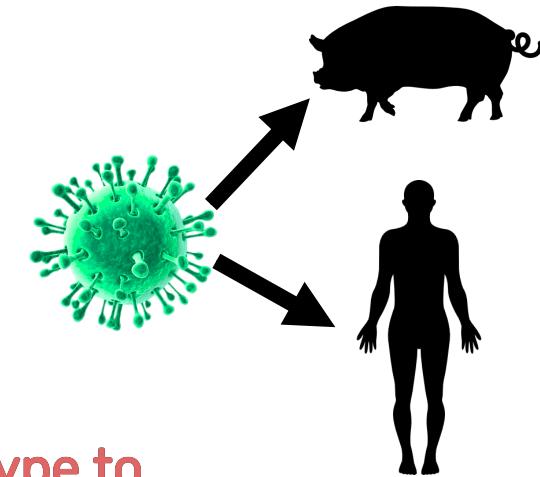
```
1 pip install quasinet
2 pip install ehrzero
3 pip install cynet
```



Problem 1: Sequence To Function Mapping in Biology

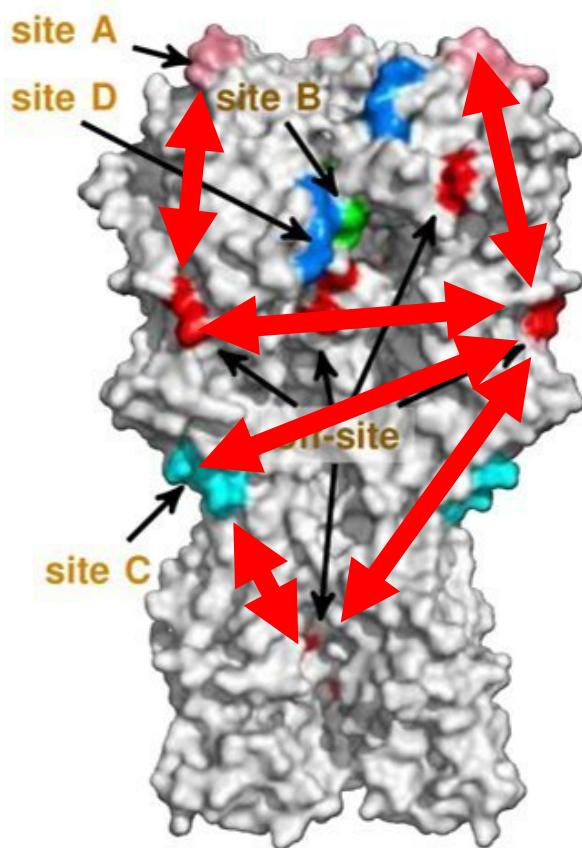


genotypic
information
+
environment
= phenotypic
outcomes

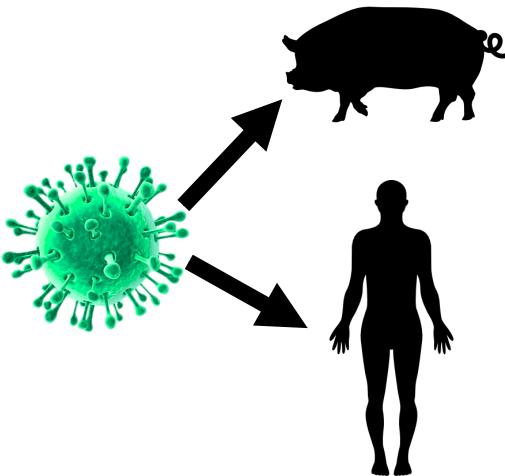


How Do We Learn The Genotype to
phenotype mappings?

Problem 1: Sequence To Function Mapping in Biology



Hemagglutinin in Influenza Virus



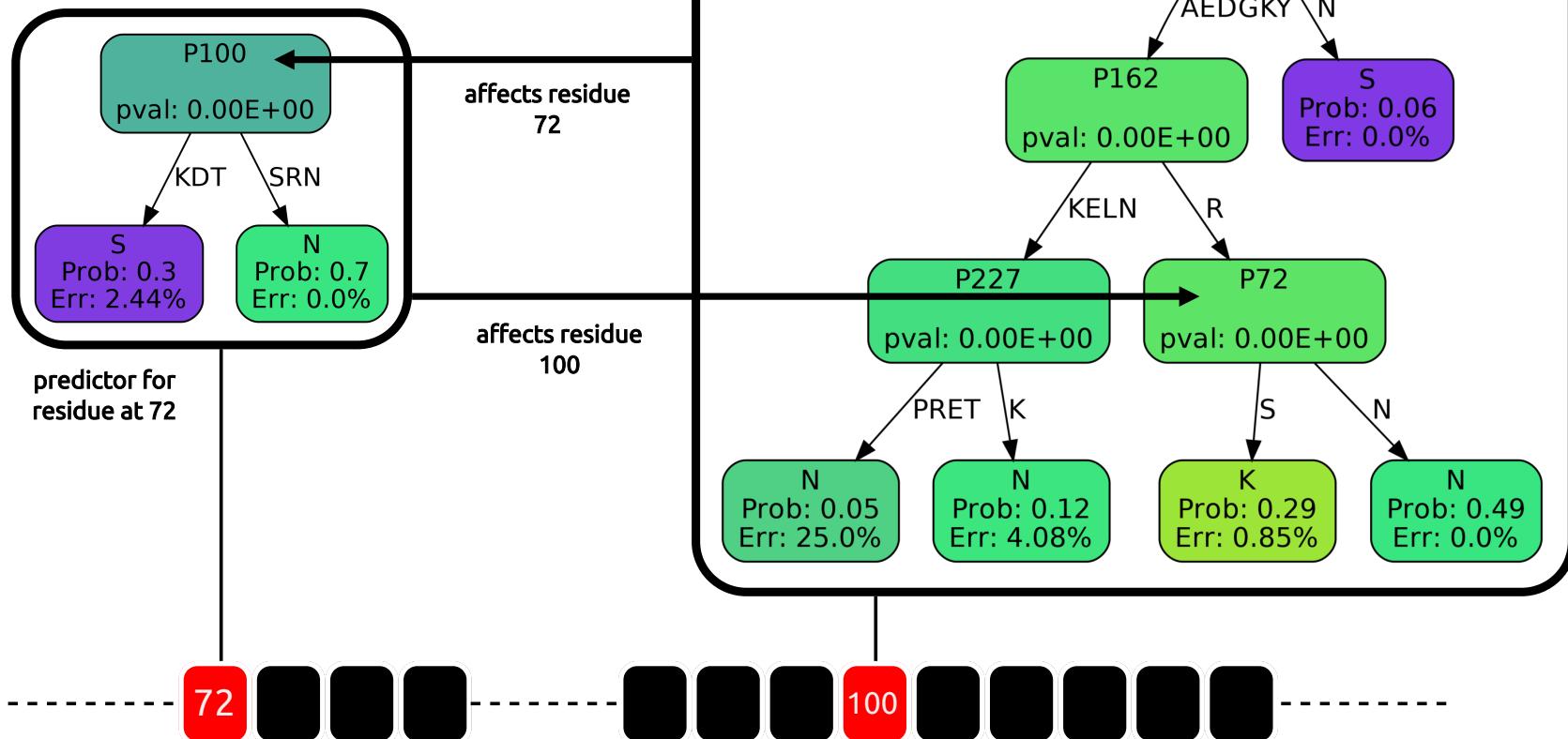
1. Must fold properly
2. Must mutate exposed sites to evade host immunity
3. Adapt to changing environments/hosts

Complicated long-range
dependencies

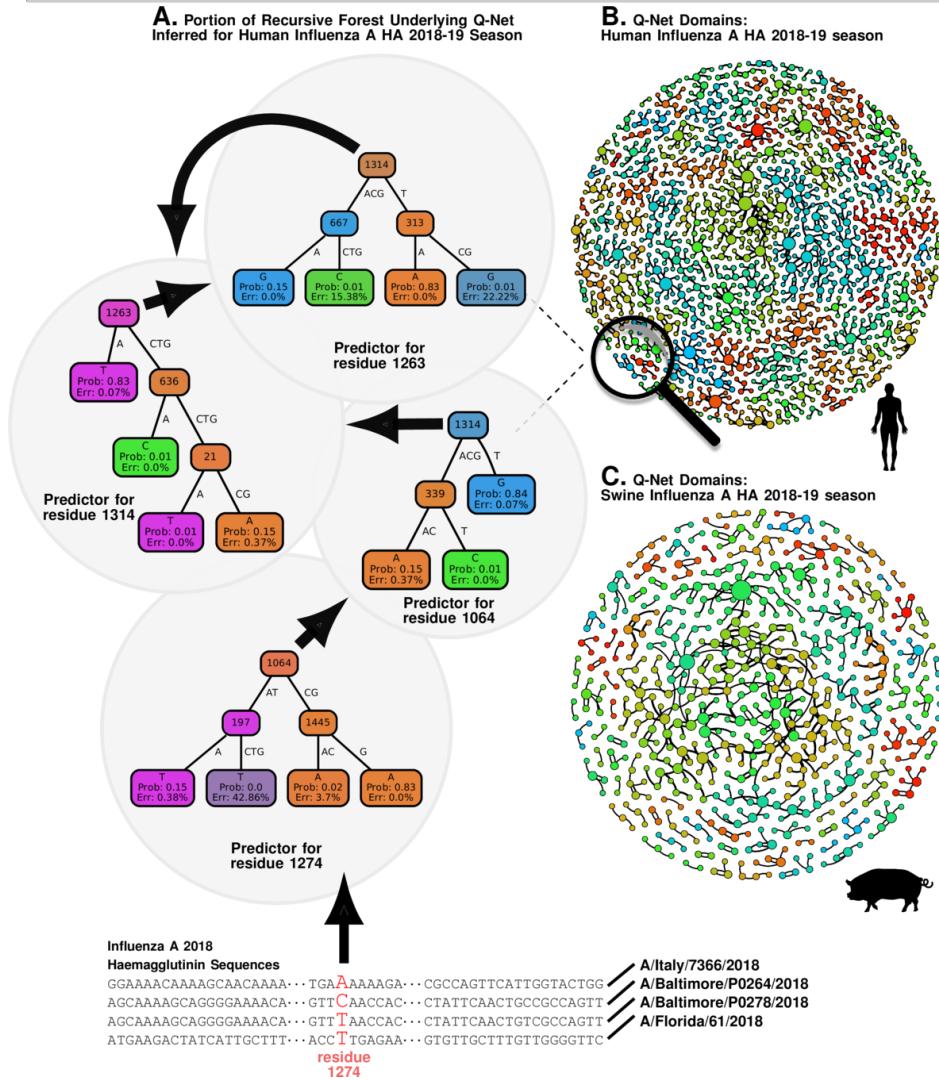
QNet: Inferring Structure From Sequence Databases



- Inferring all constraints on residue substitution
- structured/dependent "orchard" of decision trees (not random forest)



QNet: Inferring Structure From Sequence Databases



1. Not a Random Forest

- Recursive forest of conditional inference trees

2. Potentially infinite depth inference trees

3. Multiple Applications

- New distance metric in biological sequences
- Ability to chart evolutionary trajectories



Intrinsic Structure From Sequence Databases

- QNet Index Distribution

$$\zeta(x_A, i) = \text{prob. dist. at index } i|x_A$$

↑
computed efficiently by QNet

QNet computes the probability distribution of residues at some index i given the sequence x and the population A

- New Distance Via QNets Between Sequences and Populations

$$\forall x_A \in A, y_B \in B$$

$$\theta(x_A, y_B) = \mathbf{E}_i (\zeta(x_A, i) || \zeta(y_B, i))$$

$$\theta(x_A, B) = \min_{y_B \in B} \theta(x_A, y_B)$$

$$\theta(A, B) = \max_{x_A \in A} \theta(x_A, B)$$

Expected
Shannon
Jensen
Divergence

Wasserstein
metric is
better



Intrinsic Structure From Sequence Databases

- QNet Targeted Mutator

$$\mu_{A,y}(x)|_i = \begin{cases} \sigma(\zeta(x_A, i)), & \text{if } i \in \Delta(x, y) \\ x_i, & \text{otherwise} \end{cases}$$

$$\mu_{A,y}^k(x) = \mu_{A,y}(\mu_{A,y}^{k-1}(x))$$

$$\text{where } \Delta(x, y) \triangleq \{i : x_i \neq y_i\}$$

Note: that the targeted mutator maps a sequence to a distribution over sequences

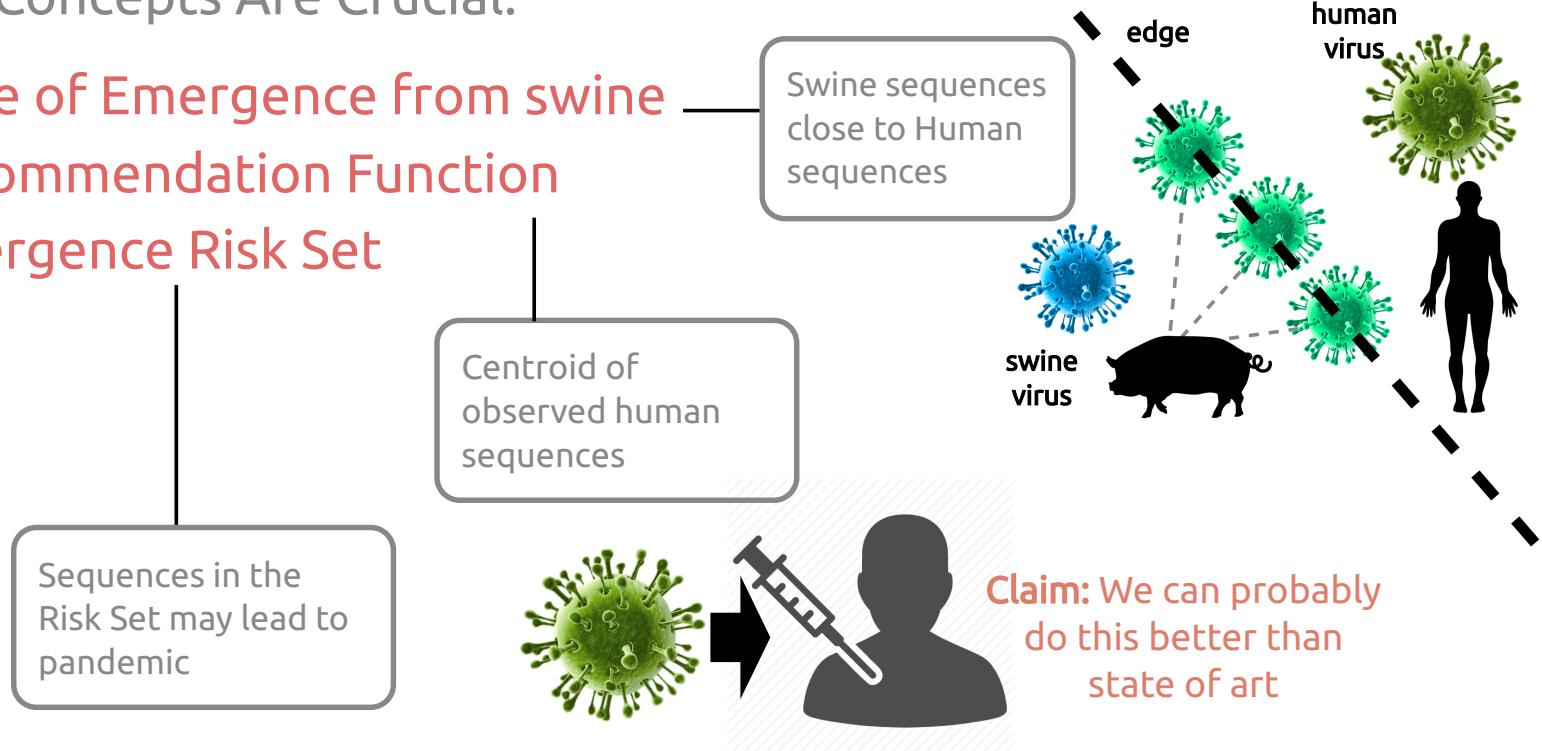
Claim: QNet mutators model actual evolutionary trajectories, including those we have not observed

Intrinsic Structure From Sequence Databases



Three Concepts Are Crucial:

- Edge of Emergence from swine
- Recommendation Function
- Emergence Risk Set



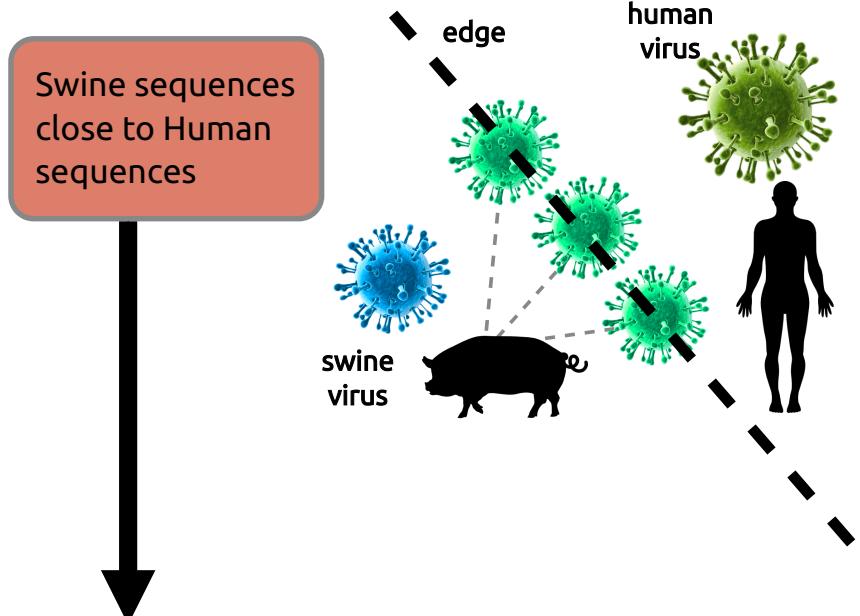
Claim: We can almost certainly do this better than the (non-existent) state of art

