IEEE GlobalSIP, Dec 7-9 2016

Learning Structural Properties of Wireless Ad-Hoc Networks Non-Parametrically from Spectral Activity Samples

U.S. NAVAL RESEARCH LABORATORY

S. Kokalj-Filipovic , C. Acosta, M. Pepe

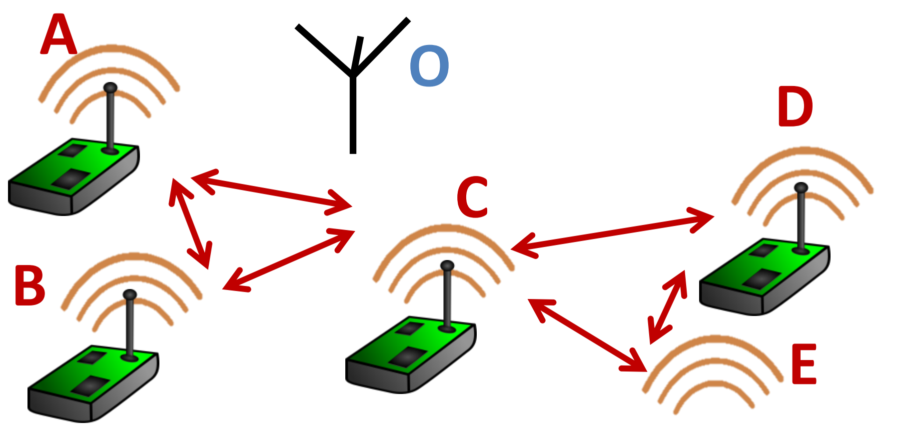
Naval Research Laboratory

System Model and Objective

U.S. NAVAL RESEARCH LABORATORY

A, B, C,…:
nodes in primary wireless network

O: Observer, a secondary user in cognitive network environment



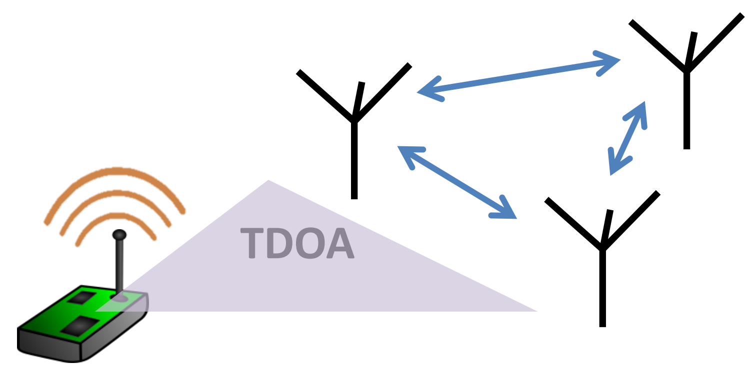
Objective:

To apply a Bayesian non-parametric model to segmenting time series of observed wireless node transmission activity in order to machine-learn routing patterns in an unknown wireless ad-hoc network, as well as its topology.

1. Observer O is the spectrum sensing agent

2. Observer is capable of associating each transmission, i.e., geo-locating sources

3. Observer may be another network



4. Observer O is the provider of transmission activity observations to the learning agent

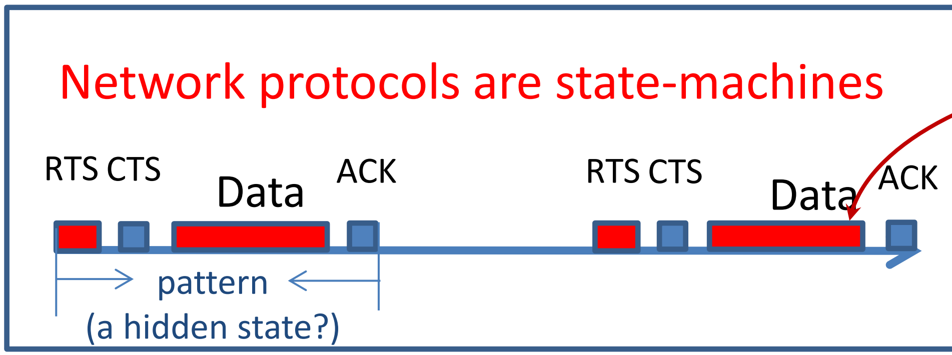
System Model – Summary

U.S. NAVAL RESEARCH LABORATORY

- We refer to the Learning Agent as Observer, event when it is another entity or network utilizing the observer's data

- Observer assumes a state-machine-like behavior of a primary node (PN), which is driven by the network inputs

  -motivates generative model (**HMM/ HSMM**)

