Learning the Nodes of a Graph from Interactions

CASIS Workshop 2019

Jose Cadena and Goran Konjevod

LLNL-PRES-774204

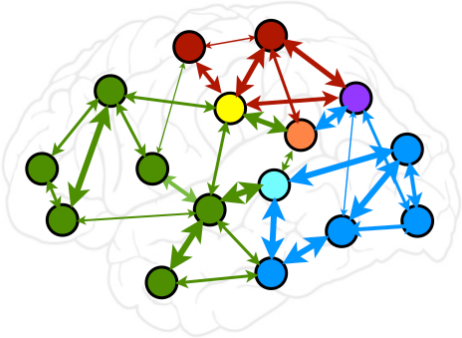
This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

Lawrence Livermore National Laboratory

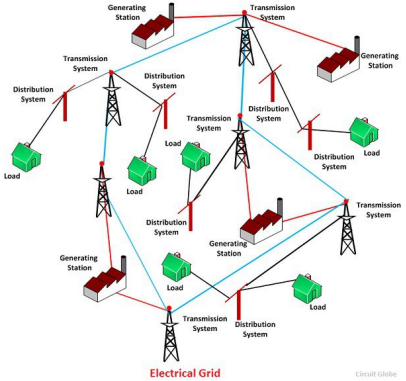
Graphs are everywhere

Cyber-security
Computer network

Social science
Collaboration network

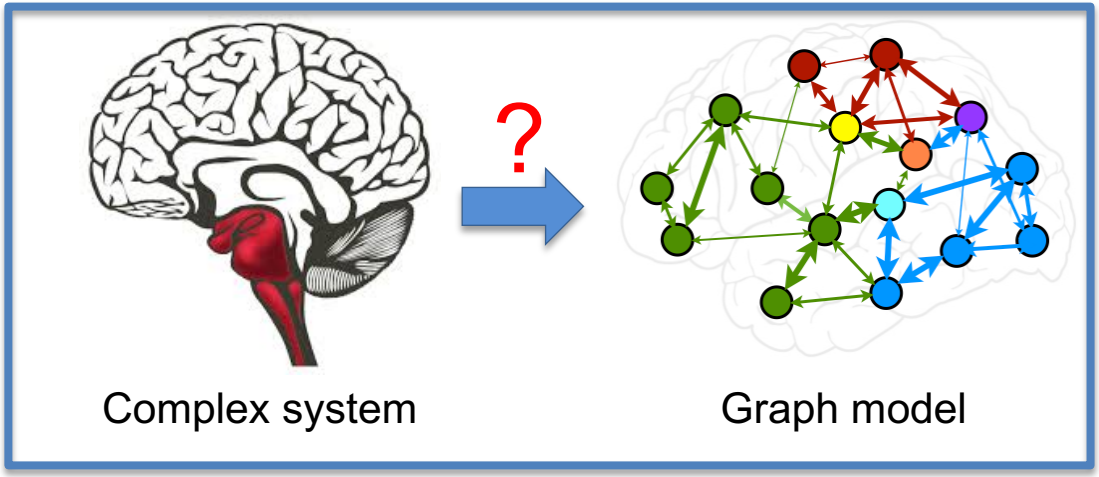


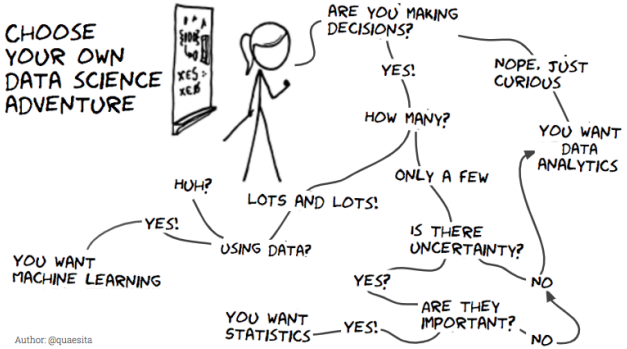
Neuroscience
Functional brain network



Energy
Power grid network

Given a graph …

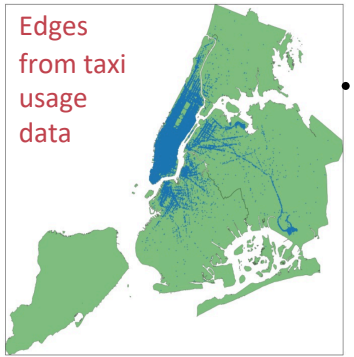


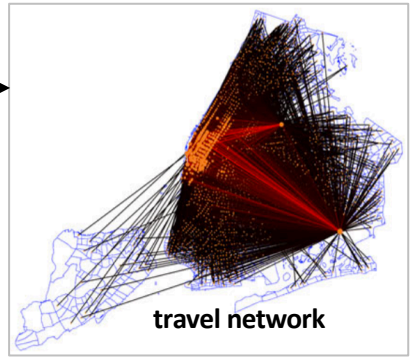