Intrinsic Structure From Sequence Databases



Three Concepts Are Crucial:

- Edge of Emergence from swine
- Recommendation Function
- Emergence Risk Set



$$\wp(X_{S_t}, X_{H_t}) = \left\{ x \in X_{S_t} : \exists y \in X_{H_t}, \forall p \in X_{S_t}, \theta(x, y) \leq \theta(p, y) \right\}$$

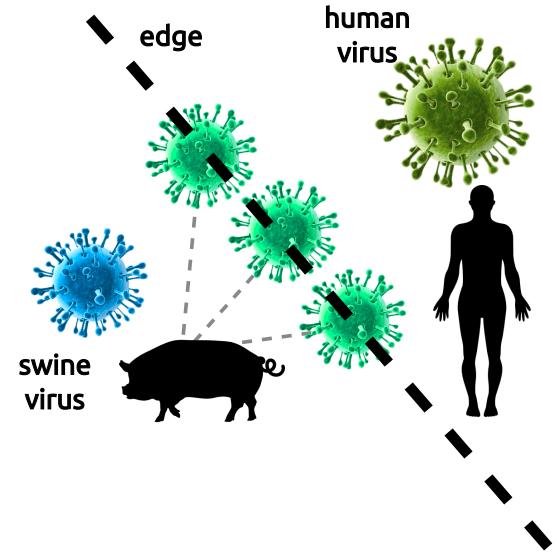


Intrinsic Structure From Sequence Databases

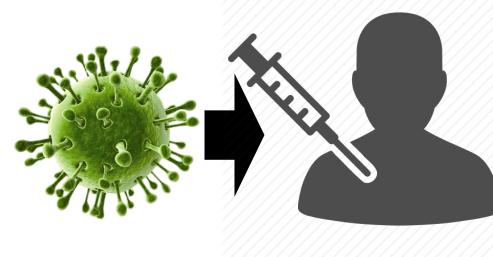
Three Concepts Are Crucial:

- Edge of Emergence from swine
- **Recommendation Function**
- Emergence Risk Set

Centroid of
observed human
sequences



$$X_{H_t}^* = \arg \min_{x \in X_{H_t}} \sum_{y \in X_{H_t}} \theta(x, y)$$





Intrinsic Structure From Sequence Databases

Three Concepts Are Crucial:

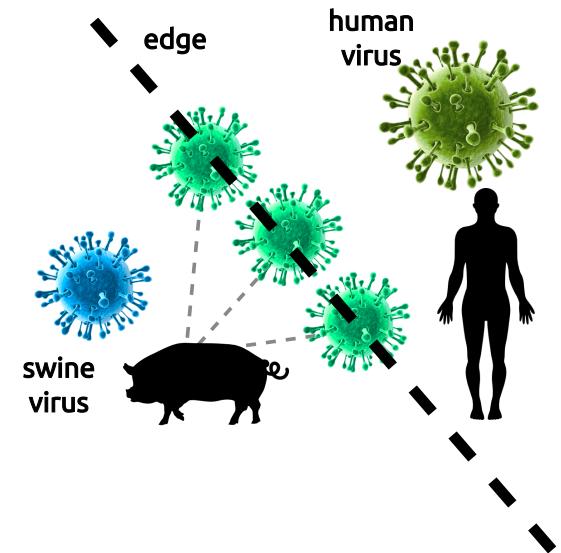
- Edge of Emergence from swine
- Recommendation Function
- **Emergence Risk Set**



Sequences in the
Risk Set may lead to
pandemic



$$\begin{aligned}\mathcal{E}_{H_t}^{S_t}(m, N, T) = \left\{ x \in \wp^m(X_{S_t}, X_{H_t}) : \forall n > N, \right. \\ \left. \theta(\mu_{X_{S_t}, X_{H_t}}^n(x), X_{H_t}^*) \geq T \right\}\end{aligned}$$



Intrinsic Structure From Sequence Databases



- Edge of Emergence from S to H

$$\wp(X_{S_t}, X_{H_t}) = \left\{ x \in X_{S_t} : \exists y \in X_{H_t}, \forall p \in X_{S_t}, \theta(x, y) \leq \theta(p, y) \right\}$$

- Recommendation Function

$$X_{H_t}^* = \arg \min_{x \in X_{H_t}} \sum_{y \in X_{H_t}} \theta(x, y)$$

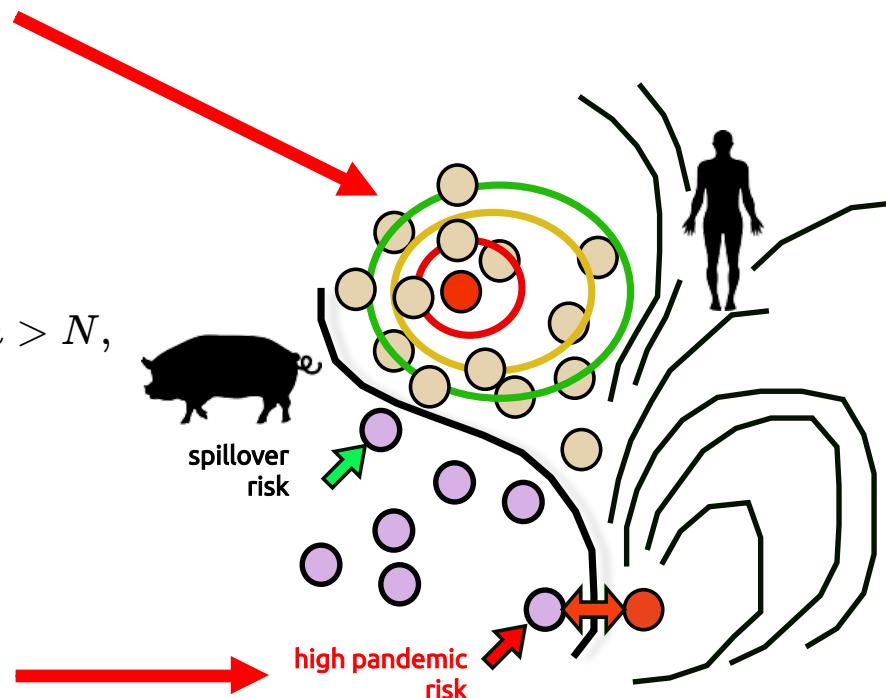
- Emergence Risk Set

$$\mathcal{E}_{H_t}^{S_t}(m, N, T) = \left\{ x \in \wp^m(X_{S_t}, X_{H_t}) : \forall n > N, \theta(\mu_{X_{S_t}, X_{H_t}}^n(x), X_{H_t}^*) \geq T \right\}$$

- We have pandemic risk if:

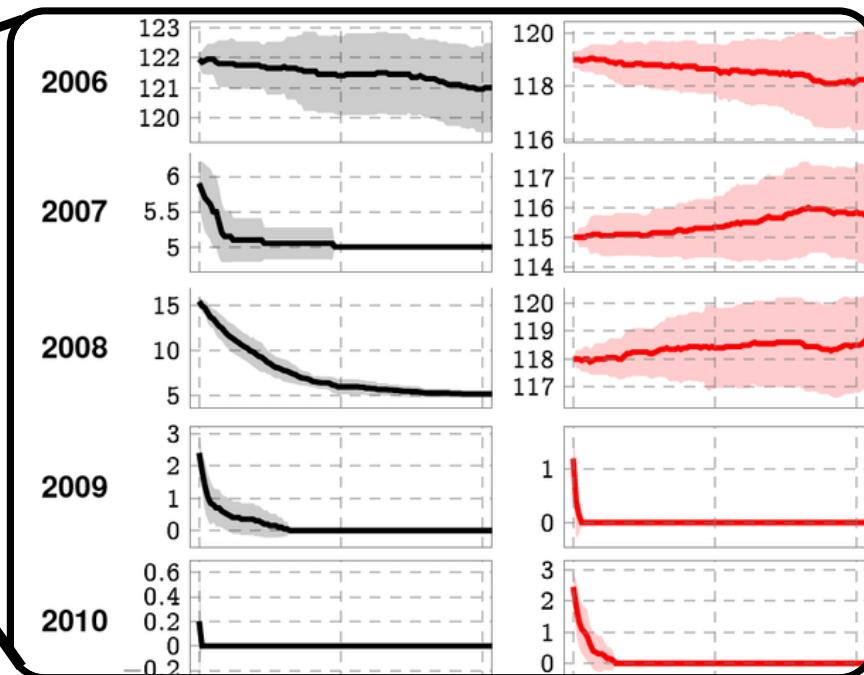
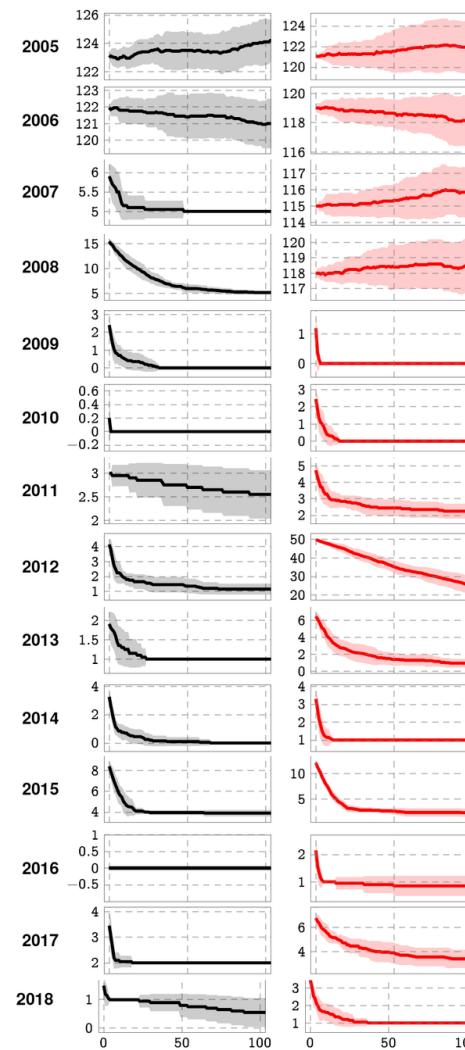
$$\mathcal{E}_{H_t}^{S_t}(m, N, T) \neq \emptyset$$

where $m \approx 10, N \geq 10^3, T \geq 100$





Could we have predicted the A/pmd09(H1N1)* strain in 2008?



low risk

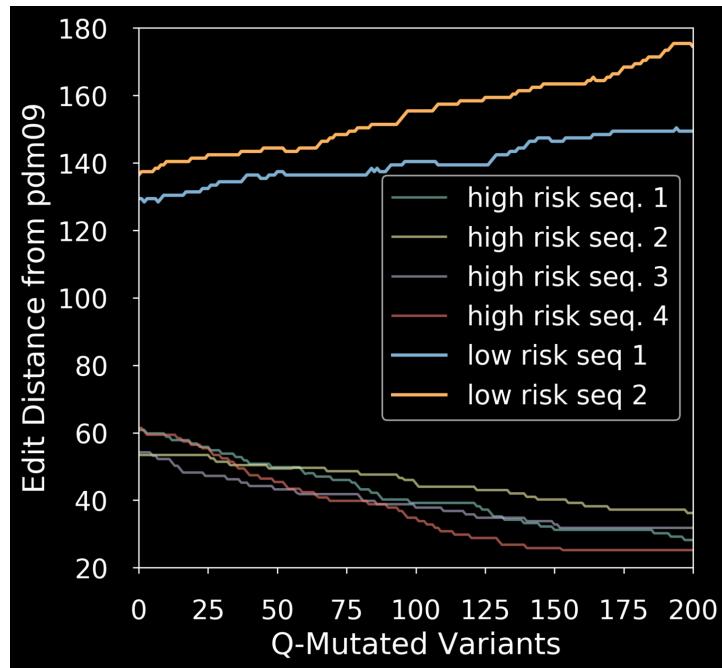


2008

2009



Could we have predicted the A/pmd09(H1N1)* strain in 2008?



Predicted High Risk Swine Strains	D0	Dpdm	Dpdm*
A/swine/Ohio/FAH2-1/2008(H1N1)	115	49.8	20.2
A/swine/Korea/VDS4/2008(H1N1)	121	54.2	19.2
A/swine/Minnesota/SG1317/2008(H1N1)	118	53.4	19.2
A/swine/Iowa/46519-3/2008(H1N1)	121	52.8	20.2
A/swine/Iowa/02096/2008(H1N1)	122	54.8	18.2
A/swine/Iowa/46519-2/2008(H1N1)	119	61.4	20.2
A/swine/Nebraska/02013/2008(H1N1)	127	60.8	21.2

D0: distance from human recommendation in 2009

Dpdm: Distance from pdm09

Dpdm*: Distance from pdm09 post evolutionary Q-simulation

Yes

We can predict:

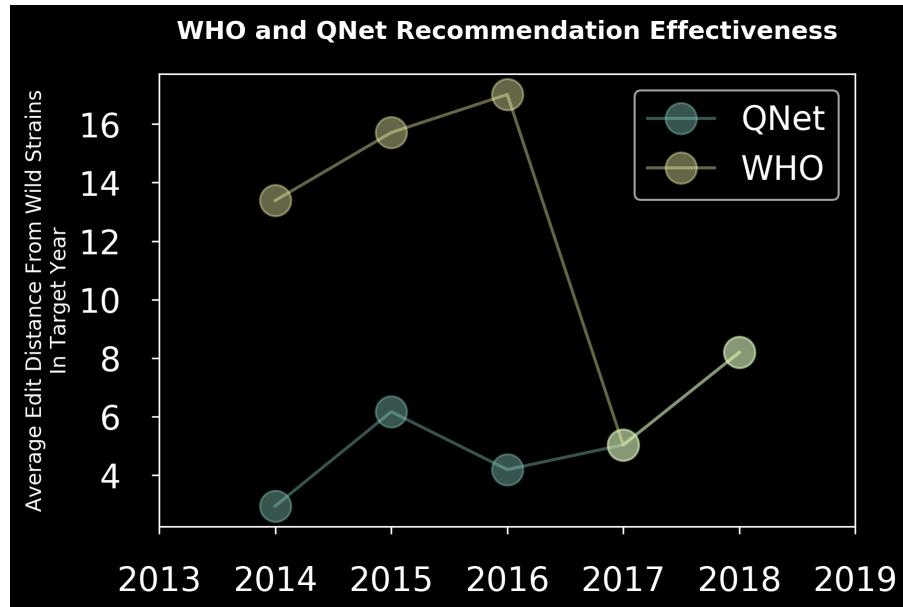
- Imminent emergence
- Close approximation of the emerging strain
- Comparison of column D0 and Dpdm* shows that our approach is not just recommending the swine sequence closest to human sequences since we are able to correctly identify swine seq. >100 mutations removed from 2008 human seq.

* novel emergent strain in 2009 H1N1 swine flu pandemic



How does WHO come up with seasonal recommendations?

Can we do better?



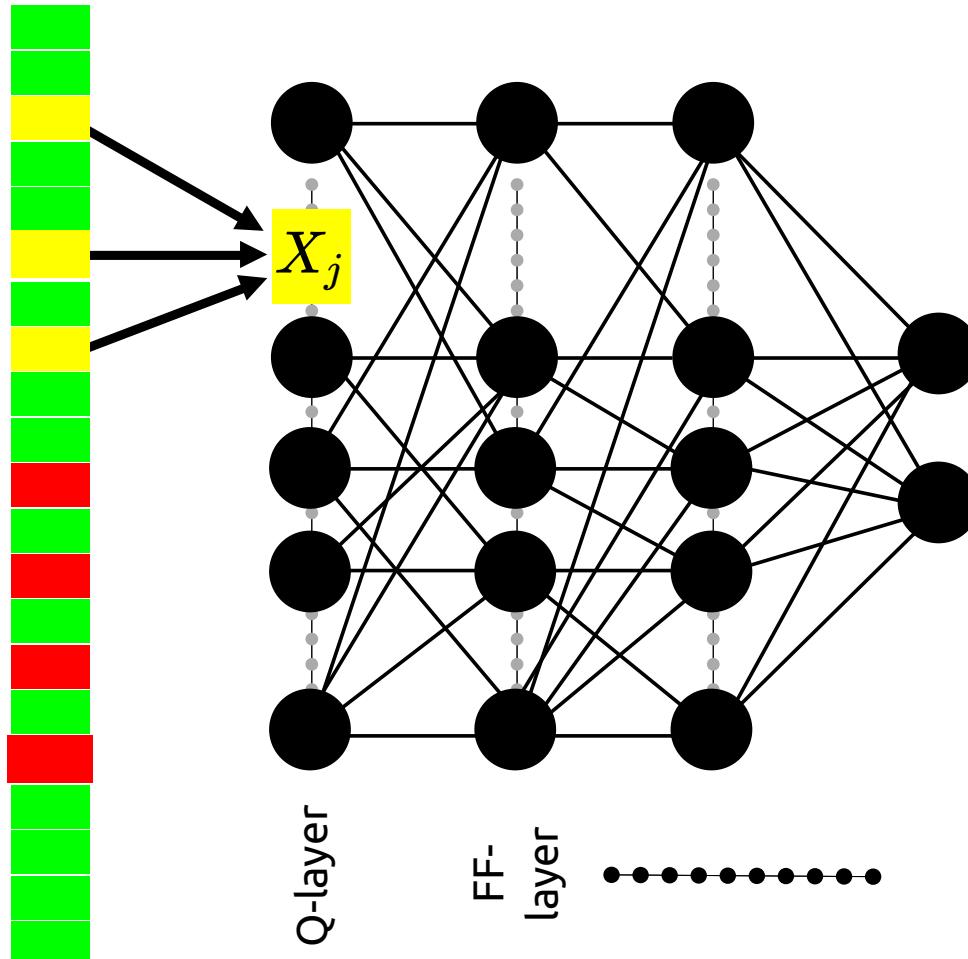
	Q-Net Recommendation	WHO Recommendation
2013-2014	A/India/Del136197/2014(H1N1)	A/California/7/2009 (H1N1)pdm09
2014-2015	A/New York/WC-LVD-14-060/2014(H1N1)	A/California/7/2009 (H1N1)pdm09
2015-2016	A/New York/WC-LVD-14-060/2014(H1N1)	A/California/7/2009 (H1N1)pdm09
2016-2017	A/California/NHRC_BRD41284N/2016(H1N1)	A/Michigan/45/2015 (H1N1)pdm09
2017-2018	A/California/NHRC_BRD41284N/2016(H1N1)	A/Michigan/45/2015 (H1N1)pdm09

Probably.