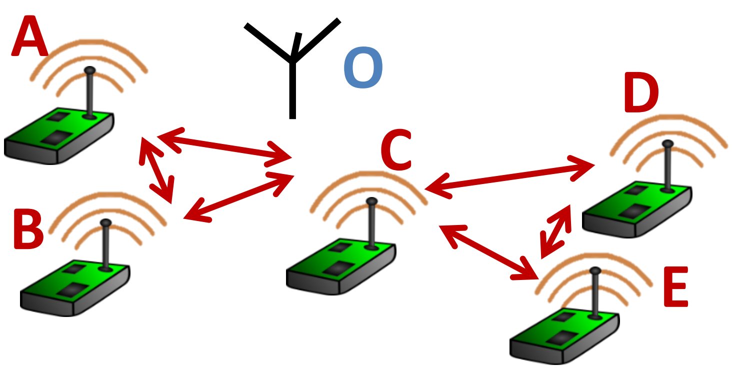


time series of packet-delimited RF transmissions obtained by energy detection: ***Tx-series***

- Observer does not have prior knowledge of the observed Primary Network

  - this motivates Bayesian non-parametric learning model (**HDP – HSMM**)

  - HDP expresses our prior beliefs about the node based on very loose assumptions



HDP: ***Hieararchical Dirichlet Process***

Analytical Model

U.S. NAVAL RESEARCH LABORATORY

EXPECTATION:

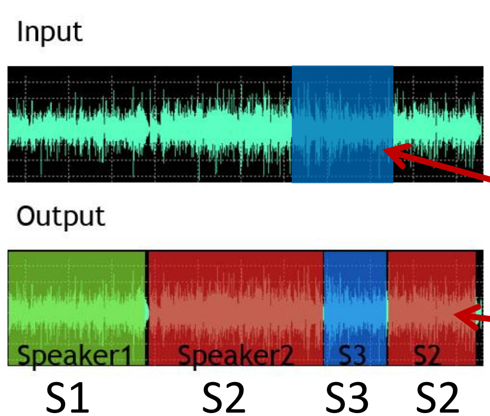
1. Learned states of the HDP-HSMM model will correspond to behavioral states of a wireless network node, and sequences thereof will correspond to activity patterns

2. Utilizing these activity patterns as features for node classification (clustering) will lead to recognition of network routing paths, including sources and sinks of the traffic, and to topology recognition, including edge and hub nodes.

CHALLENGES:

Segmenting the time series of observed outputs blindly, and labeling their latent sources as states of an unknown Markov Chain