As shown by the average number of mutations between the strains that showed up in target year and the recommendation



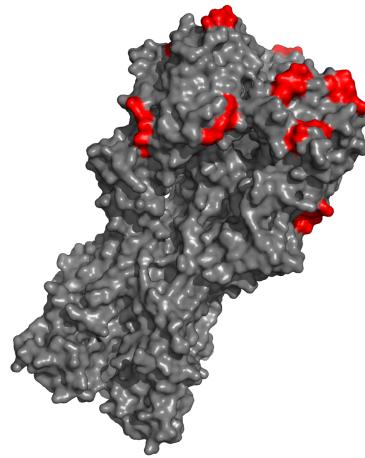
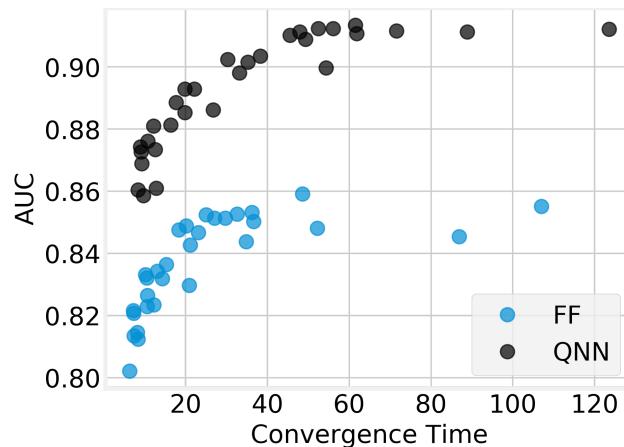
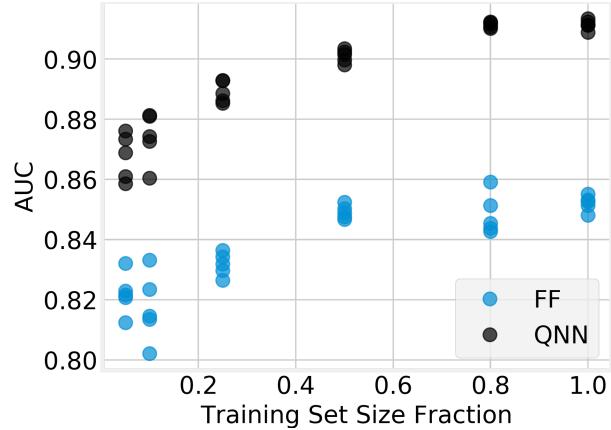
QNN: Neural Nets Augmented With QNet Structure



- NN structure reflects distinctive biology and evolutionary patterns
- Same question asked on different species (HIV vs influenza) has different NN structure
- Leveraging unlabeled data from sequence databases



QNN: Neural Nets Augmented With QNet Structure



- Human Influenza A Virus Hemagglutinin (HA) for years 2010-2018
- 40,000 sequences from NCBI Database
- **Predict Year of Collection**

Genotype to Phenotype Problem

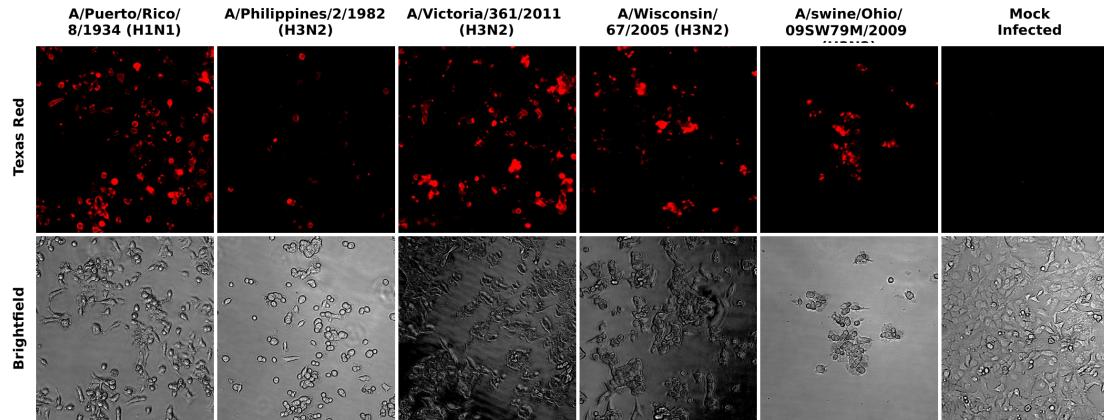
QNet Inference

Tensor Flow

Experimental Validation of Predictions

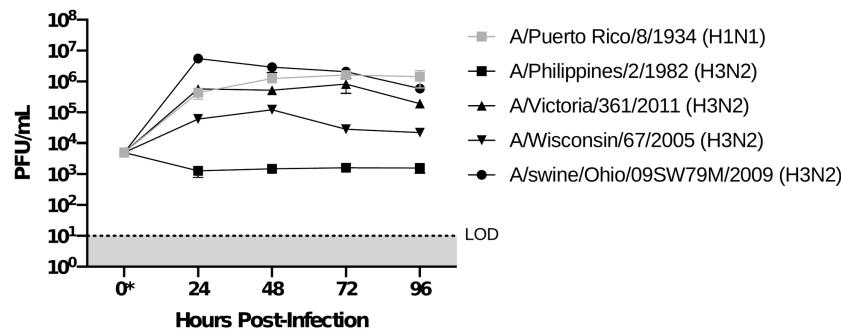


Micrographs of Infected MDCK Cells



MDCK Cells | MOI~0.01 | 72 HPI
1 ug/mL TPCK-treated Trypsin

A549 Cells



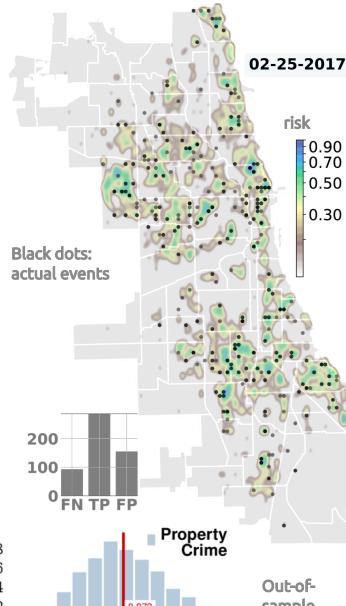
A549 Cells
MOI~0.01 (5,000 PFU)
LOD: 10 PFU
0* - Calculated Input; Not Titrated

Hypothesis: Decomposing the learning problem, to first estimate unknown problem structure, followed by standard inference leads to faster, better, efficient learning

Key Insight: Notion of structure changes between problem domains, but basic idea remains valid

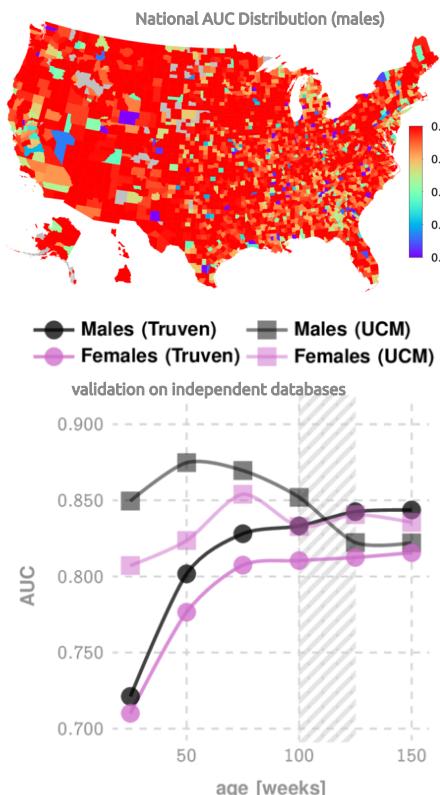
XPFSAs

1. Predicting Crime in Chicago



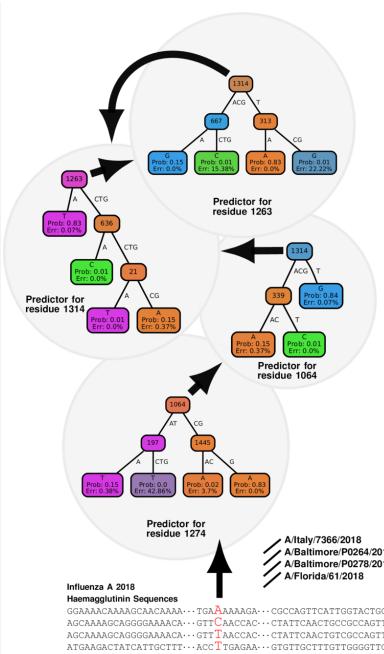
PFSA

2. Predicting Autism Diagnosis from Electronic Health Records



QNN

3. Genotype to Phenotype Mapping



Publications

1. "Beyond edit distance: A Function aware adaptive distance metric for evolving populations", T. Li, P. Leon, A. Weinberger, and I. Chattopadhyay **In Preparation**

2. "QNN: Using Quasi-species Sequence Variations to Inform Neural Net Modeling of The Genotype to Phenotype Problem", T. Li, I. Chattopadhyay **In preparation**

3. "Zero-burden Digital Biomarkers For Autism: Exploiting Co-morbidity Patterns To Drive Early Intervention", D. Onischenko, Y. Huang, P. Smith, and I. Chattopadhyay **Under Review in Nature Medicine**

4. "Distilling Event Logs For Actionable Prediction of Urban Crime", T. Li, Y. Huang, J. Evans, and I. Chattopadhyay **Submitted to Science**

Patent:

Provisional patent application submitted for EHR based autism diagnosis

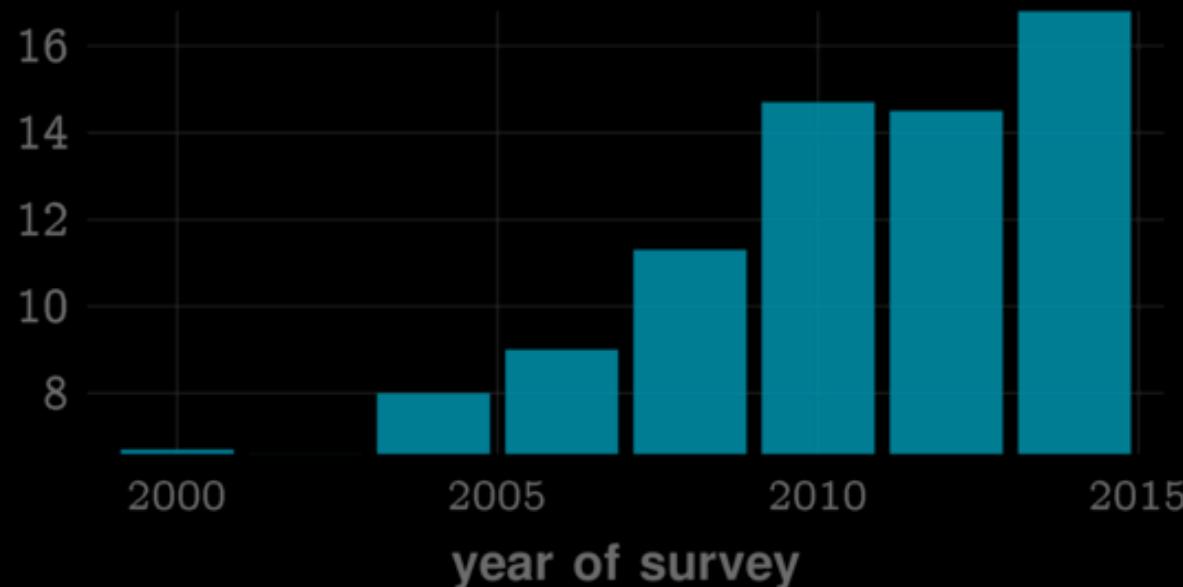
Software: quasinet ehrzero cynet

1. Release QNN framework
2. Expand disease predictors
3. Mapping predictions to interventions in social modeling

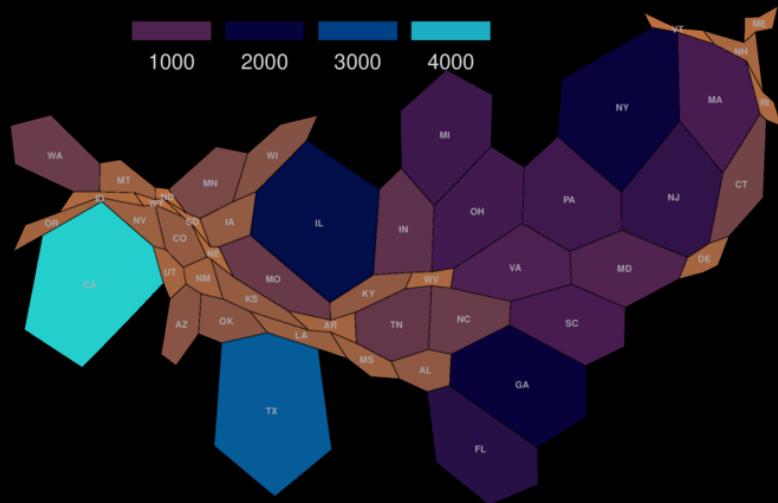
```
1 pip install quasinet
2 pip install ehrzero
3 pip install cynet
```

Autism Spectrum Disorder

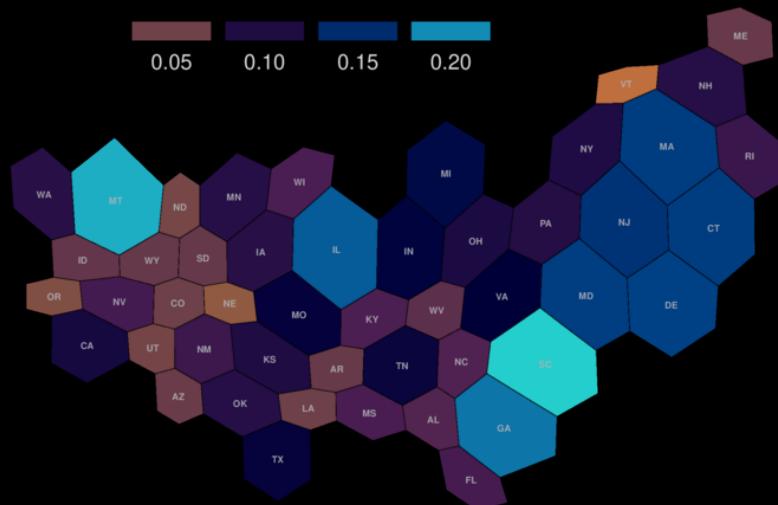
1 in 59



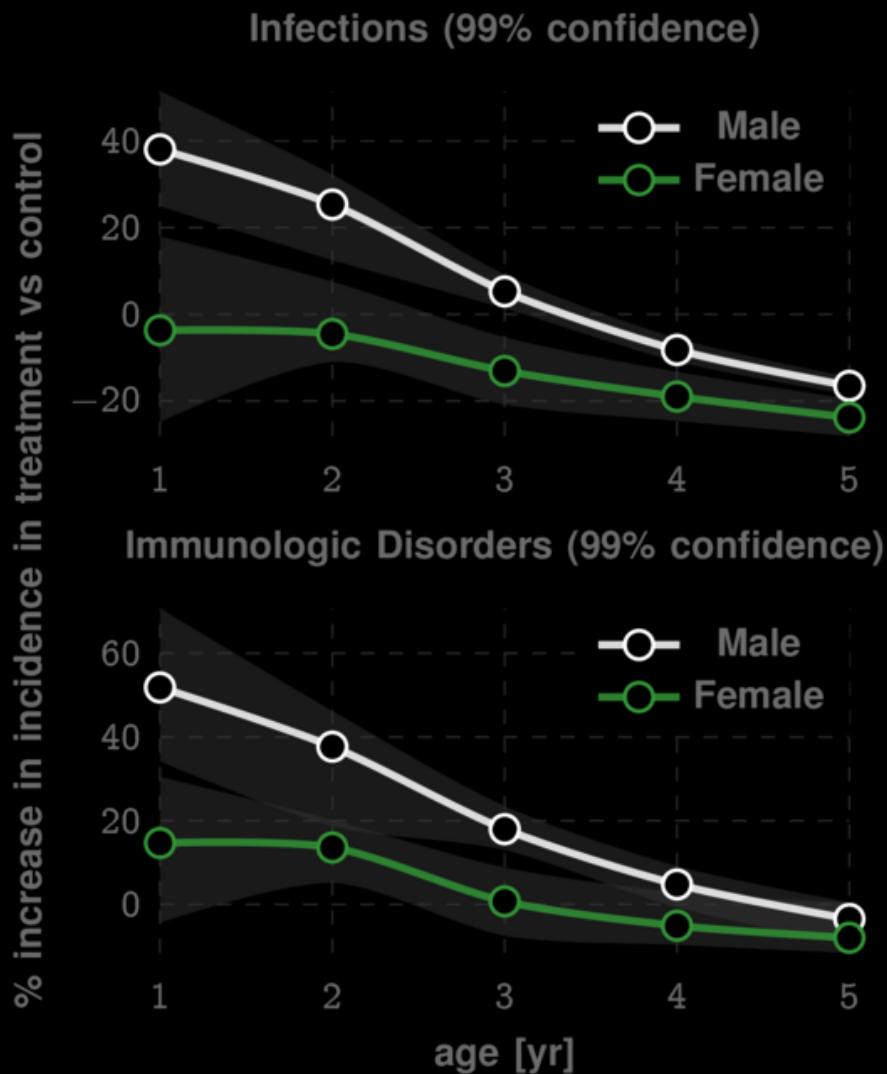
A. Autism Insurance Claims (2003-2013)