HDP-HMM: Used for Speaker Diarization

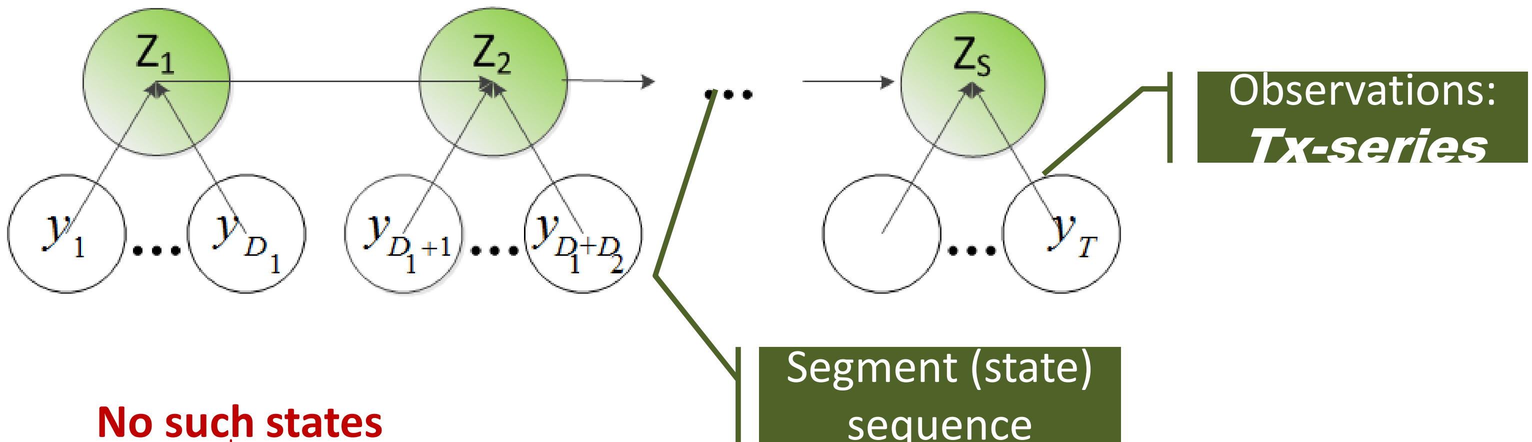


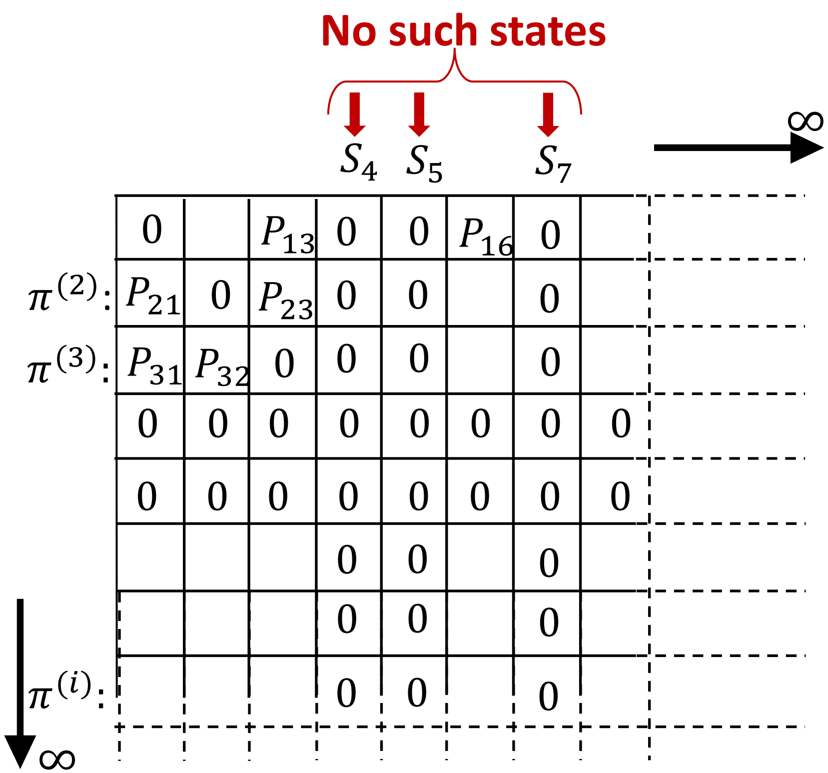
Key questions:
How do we know
Speaker 3 (S3)
is not this?

How do we know
this is S2 and not S5?

HDP concept applied to hidden matrices allows for learning of the number of states and what they are

HDP-HMM: Learning State Dynamics Non-Parametrically





S_i: defined by distribution of observations $\overrightarrow{y_{S_i}}$

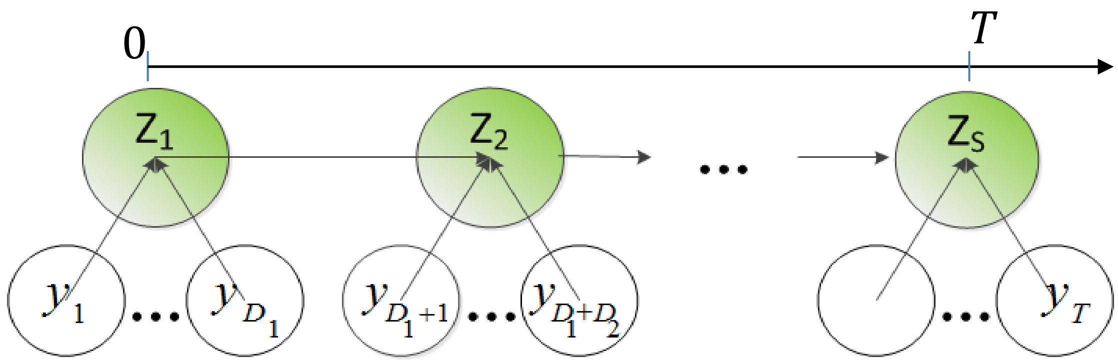
$\forall\pi^{(i)}$ is a DP (Dirichlet Process) with prior G_i

$\pi^{(i)}$s are coupled by the same prior G_o into a hierarchy of state transition probabilities (HDP):

$G_o \sim DP(\gamma, H)$; $G_i \sim DP(\alpha, G_o)$

H-base measure, γ – concentration parameter

HDP-HMM:
Learning State Dynamics Non-Parametrically



$$T = \sum_{i=1}^{s} D_i$$

$\overrightarrow{y_{S_i}}$: 2-D discrete samples described by multinomial prob. distribution specific to state S_i.

1st multinomial process outputs quantized packet lengths:

$$\widehat{L^{(i)}} = Q_L(L^{(i)})$$

2nd outputs quantized inter-packet times:

$$I^{(i)} = T_{st}^{(i)} - T_{end}^{(i-1)}$$

$$\widehat{I^{(i)}} = Q_G(I^{(i)})$$

Q_G, Q_G : quantization functions for packet lengths and inter-packet gaps

~~$D_i \sim Poiss(\lambda_{z_{i+1}})$~~

$D_i \sim NB(r, p)$

What is NB?