B. Autism Prevalence in US (Pop. Normalized)

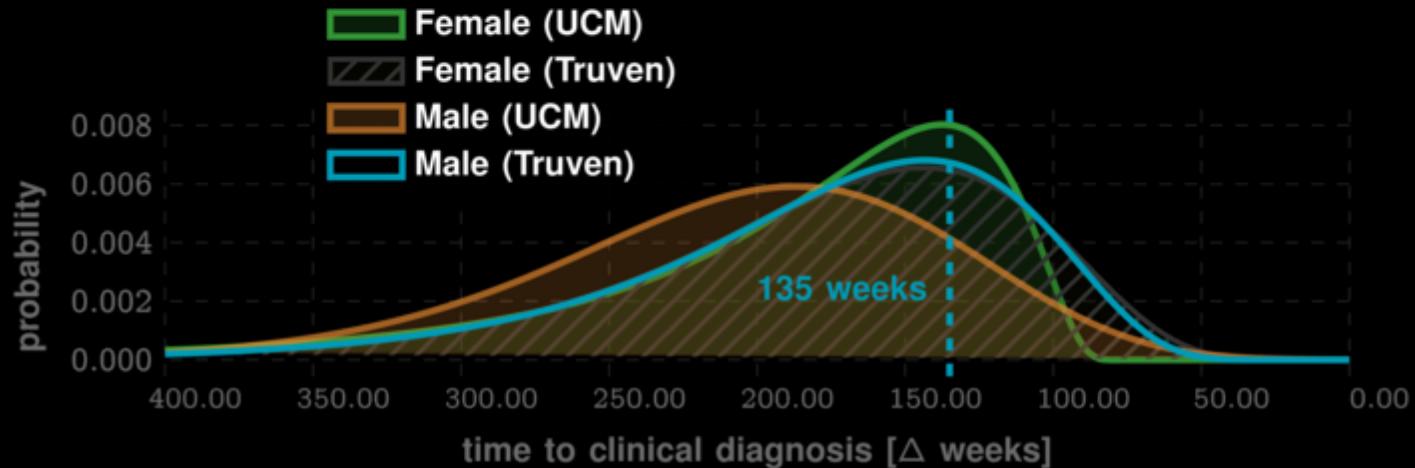


Median
Diagnosis
Age
~ 4 yrs



Autistic children
experience higher
co-morbidities

Can we exploit these
patterns to predict
diagnosis?



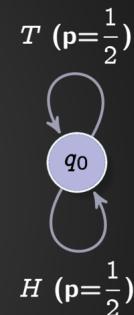
Learn from **stochastic**, highly
sparse, medical histories

Diagnosis Age
~4 years
↓
<2 years

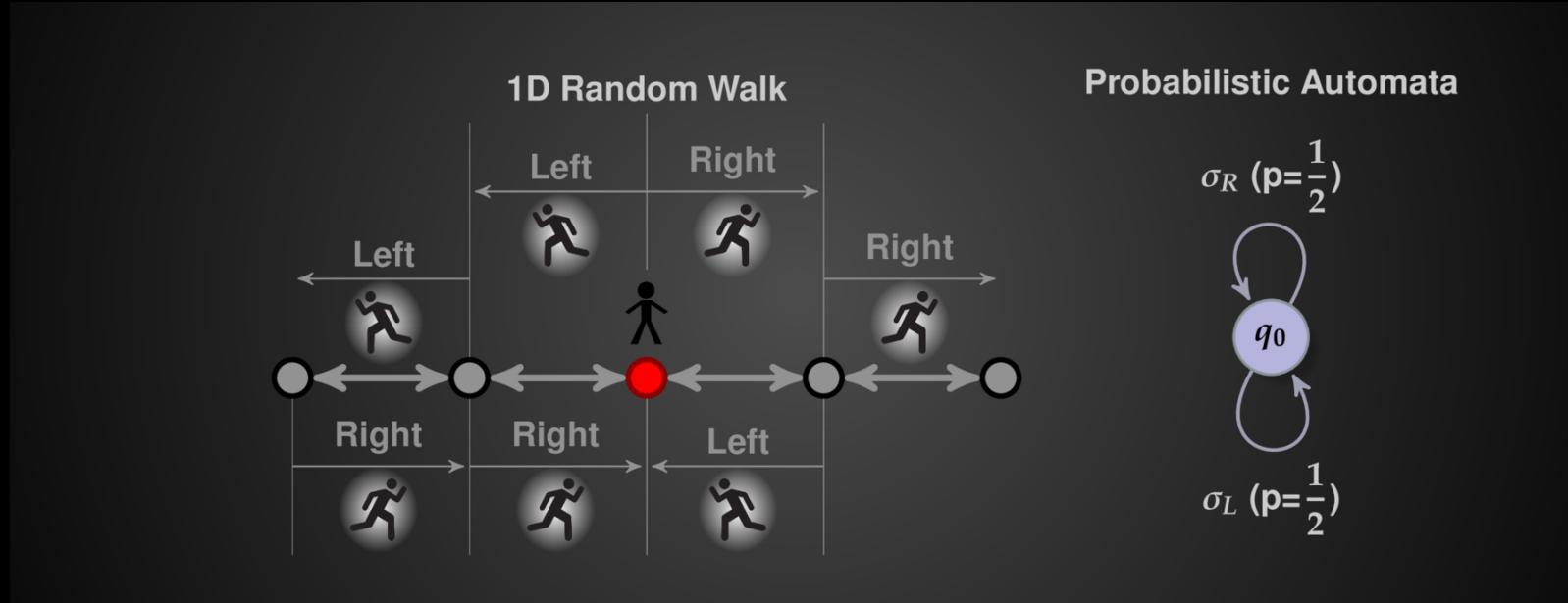
Structure of Finite Valued Ergodic Stationary Processes



Exact Model
for unbiased coin



Structure of Finite Valued Ergodic Stationary Processes





Structure of Finite Valued Ergodic Stationary Processes



Structure of Finite Valued Ergodic Stationary Processes



- 1. states are equivalence classes of histories**

- 2. More states, and more complex wiring indicates complicated dependency structure**

Structure of Finite Valued Ergodic Stationary Processes

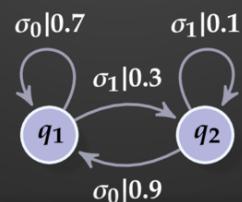


0100100010010000100000100001010000100001010010000000100000100000010010010000000000010
 00101000101000100010001000000010001010110000101000101000110010000001000000010101000000010001
 0010100100011000110010100010000000010001000010010000110010001100100001010000001010000000101
 01000100000000000111100000001000010100010001000000001000001000001010001010101001001000101
 00100000001000011000100000000100101000000001000000100000001001010000100001000010000100001000 ...

GenESS

$$\phi_x|_{\sigma} = Pr(x\sigma|x)$$

$x \in \Sigma^*$: immediate past



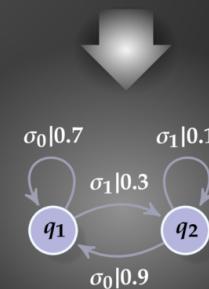
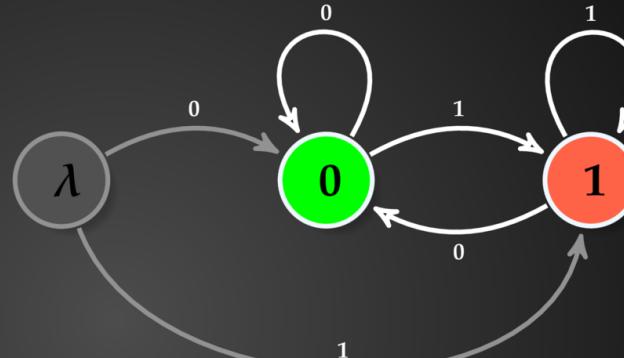
generator Extraction from Self-Similar Semantics

I. Chattopadhyay and H. Lipson, "Abductive learning of quantized stochastic processes with probabilistic finite automata", Philosophical Transactions of The Royal Society A, Vol. 371(1984), Feb 2013, pp 20110543

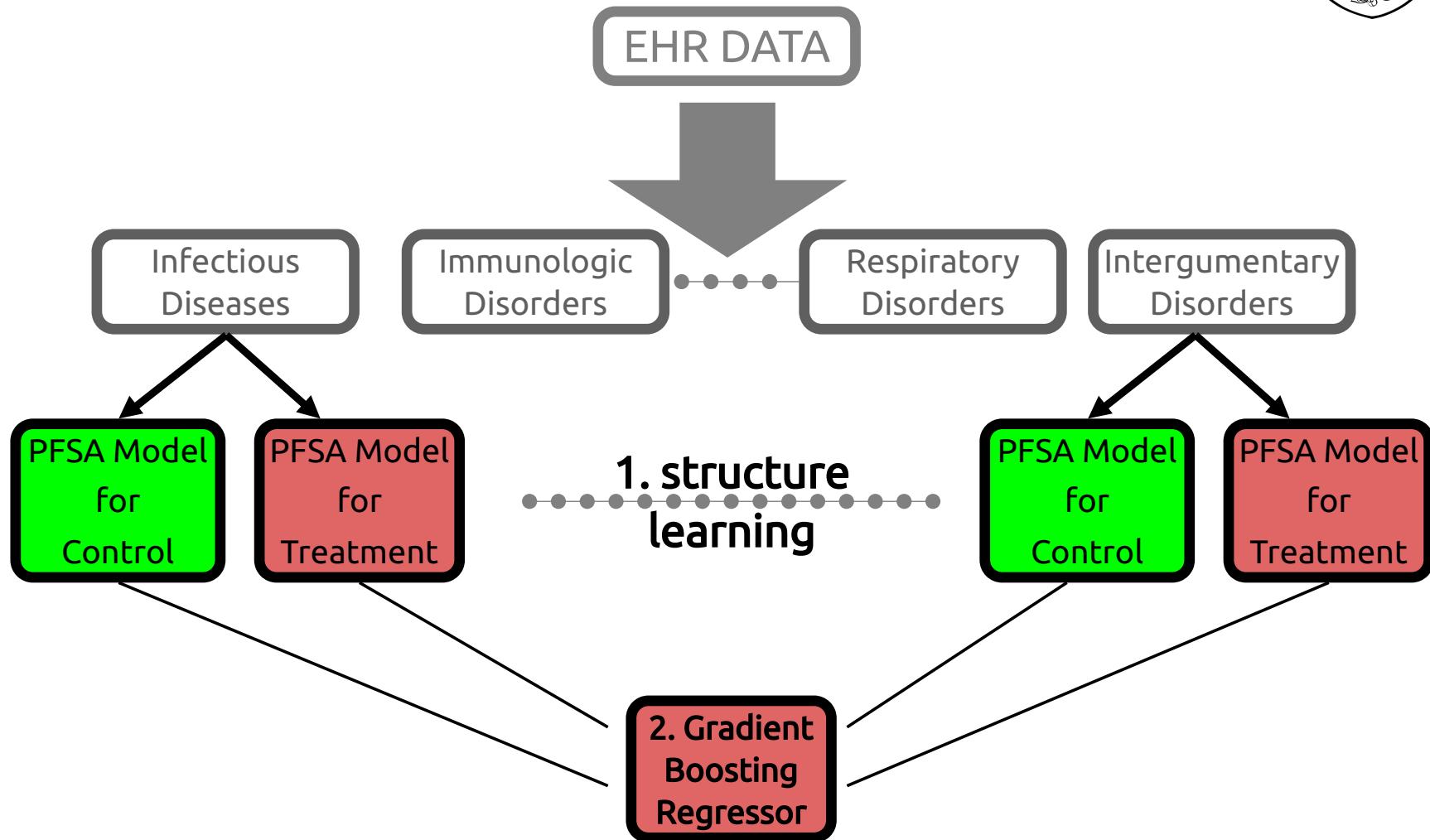
Structure of Finite Valued Ergodic Stationary Processes



λ	0.750285	0.249715
0	0.700112	0.299888
1	0.901009	0.0989909
00	0.699844	0.300156
01	0.899111	0.100889
10	0.700711	0.299289
11	0.918285	0.0817152
000	0.699004	0.300996
001	0.898769	0.10123
010	0.701038	0.298962
011	0.917181	0.0828194
100	0.701763	0.298237
101	0.899911	0.100089
110	0.697797	0.302203
111	0.930693	0.0693069
0000	0.699284	0.300716
0001	0.902025	0.0979754
:	:	:



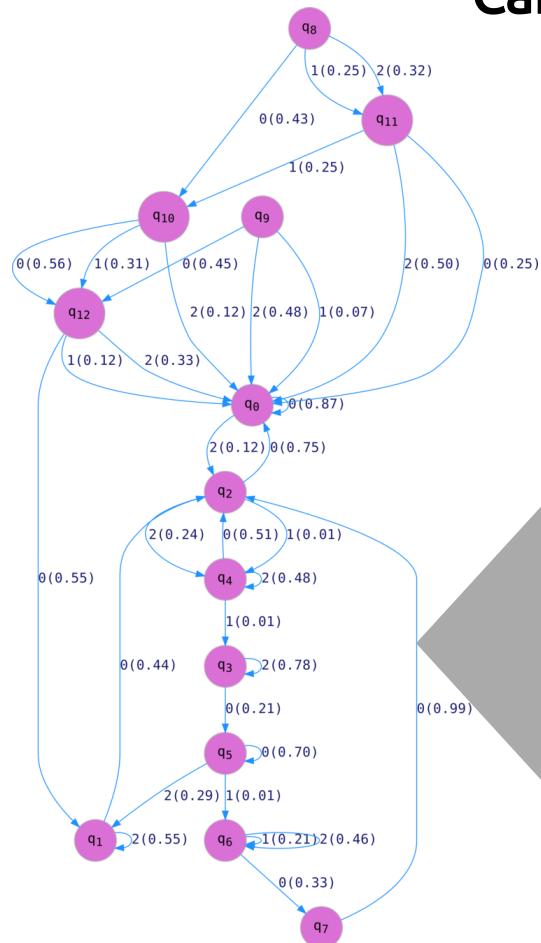
Predictive Diagnoses



Learning Structure in Predictive Diagnoses



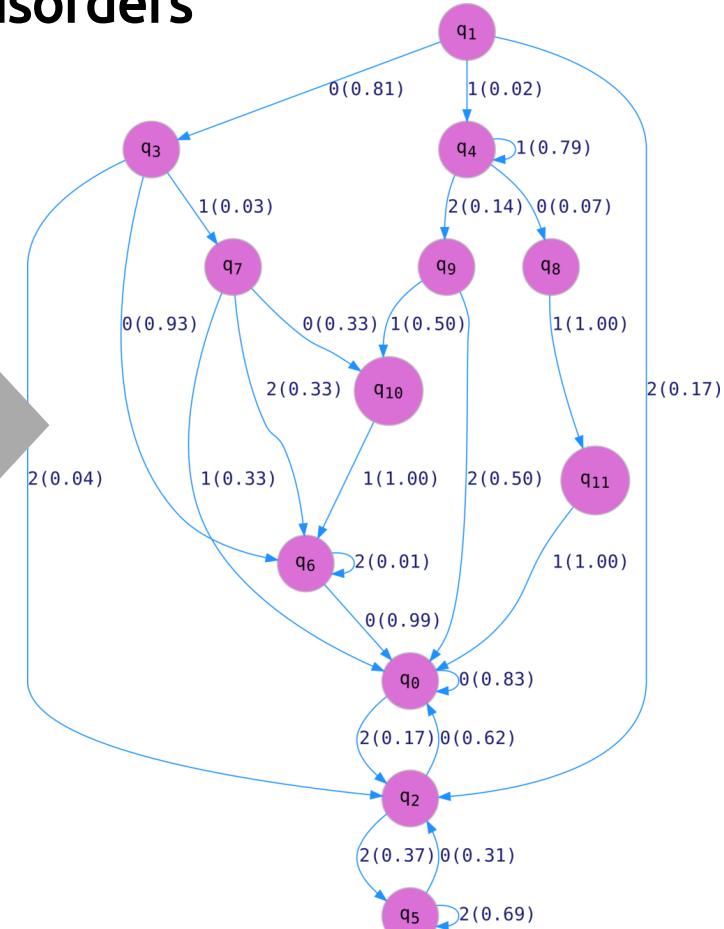
Cardio-vascular Disorders



Female.
Treatment



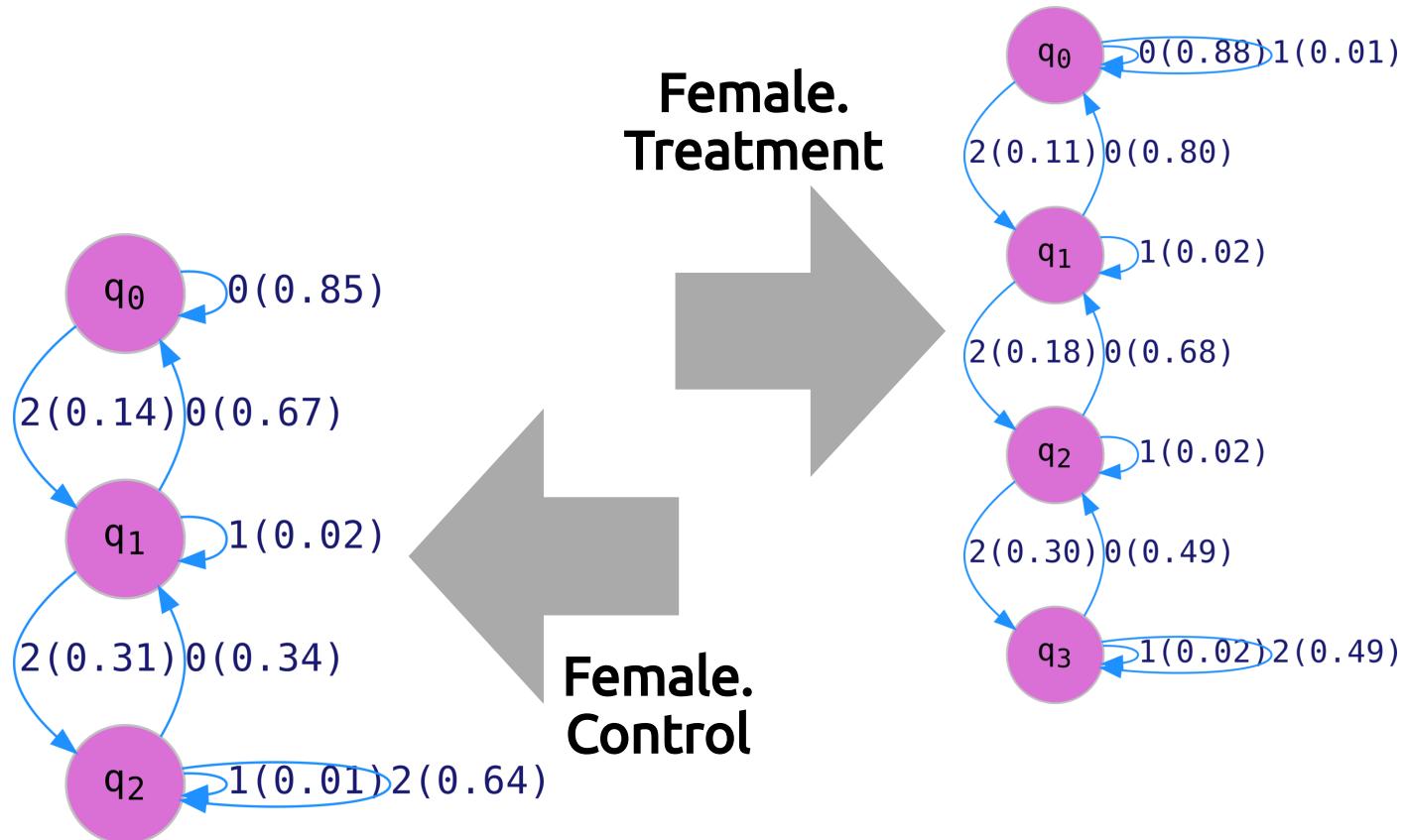
Female.
Control

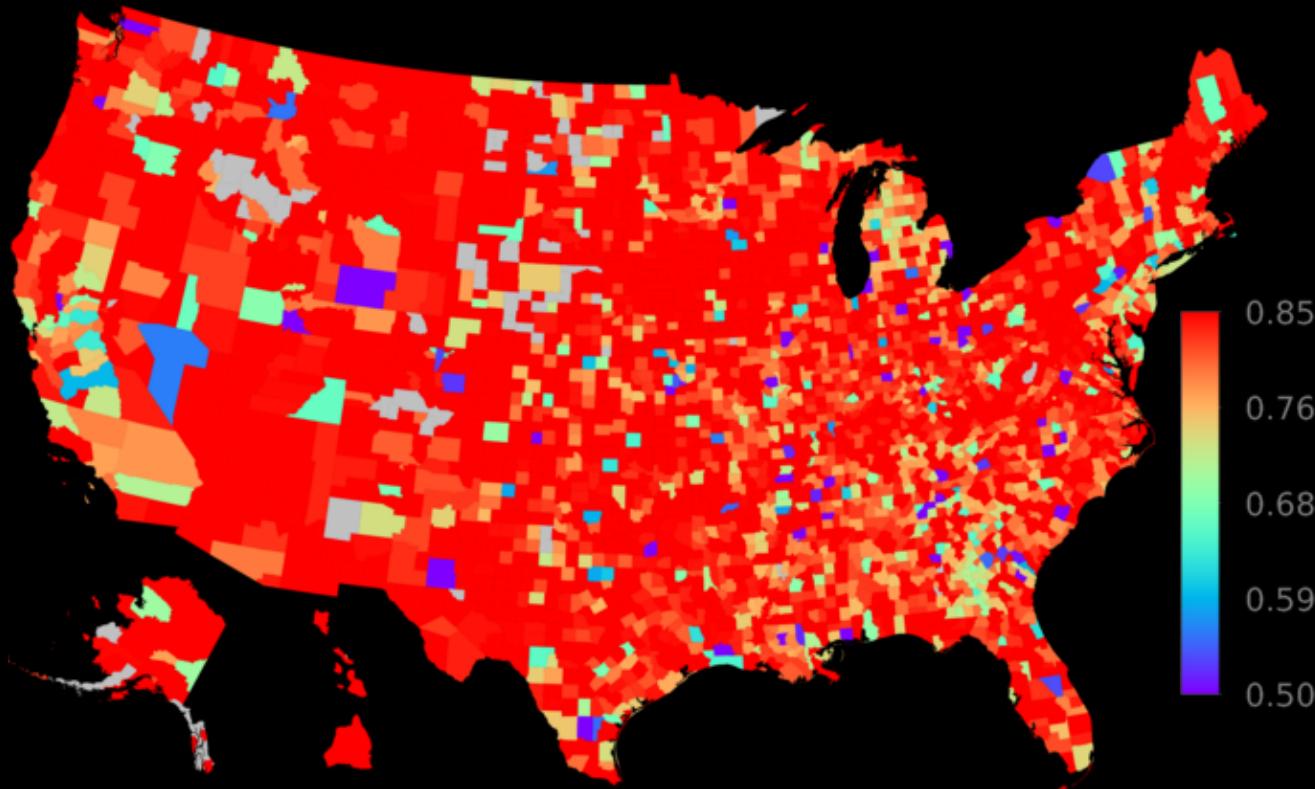


Learning Structure in Predictive Diagnoses



Immunological Disorders

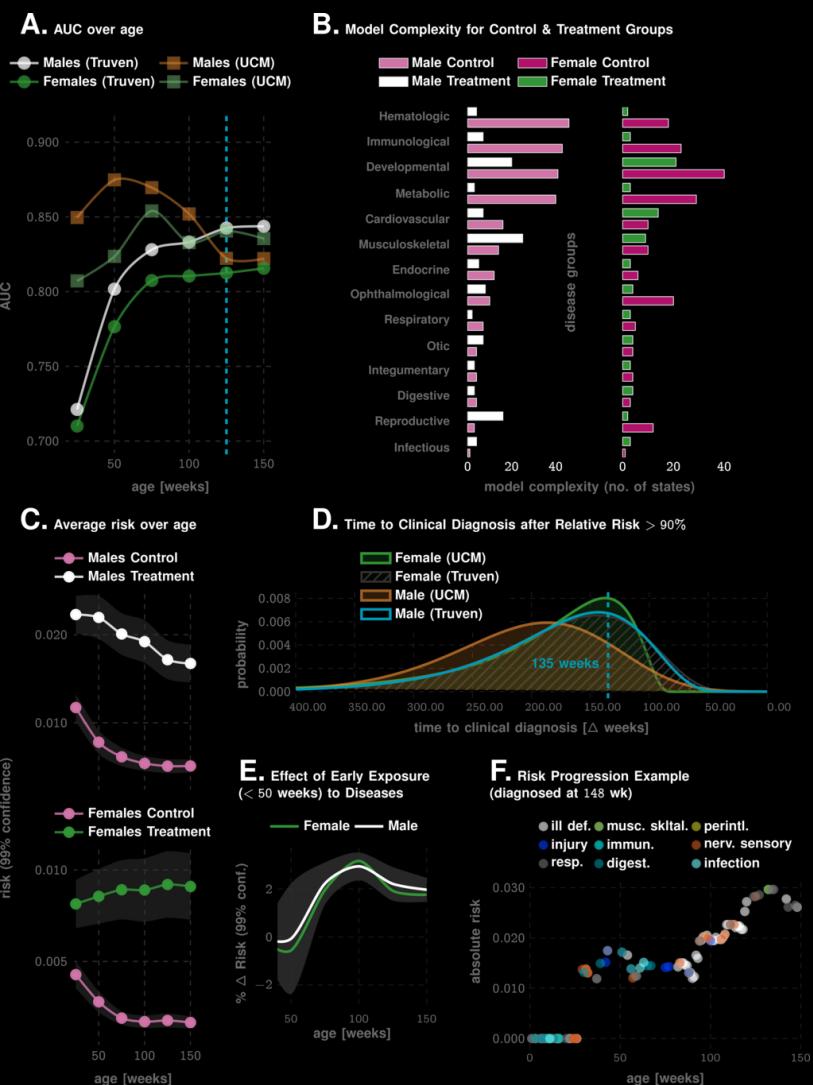




National
AUC
Distribution

- No blood work
- **Validated on millions of patients**

n=5M (insurance claims) , 70K (UChicago Medical Center)



Zero-burden Digital Biomarkers For Autism: Exploiting Co-morbidity Patterns To Drive Early Intervention

Dmytro Onishchenko¹, Yi Huang¹, Peter J. Smith² and Ishanu Chattopadhyay^{1,*}

¹Institute of Genomics and Systems Biology and Department of Medicine

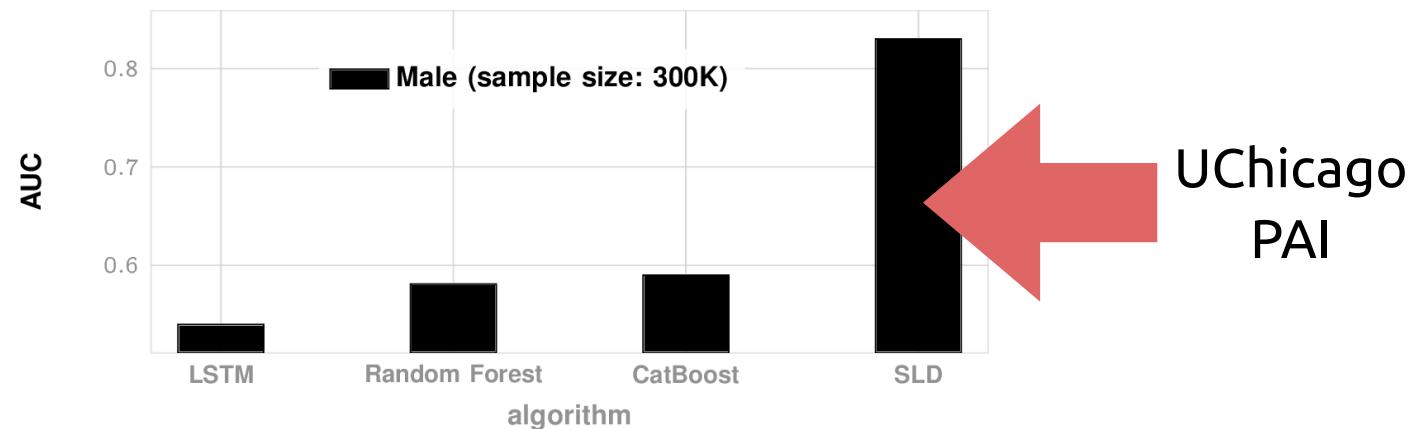
²Department of Pediatrics, University of Chicago, Chicago, IL, 60637, USA

*To whom correspondence should be addressed: e-mail: ishanu@uchicago.edu.

State of the Art Comparison

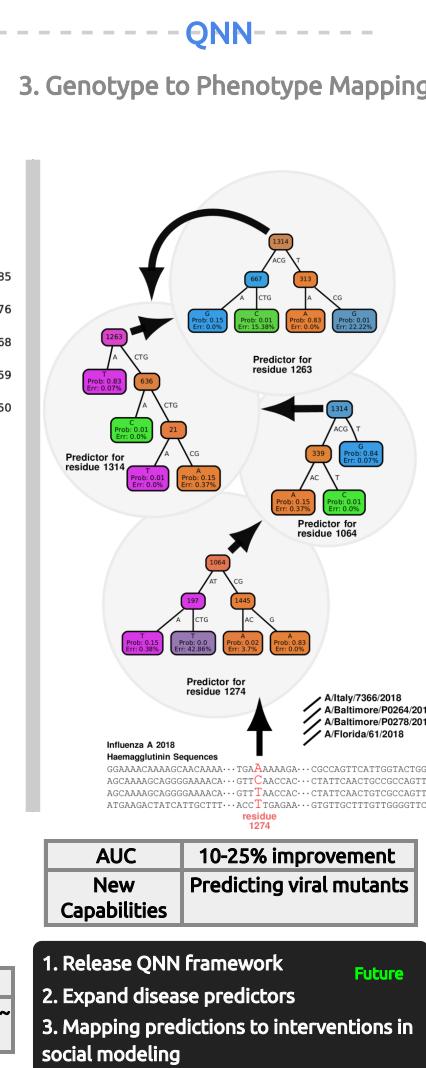
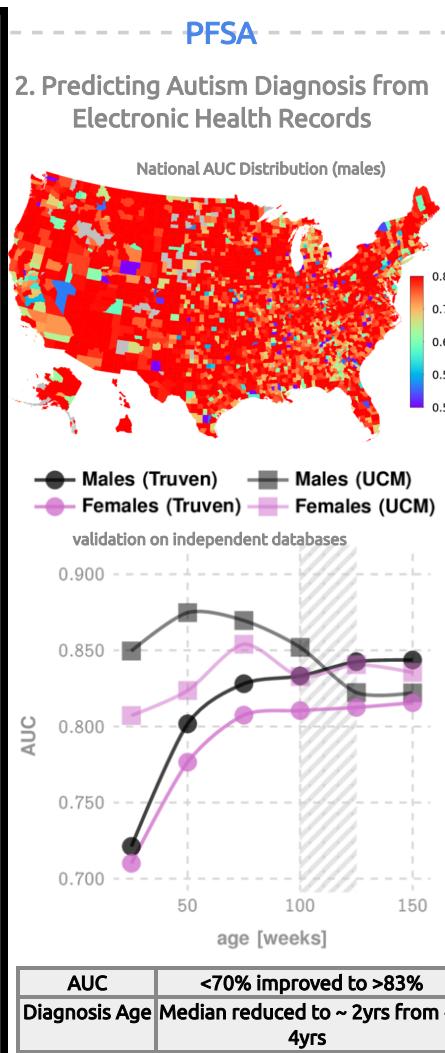
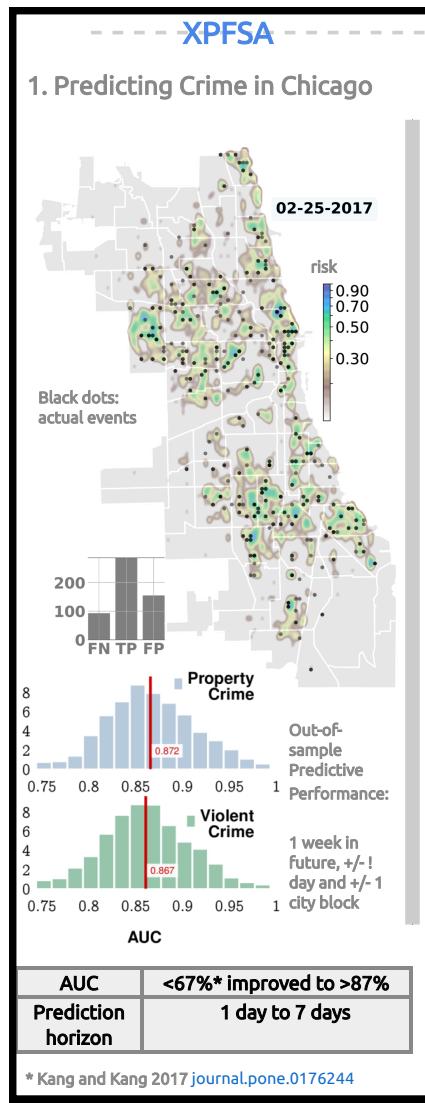


Genetic Screening	Deep Patient	UChicago PAI
AUC~ 83%	AUC for neuropsychiatric conditions: < 70%	AUC: ~85%
Blood reqd.	Blood not reqd.	Blood not reqd.
n~ 1K	n~ 1M	n~ 1M



Hypothesis: Decomposing the learning problem, to first estimate unknown problem structure, followed by standard inference leads to faster, better, efficient learning

Key Insight: Notion of structure changes between problem domains, but basic idea remains valid



- Publications**
- "Beyond edit distance: A function aware adaptive distance metric for evolving populations", T. Li, P. Leon, A. Weinberger, and I. Chattopadhyay **In Preparation**
 - "QNN: Using Quasi-species Sequence Variations to Inform Neural Net Modeling of The Genotype to Phenotype Problem", T. Li, I. Chattopadhyay **In preparation**
 - "Zero-burden Digital Biomarkers For Autism: Exploiting Co-morbidity Patterns To Drive Early Intervention", D. Onischenko, Y. Huang, P. Smith, and I. Chattopadhyay **Under Review in Nature Medicine**
 - "Distilling Event Logs For Actionable Prediction of Urban Crime", T. Li, Y. Huang, J. Evans, and I. Chattopadhyay **Submitted to Science**