$$x = 1 + \sum_{i=1}^{r} v_i$$

$$P(v_i = k) = p^k(1-p)$$

$D_i \sim x$

Inference for HMM Models -outline-

U.S. NAVAL RESEARCH LABORATORY

- Evaluate observation sequence given the model $\vartheta = (A, B, \pi_o)$
- Find maximum likelihood parameters of the model ϑ
- Find the best fitting state sequence given your observations and your model ϑ

- z_t: state sequence (hidden)
- y_t : observation sequence (*Tx-series*)

HMM model

$\vartheta = (A, B, \pi_o)$ *A: state-transition probabilities*, *B state-dependent observation probabilities*

- $A = \{a_{ij}\}$ $a_{ij} = P(q_{t+1} = S_j \mid z_t = S_i)$, $1 <= i, j <= N$
- $B = \{b_{ij}\}$ $b_{ij} = P(y_t = v_j \mid z_t = S_i)$, $1 <= i <= N$

  $1 <= j <= M$ *(observable alphabet)*
- $\pi_o = \{\pi_{1o} \dots \pi_{No}\}$ $\pi_i = P(z_1 = S_i)$ *for* $1 <= i <= N$

  -initial probability of states-

- Our model is more complicated
  - $\{\theta_i^{\{(k)\}}, \lambda_i^{\{(k)\}}\}$ emission parameters (for 2-D Gaussian) and state duration parameters
  - We cannot easily integrate out parameters and create close form solutions
  - We use Gibbs sampling instead

How to Learn Posterior Probability of a HDP HSMM Model?

U.S. NAVAL RESEARCH LABORATORY

Markov Chain Monte Carlo (MCMC) methods used to derive the posterior:

- $B \sim GEM(\gamma)$
- $\pi^{(i)} \sim DP(\alpha, B)$, for any i
- $(\theta_i,\ \lambda_i) \sim (H \times G)$, for any i (denotes state)
- $z_s \sim \pi^{(i)}$

- H: conjugate prior of Multinomial (Dirichlet)
- $\theta_i = \eta_{1i}, \eta_{2i}$: Dirichlet Parameters for packet lengths and gaps

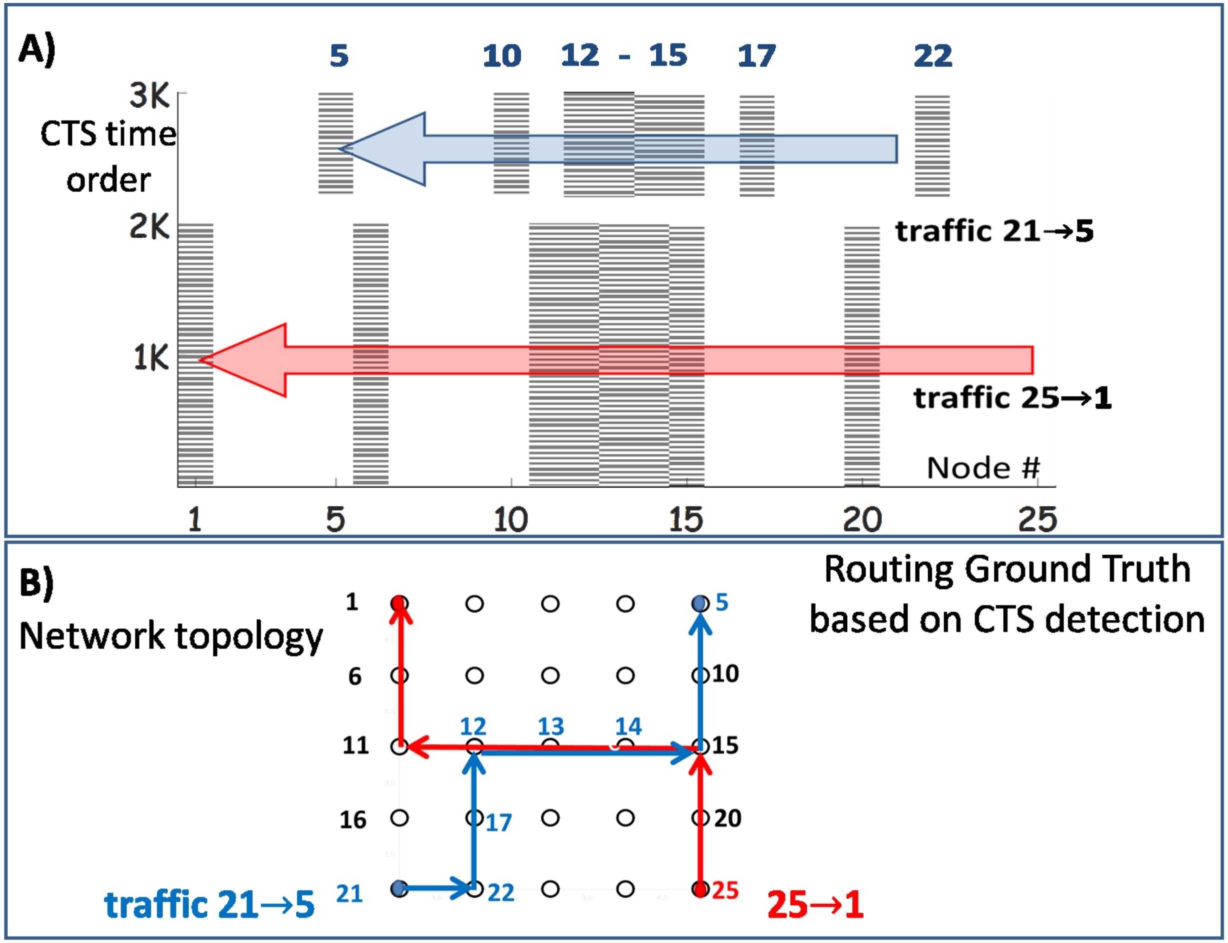
- G: *Beta (p),Gamma (r)* (conjugate priors of Negative Binomial)
- $\lambda_i = r_i, p_i$: Negative Binomial parameters