Patent:
Provisional patent application submitted for EHR based autism diagnosis

Software: quasinet ehrzero cynet

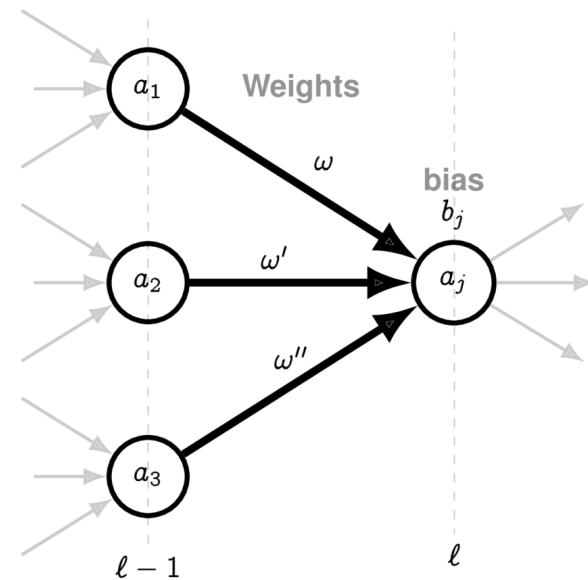
```

1 pip install quasinet
2 pip install ehrzero
3 pip install cynet

```

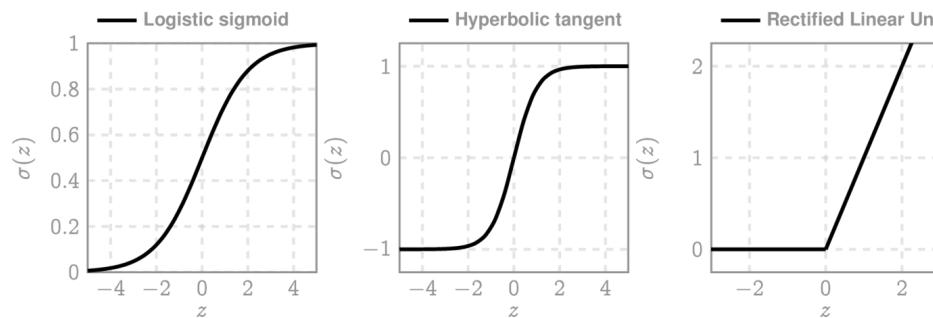


Problem 3: Spatio-temporal Learning



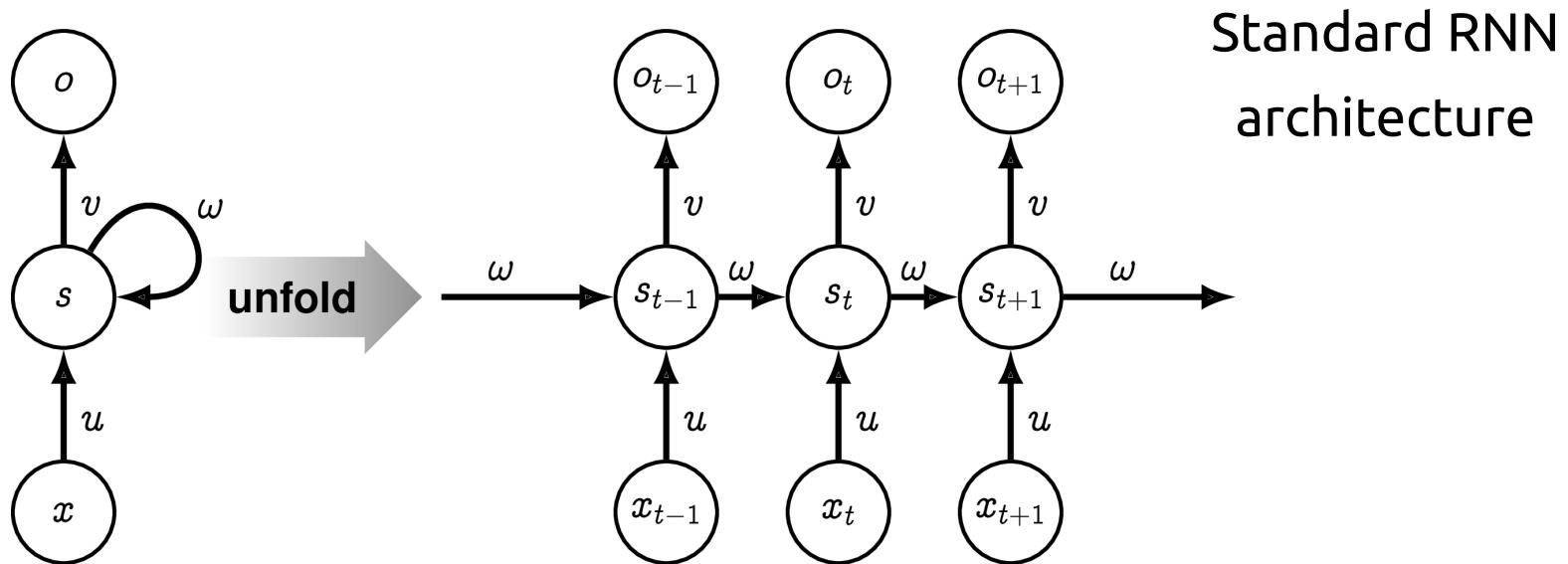
Standard NN
architecture

$$a_j^\ell = \sigma \left(\sum_k \omega_{jk}^\ell a_k^{\ell-1} + b_j^\ell \right)$$
$$a^\ell = \sigma(\omega^\ell a^{\ell-1} + b^\ell)$$





Spatio-temporal Learning

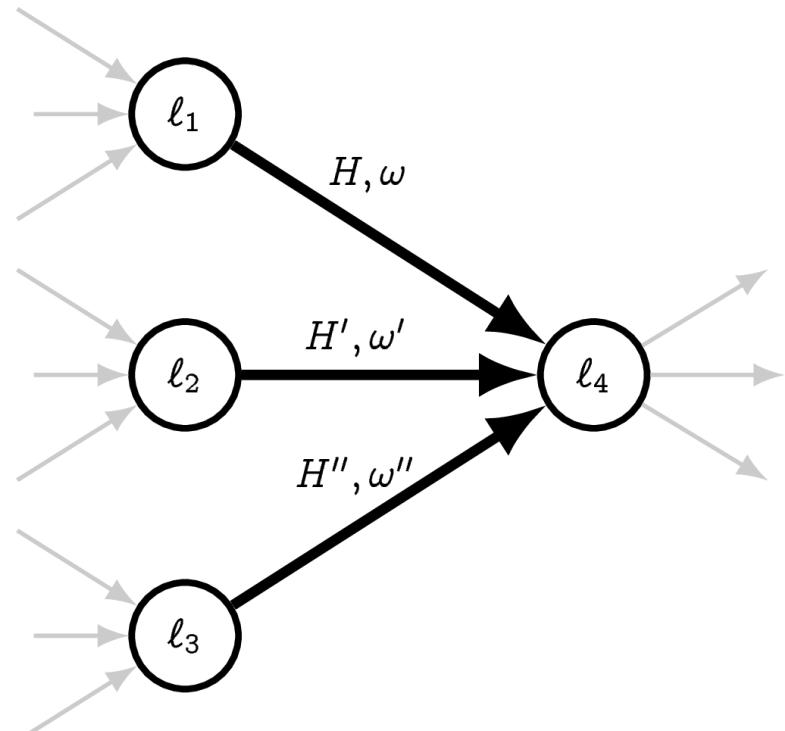


$$s_t = \sigma(b + \omega s_{t-1} + ux_t)$$

$$o_t = c + vs_t$$



Spatio-temporal Learning



Compare:

$$a^\ell = \omega^\ell H^\ell(a^{\ell-1}) + b^\ell$$

↑
non-linearities
↓

$$a^\ell = \sigma(\omega^\ell a^{\ell-1} + b^\ell)$$

1. Identify the local activation functions

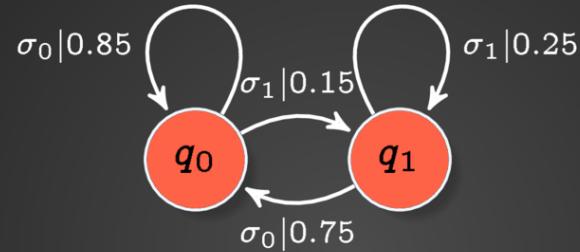
H

2. Learn the linear combination weights



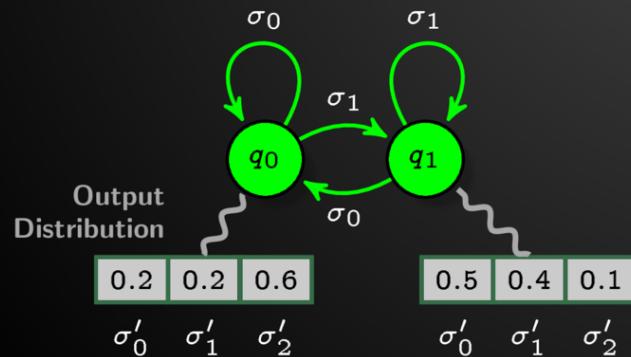
Spatio-temporal Learning

A. Probabilistic Finite State Automata

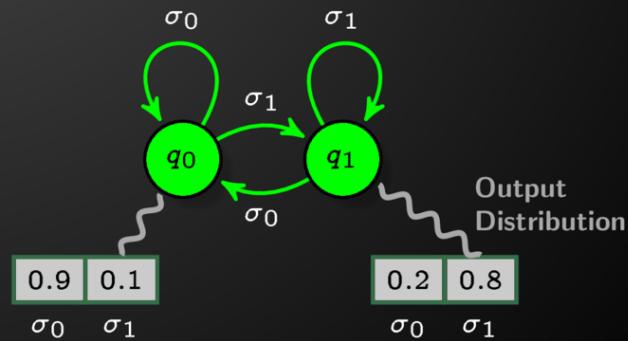


Crossed Probabilistic Finite State Automata (XPFSA)

B. (3-letter output alphabet)

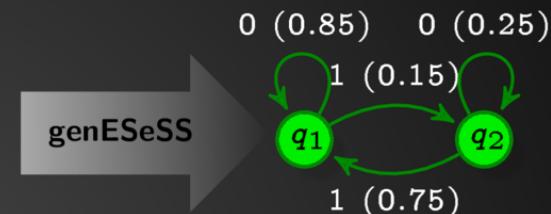


C. (2-letter output alphabet)



Spatio-temporal Learning

```
01100000000011000010011010001100000000110000011000101101100
011000001011011000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
001100101010000011100000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000000111001100000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000101000011000110010000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000000000000000
000010100001100011001000000000000000000000000000000000000000000
```



A. Stochastic process on two letter alphabet

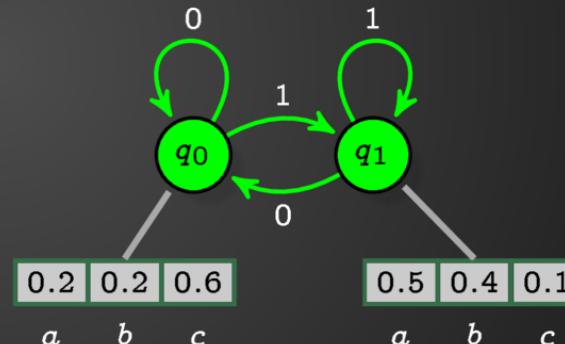
xgenESeSS

Predict

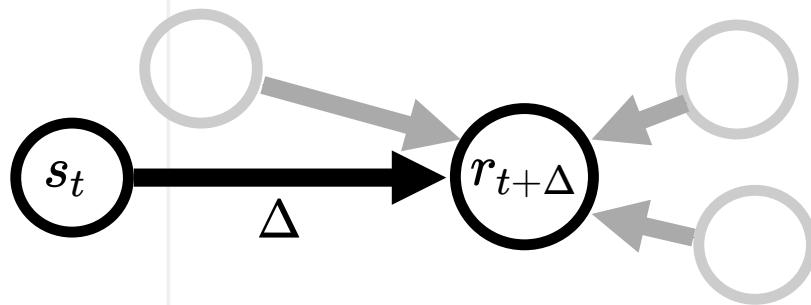


aaababbacababababcacacacccaaaabb ? bbababbbcaccacacccabbbacccacabb

cabbaccccabbacbacbabacacabbababccacababcacaccaabbababacaccabbacac
abcabcbbccaccacabcbbabababbacababababcacacacccccaaaabbbbacccacabb
ababbbcaccacacccabbacccabbacacbbacccaacacabbabccacababcacac
caabbacaccabbbacacabcbbccbccaccacabcababababcacacacccccaaaabbbbacccaca
cababcacaccaabbababacaccabbacacabcbbccbccaccacabcbbabababbacabab



B. Stochastic process on 3 letter alphabet



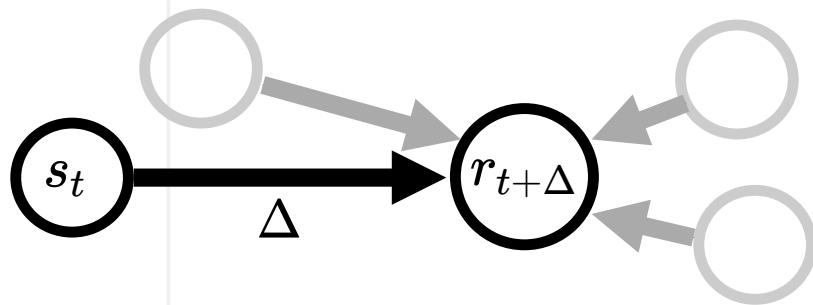
Weighted
probabilistic
transducers

$$G = s^{\models} \vec{\otimes} H_{r, \Delta-j}^s \quad (1)$$

$$\bar{r}_{t+\Delta}^{\bar{s}_{t_0, t+j}} \doteq [\lambda]^G \left(\prod_{\sigma \in \bar{s}_{t_0, t+j}} \Gamma_\sigma^G \right) \widetilde{\Pi}^{H_{r, \Delta-j}^s} \quad (2)$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \in (-t_0 + t, 0]}} \omega_{j,s} \bar{r}_{t+\Delta}^{\bar{s}_{t_0, t+j}} \quad (3)$$

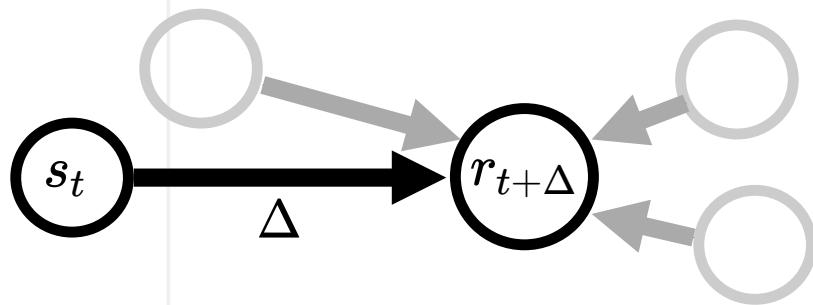




$$\bar{r}_{t+\Delta}^{\bar{s}_{t_0,t+j}} \doteq [\lambda]^G \left(\prod_{\sigma \in \bar{s}_{t_0,t+j}} \Gamma_\sigma^G \right) \widetilde{\Pi}^{H_{r,\Delta-j}^s} \quad (2)$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \in (-t_0+t, 0]}} \omega_{j,s} \bar{r}_{t+\Delta}^{\bar{s}_{t_0,t+j}} \quad (3)$$

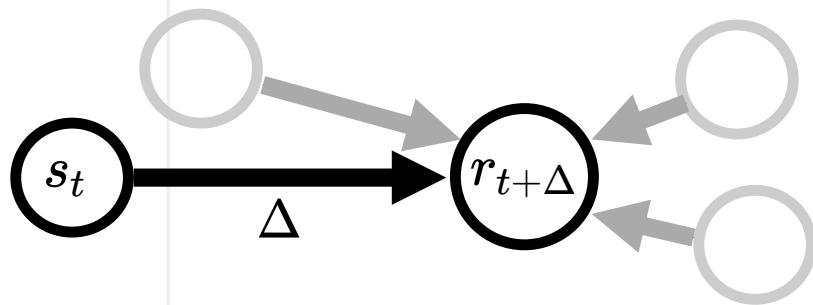




$$\bar{r}_{t+\Delta}^{\bar{s}_{t_0,t+j}} \doteq [\lambda] \left(\prod_{\sigma \in \bar{s}_{t_0,t+j}} \Gamma_\sigma \right) \widetilde{\Pi}^{H_{r,\Delta-j}^s} \quad (2)$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \in (-t_0+t, 0]}} \omega_{j,s} \bar{r}_{t+\Delta}^{\bar{s}_{t_0,t+j}} \quad (3)$$

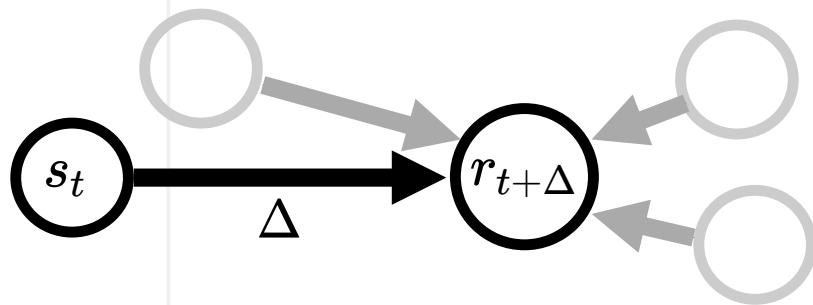




$$r_{t+\Delta}^{s_{t_0,t+j}} \doteq [\lambda] \left(\prod_{\sigma \in s_{t_0,t+j}} \Gamma_\sigma \right) \widetilde{\Pi}^{H_{r,\Delta-j}^s} \quad (2)$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \in (-t_0+t, 0]}} \omega_{j,s} r_{t+\Delta}^{s_{t_0,t+j}} \quad (3)$$

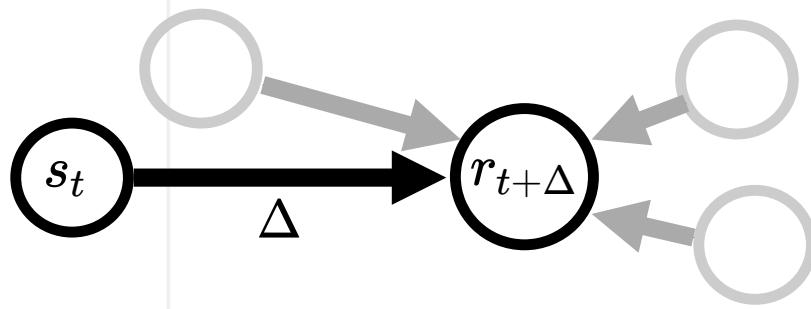




$$r_{t+\Delta}^{s_{t_0,t+j}} \doteq [\lambda] \Gamma_{s_{t_0,t+j}} \widetilde{\prod} H_{r,\Delta-j}^s \quad (2)$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \in (-t_0+t, 0]}} \omega_{j,s} r_{t+\Delta}^{s_{t_0,t+j}} \quad (3)$$





$$r_{t+\Delta}^s \doteq [[\lambda]] \Gamma_s \tilde{\Pi}^H$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \leq 0}} \omega_{j,s} r_{t+\Delta-j}^s$$



$$r_{t+\Delta}^s \doteq [\![\lambda]\!] \Gamma_s \widetilde{\Pi}^H$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \leq 0}} \omega_{j,s} r_{t+\Delta-j}^s$$



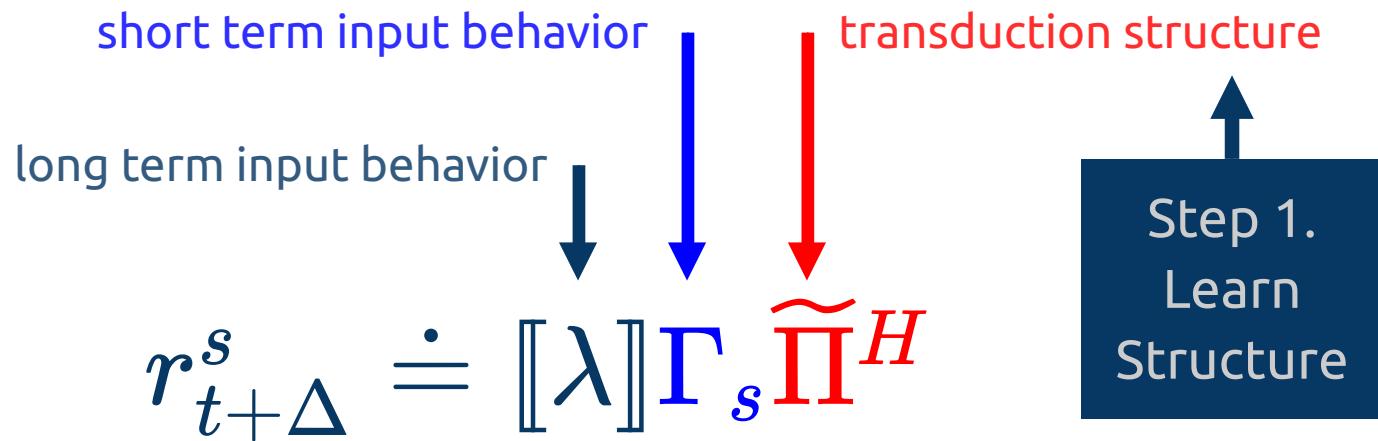
$$r_{t+\Delta}^s \doteq [[\lambda]] \Gamma_s \widetilde{\Pi}^H$$

↓ transduction structure

Step 1.
 Learn
 Structure

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \leq 0}} \omega_{j,s} r_{t+\Delta-j}^s$$





$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \leq 0}} \omega_{j,s} r_{t+\Delta-j}^s$$



$$r_{t+\Delta}^s \doteq [\lambda] \Gamma_s \widetilde{\Pi}^H$$

$$r_{t+\Delta} = \sum_{\substack{s \in \mathcal{S} \\ j \leq 0}} \omega_{j,s} r_{t+\Delta-j}^s$$

↑

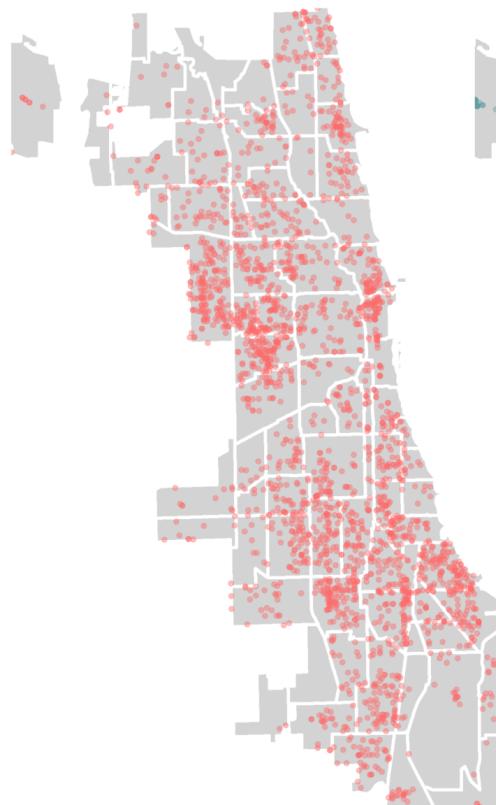
Step 2.
Learn
weights



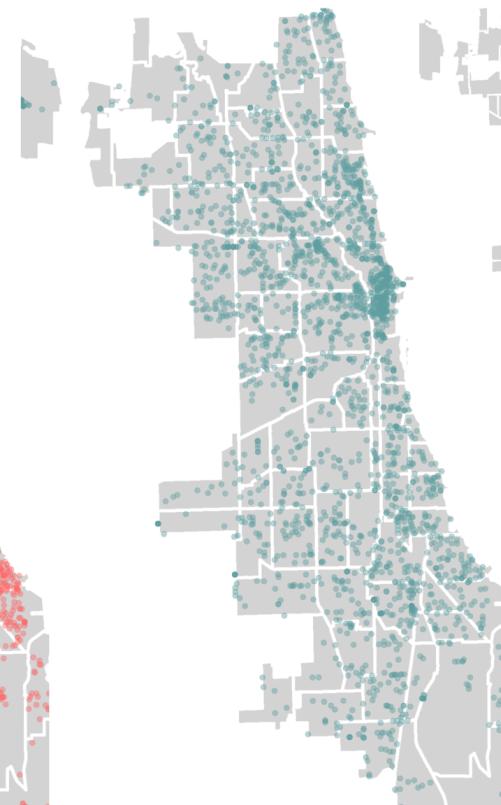


Predicting Crime in Chicago

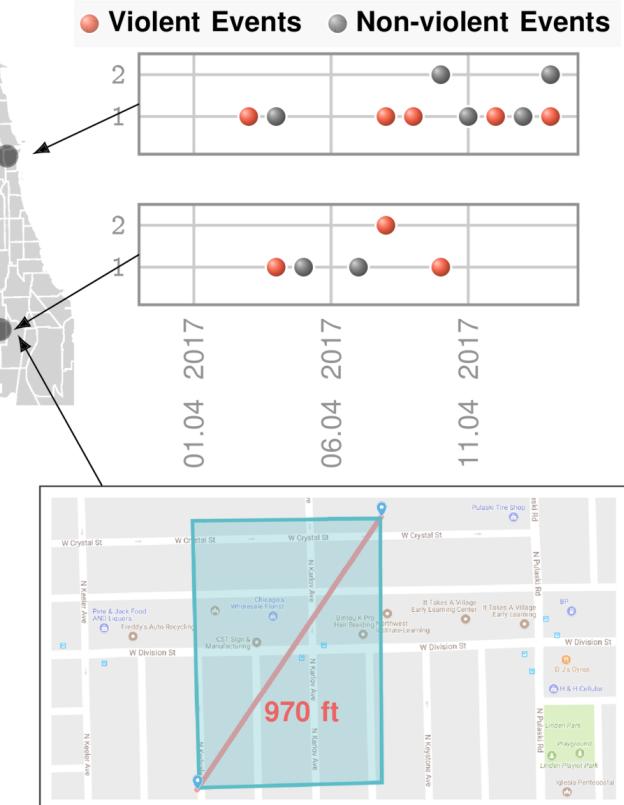
A. Violent Crimes including
Assaults, Battery & Homicides
(April 1-15, 2017)



B. Non-violent Crimes including
Thefts & Burglaries
(April 1-15, 2017)



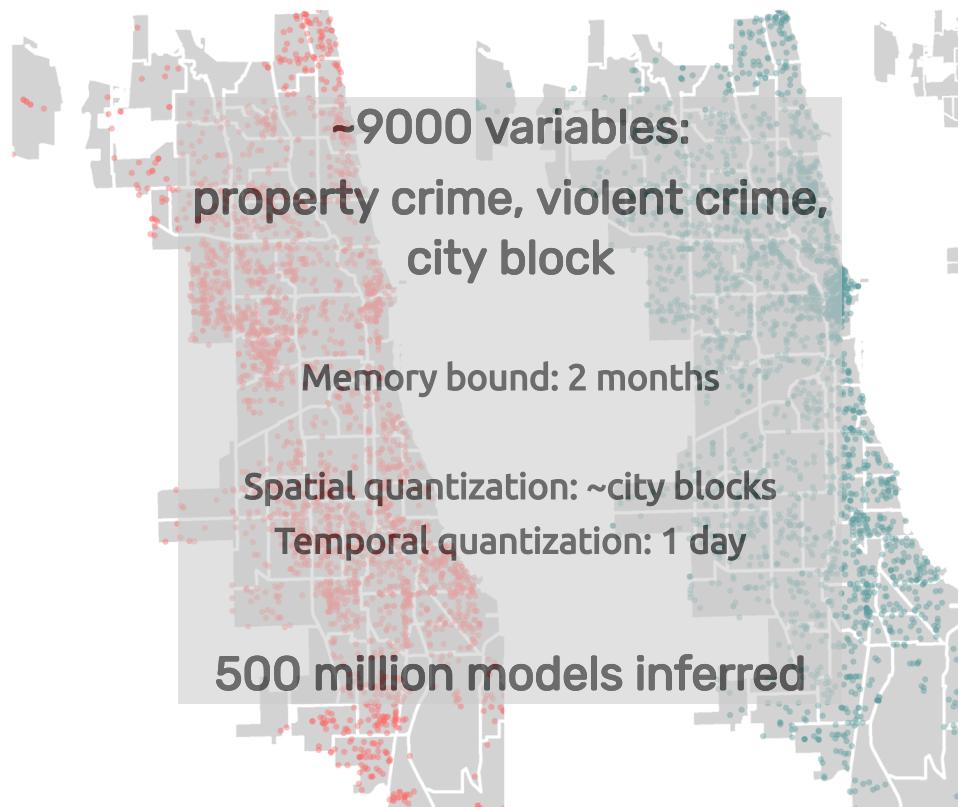
C. Spatio-temporal Modeling Approach Using
Daily Event Counts &
Spatial Tiles ≈ City Blocks





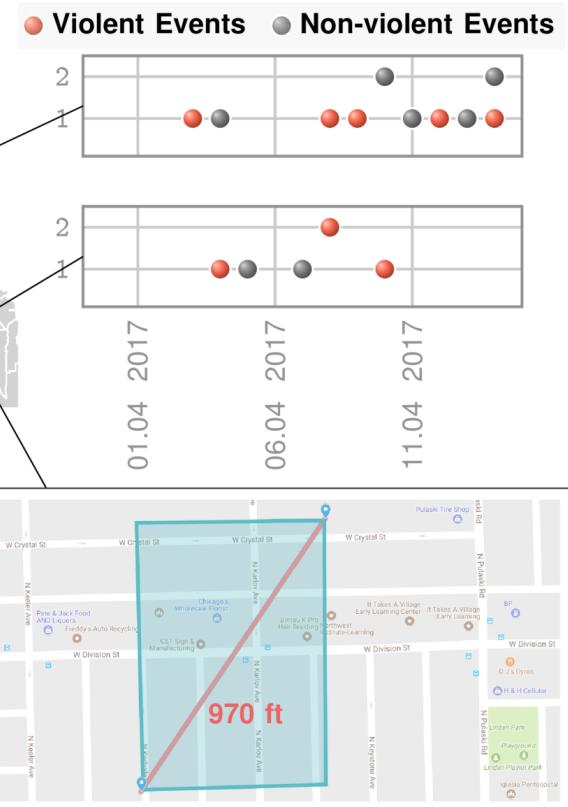
Predicting Crime in Chicago

A. Violent Crimes including
Assaults, Battery & Homicides
(April 1-15, 2017)



B. Non-violent Crimes including
Thefts & Burglaries
(April 1-15, 2017)

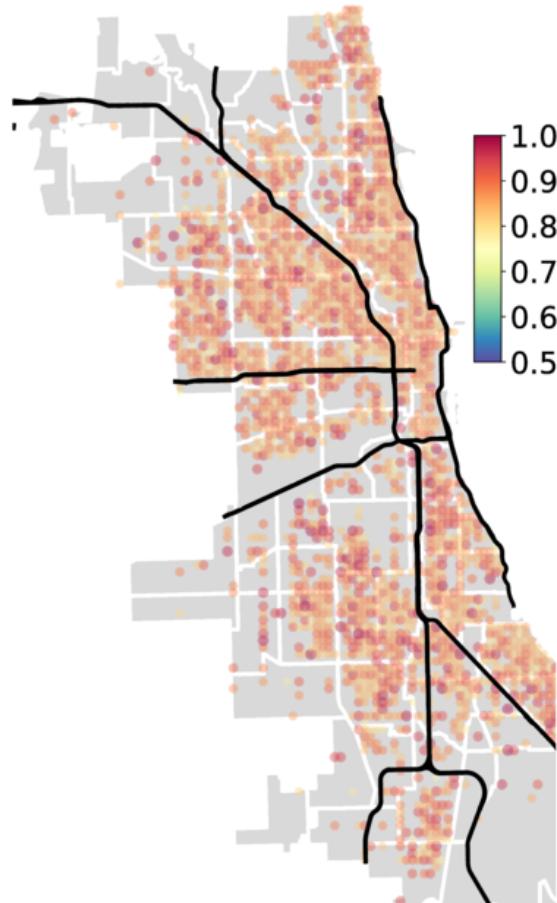
C. Spatio-temporal Modeling Approach Using
Daily Event Counts &
Spatial Tiles ≈ City Blocks



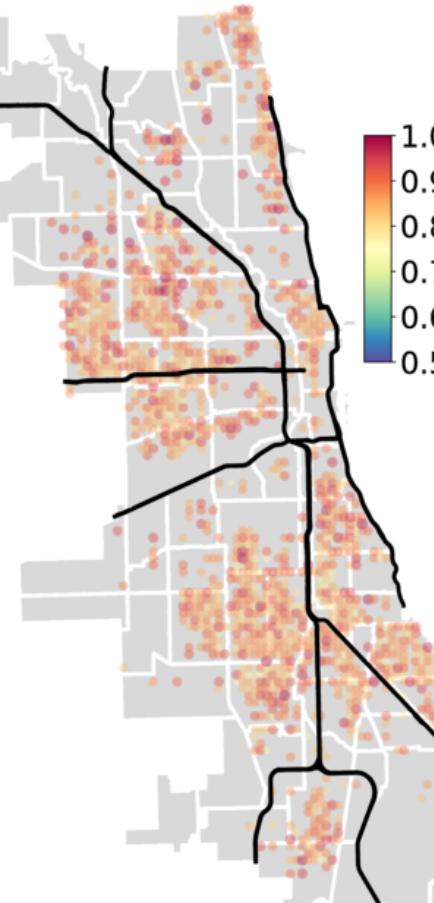


Predicting Crime in Chicago

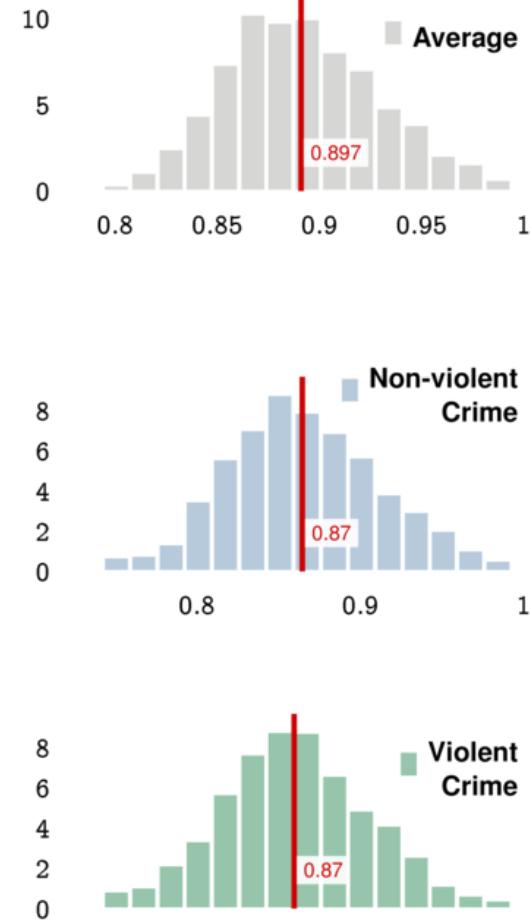
A. Spatial Distribution of AUC for Property crimes