Gibbs sampling is a generic MCMC method that relies on knowing only the conditional marginal probabilities of the unknown parameters, e.g., the conditional probability of π given the current estimate of $\{\lambda_i\}, \{\theta_i\}$, and $\{z_k\}$.

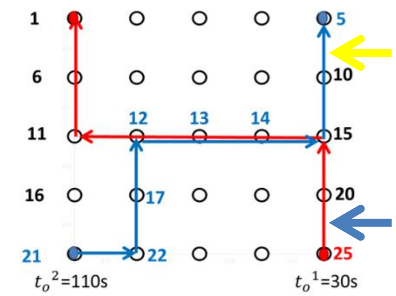
∀ sampling iteration outputs for ∀ node the latest $\pi, \{\lambda_i\}, \{\theta_i\}$, and $\{z_k\}$

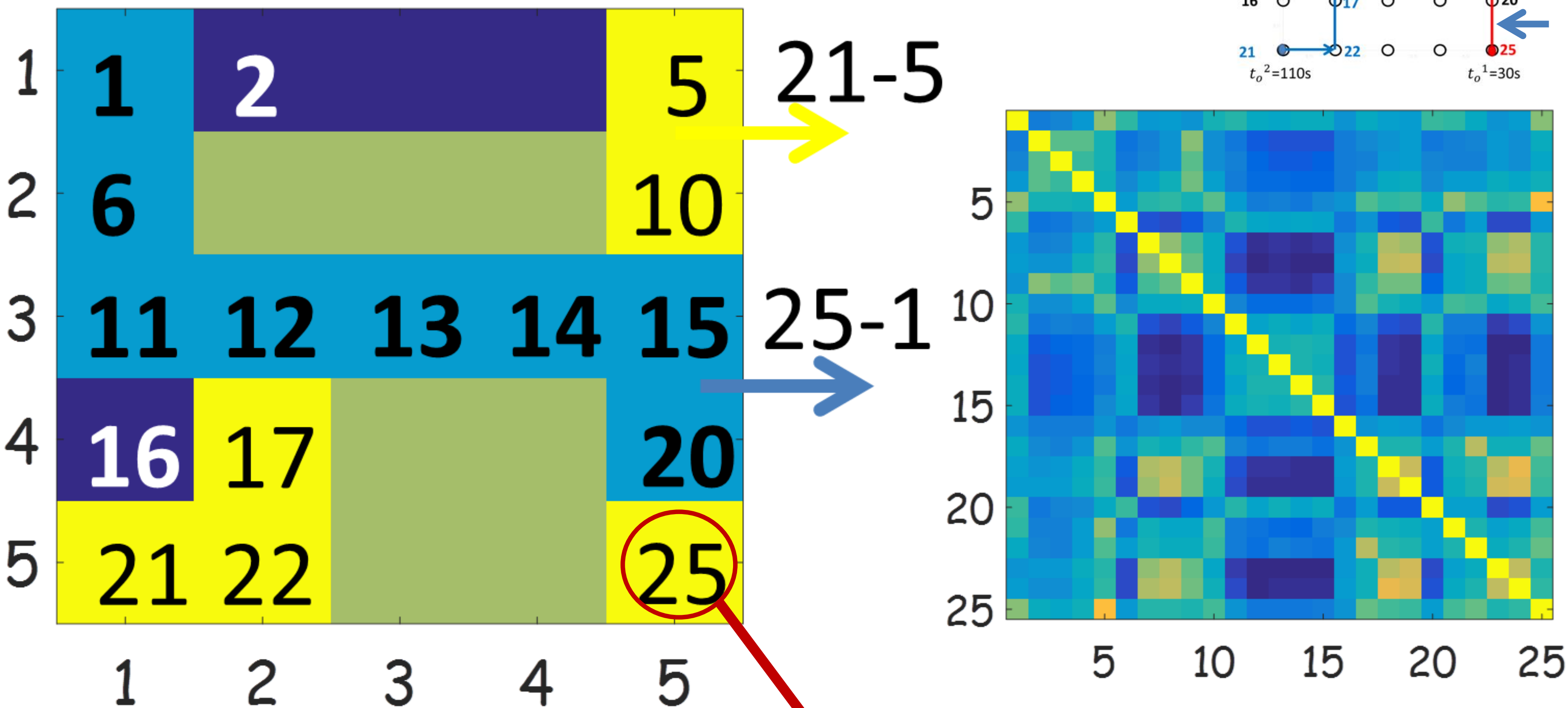
NS3 Simulations
Ground Truth

U.S. NAVAL RESEARCH LABORATORY



NS3 Simulations
Machine-Learned paterns





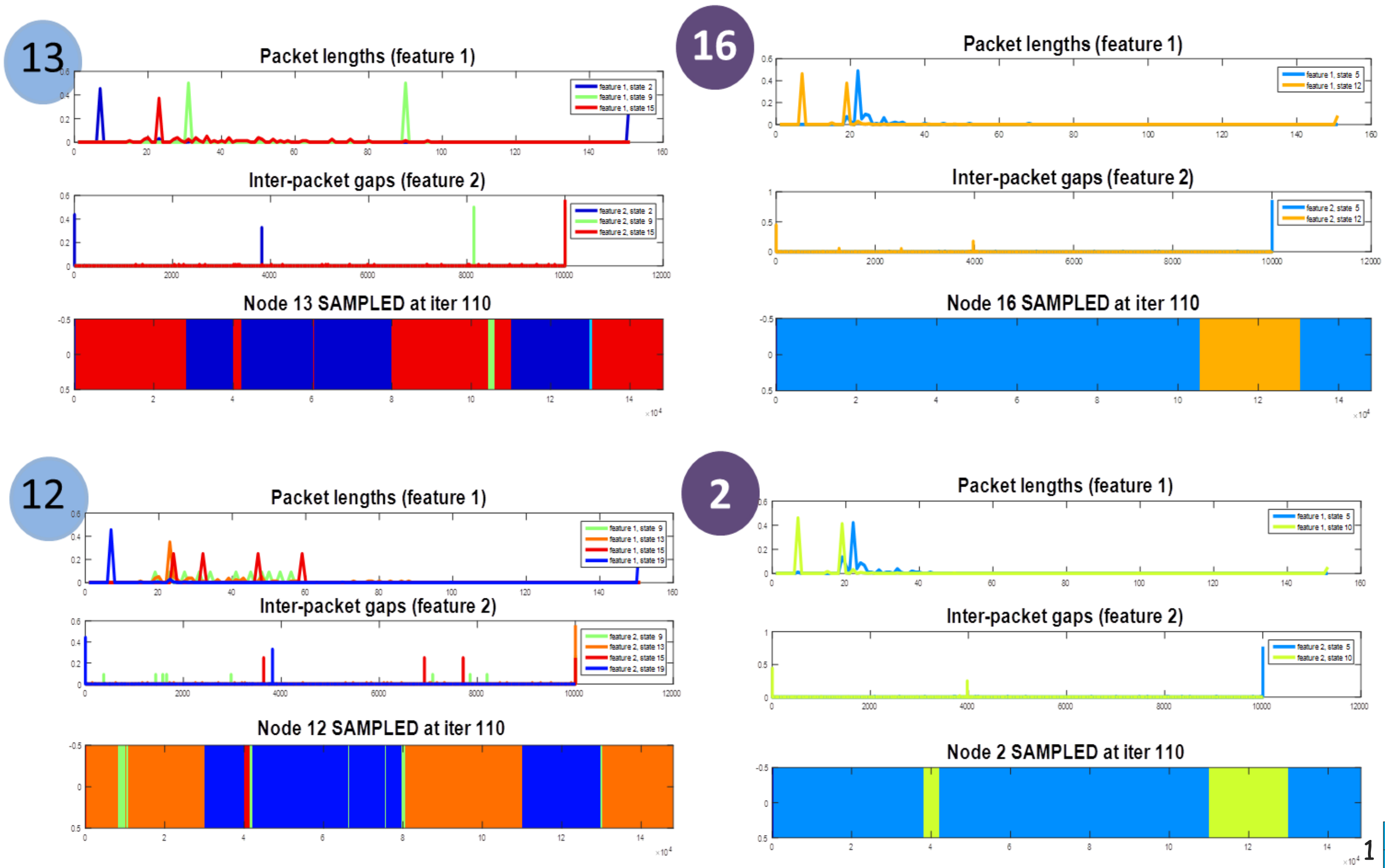
Color-coded clusters of similar nodes obtained by HDP-HSMM learning of their hidden states, followed by A-P clustering of the states, and a distance metric between new state sequences

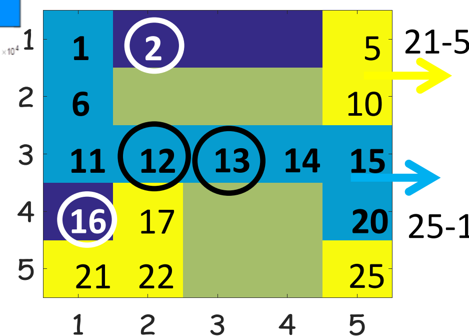
Heat map of max cross-correlation between Tx-series

misclassification due to small number of features, and feature selection

A-P Clustering for Common State-labeling Across Nodes

U.S. NAVAL RESEARCH LABORATORY





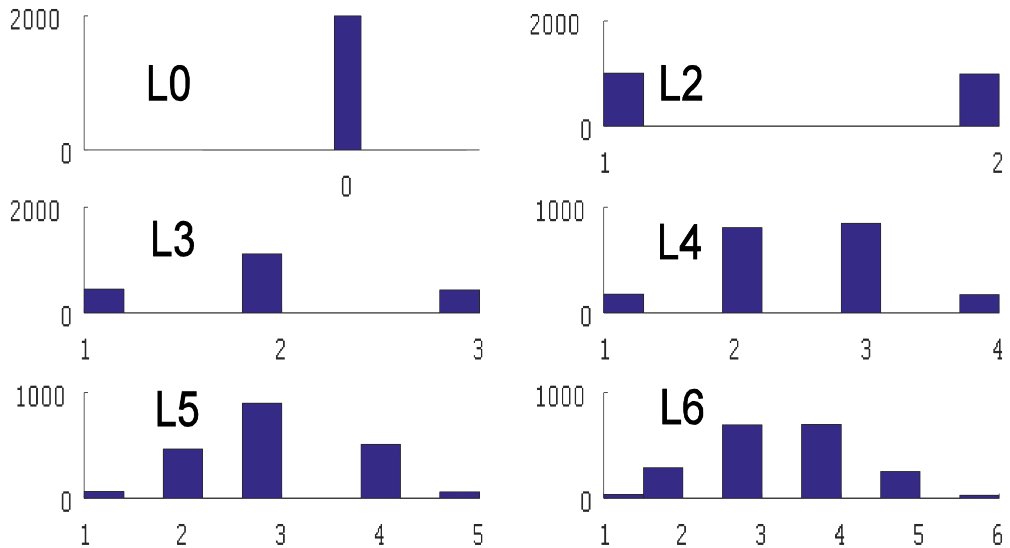
A-P*: Affinity propagation

* B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, 2007.

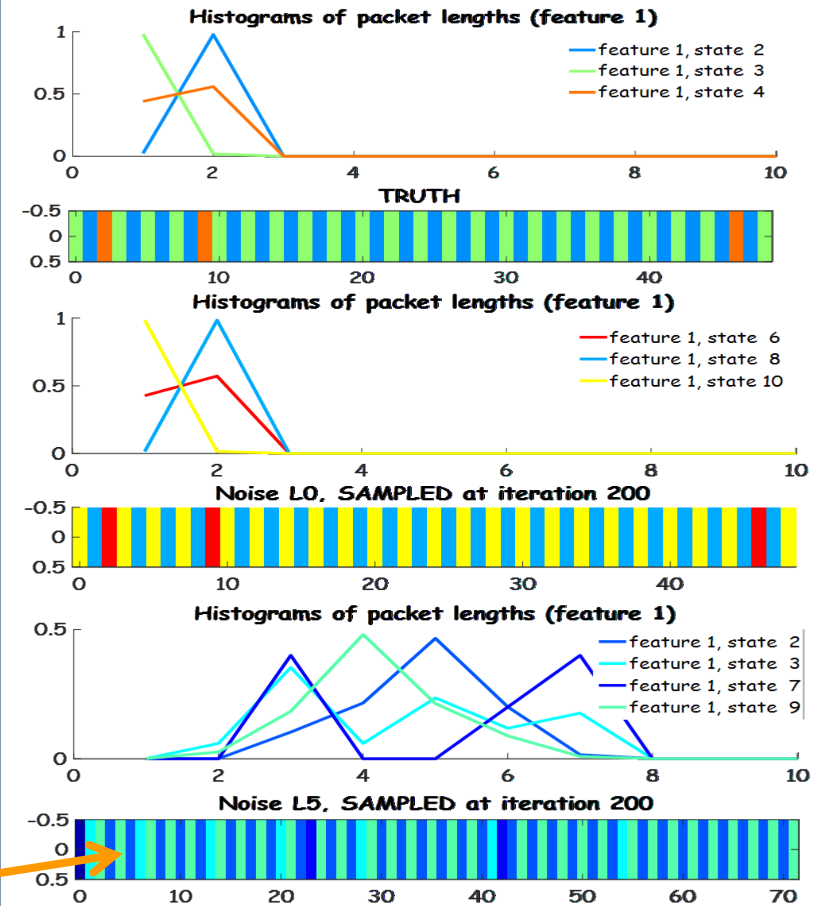
Synthetic Data: Noise Effect on Quality of Learning in HDP HSMM