B. Spatial Distribution of AUC for Violent Crimes



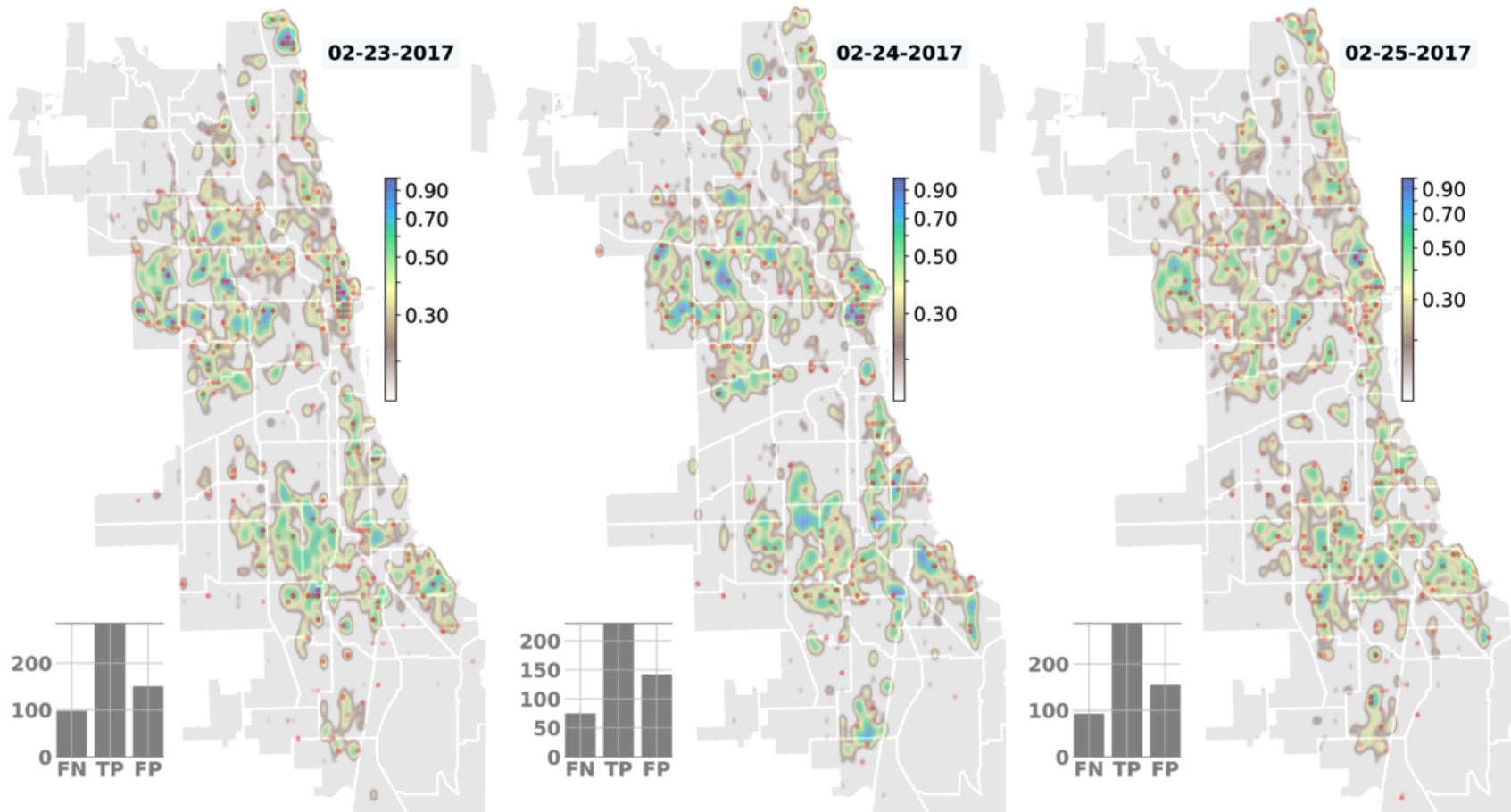
C. Distribution of out-of-sample AUC Across Spatial Tiles





Predicting Crime in Chicago

1 week in advance, +/- 1 day, +/- 1 City Blocks

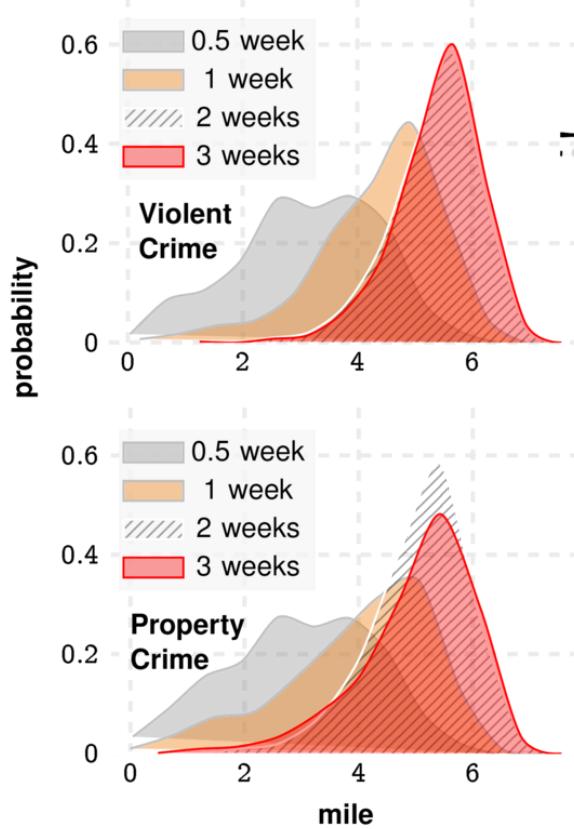




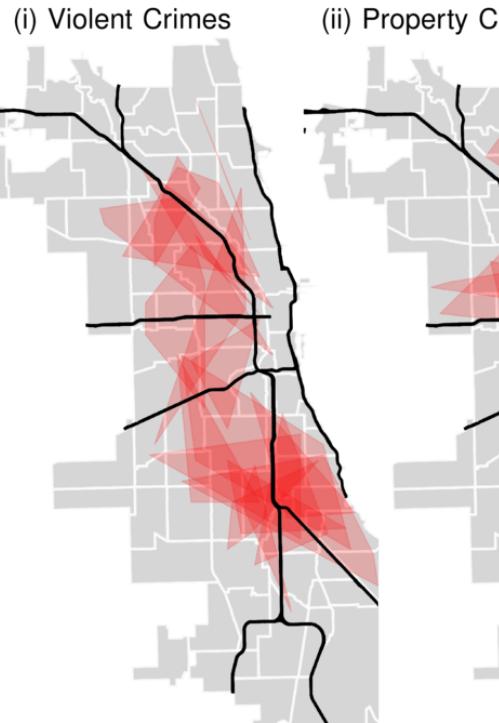
Predicting Crime in Chicago

Spatial Scales

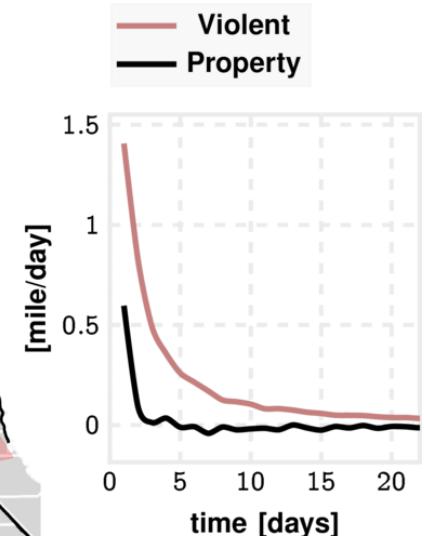
A. Distribution of Influence Radius



B. Inferred Neighborhood Samples
(≤ 3 day influence period)



C. Inferred Influence Diffusion Rates

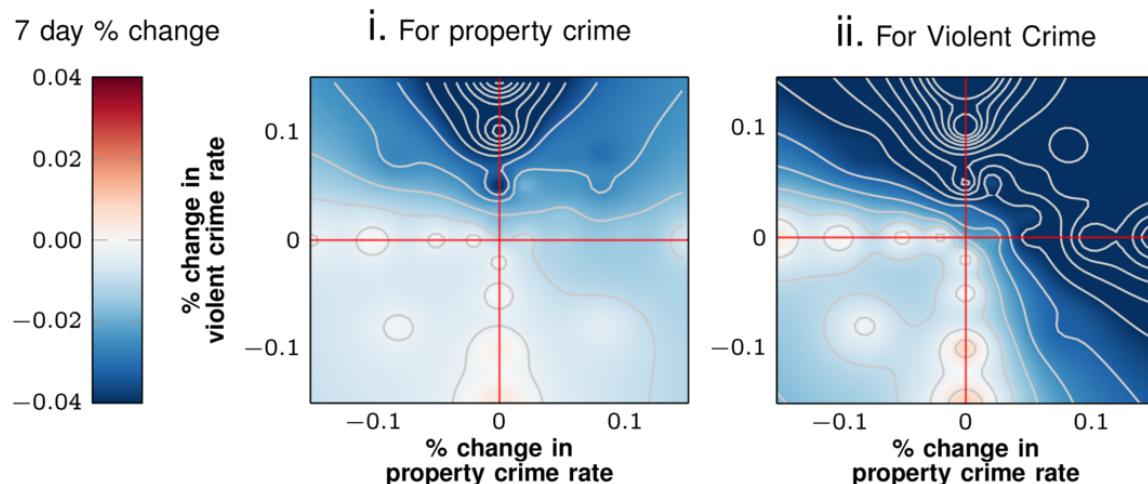




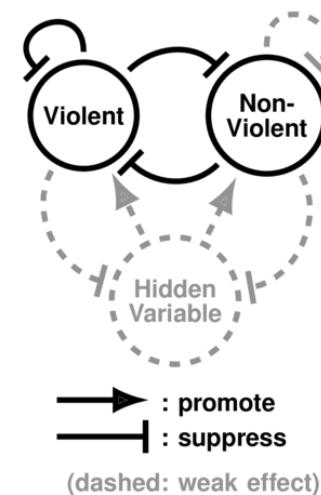
Predicting Crime in Chicago

Perturbation Space

D. Perturbation Space For the Dynamics of Criminal Infractions



E. Inferred Dependency Model

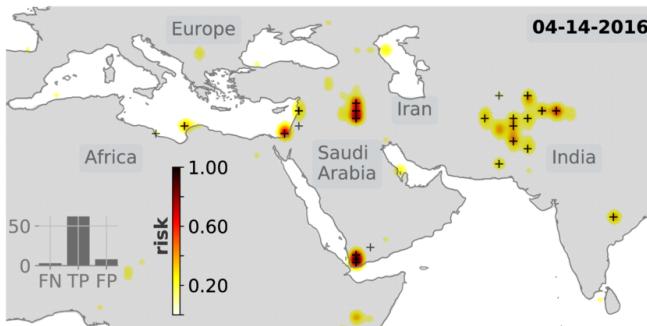
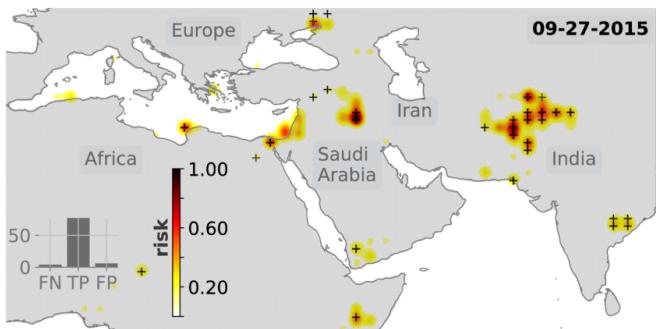


1. Some degree of stability in the emergent dynamics
2. Future: Intervention/policy design based on inferred dependencies

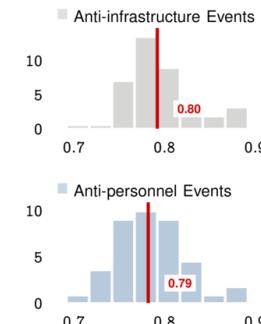


Modeling Terrorism

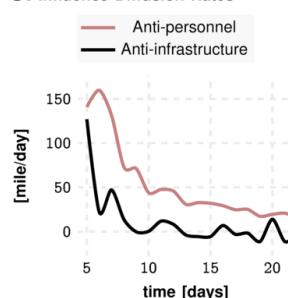
A. Out-of-sample Terror Event Predictions (Temporal res.: 1 day, spatial res.: $1^\circ \times 2^\circ$)



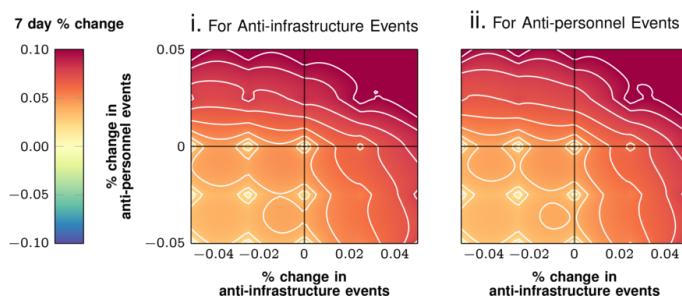
B. Out-of-sample AUC



C. Influence Diffusion Rates



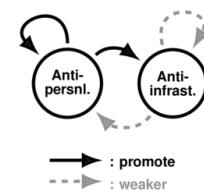
D. Perturbation Space for Terrorism Event Dynamics



Source: Global
Terrorism Database

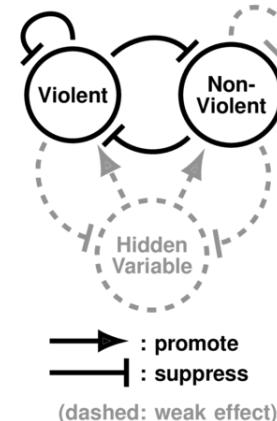
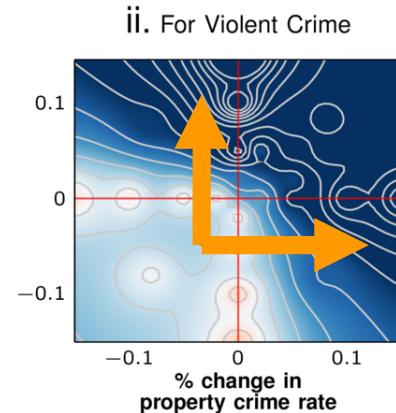
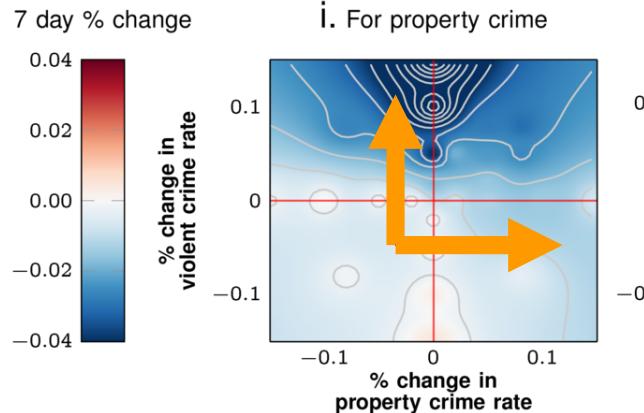
Spatial Discretization: 1
X 2 degrees

Temporal
Discretization: 1 day

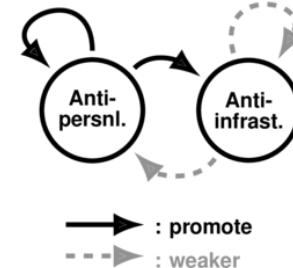
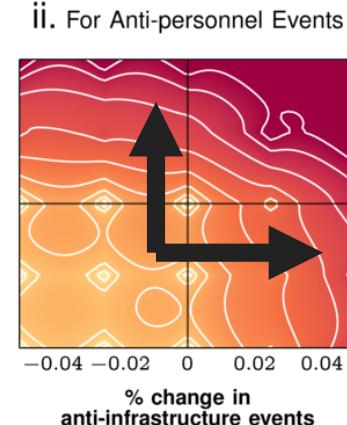
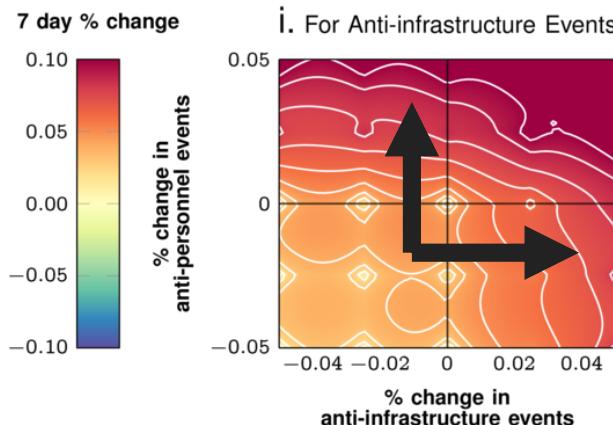




Comparative Insights



Crime



Terrorism



State of the Art Comparison

	SOA	UChicago PAI
Crime Prediction:	AUC: ~67% with CNN * TPR < 10% **	AUC: ~85%-99% (Individual block performance, with 3 years of data) TPR > 80% with 20% FPR
Terrorism:	NA	AUC: ~80%

* Kang, H.-W., & Kang, H.-B. (2017). Prediction of crime occurrence from multi-modal data using deep learning. Plos One, 12(4), e0176244. <https://doi.org/10.1371/journal.pone.0176244>

** Mohler, G. O. et al. Randomized controlled field trials of predictive policing. Journal of the American Statistical Association 110, 1399–1411 (2015).

Software



python packages

quasinet

pypi package 0.0.34

downloads 36/month

ehrzero

pypi package 1.0.36

downloads 123/month

cynet

pypi package 1.2.0

downloads 229/month

```
1 pip install quasiynet --user
2 pip install ehrzero --user
3 pip install cynet --user
4 pip install entropyrate --user
```



Baking in
Structure helps
a lot!