U.S. NAVAL RESEARCH LABORATORY

We gradually apply measurement noise on synthesized Tx-series:



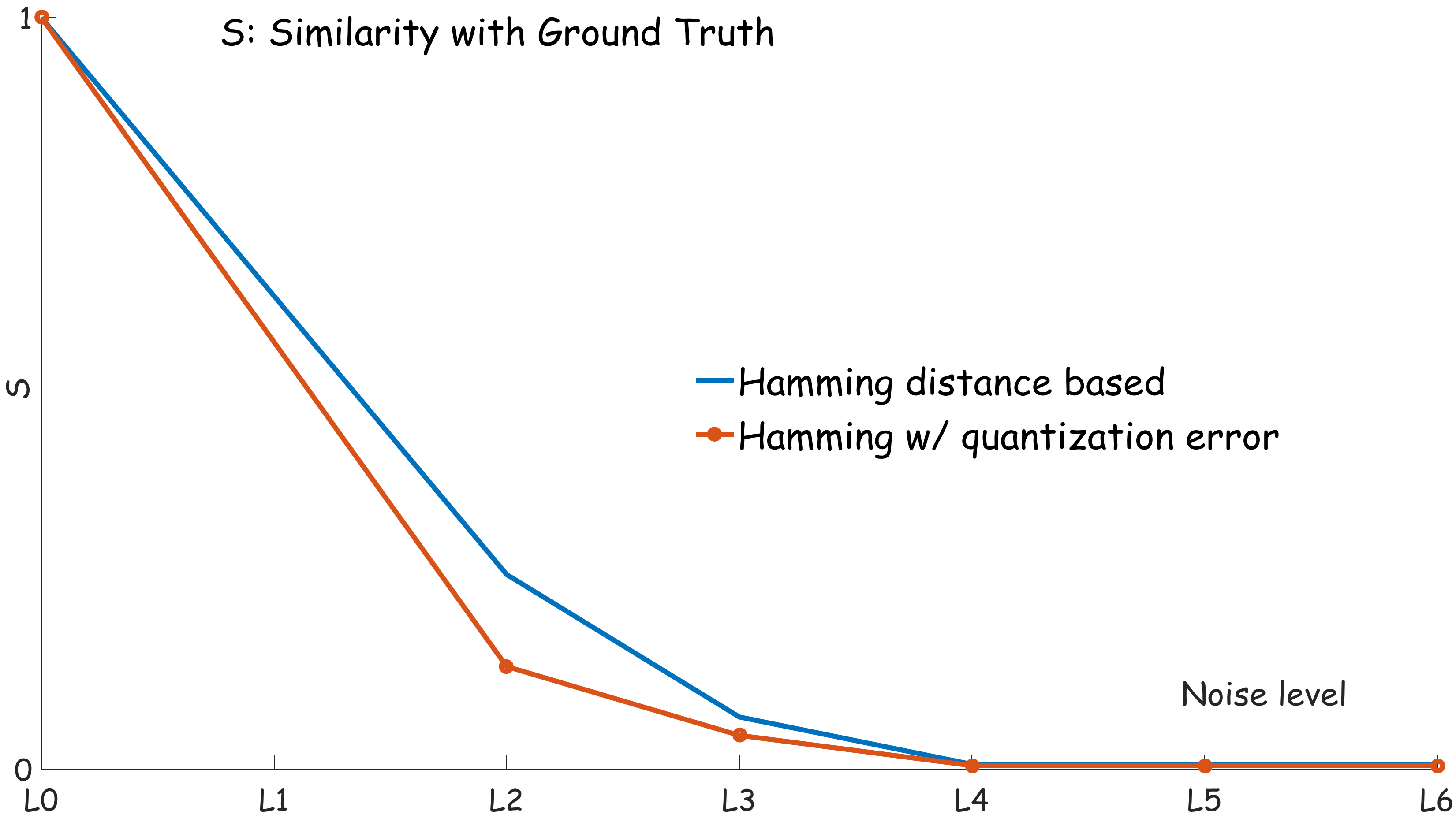
Levels of discrete measurement noise to multinomial samples L0-L6 (multinomials with mean 12; L0 corresponds to no noise)



incorrectly learned state sequence

L5 measurement noise causes both duration and emission distributions to be incorrectly learned

Quantified Noise Effect on Quality of Learning in HDP HSMM



Are there noise-resilient features, derived from $\widehat{L^{(i)}}$ and $\widehat{I^{(i)}}$?

Thanks!

- Questions?