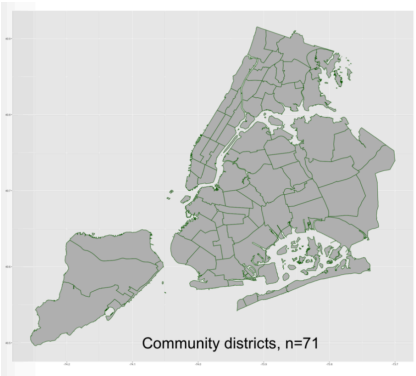
Analysis

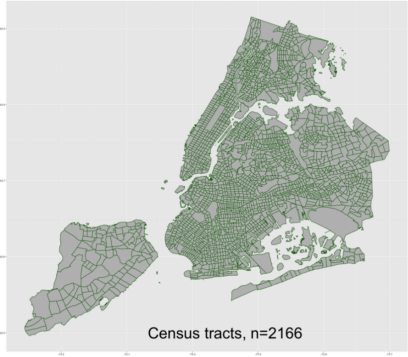
- Obvious nodes and edges (power grid, computer network)
- Obvious nodes, but not edges (social networks)
- Nodes not given (brain, mobility networks)

How to define nodes when not given?



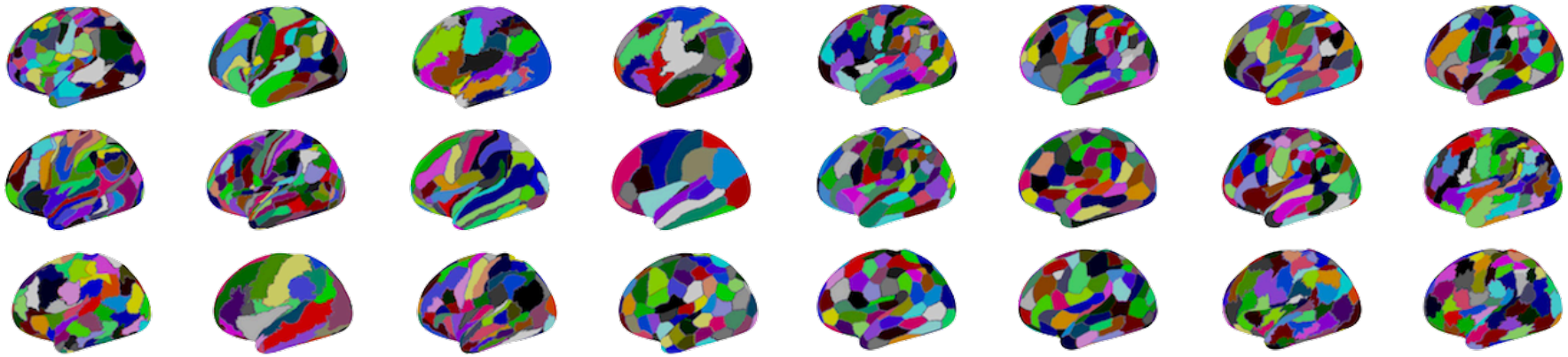






What makes a
“good” set of nodes?

Many brain parcellations have been proposed

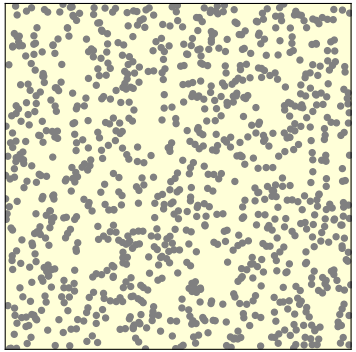


Arslan, S., Ktena, S.I., Makropoulos, A., Robinson, EC., Rueckert, D., Parisot, S., 2017, Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex, NeuroImage, doi: 10.1016/j.neuroimage.2017.04.014

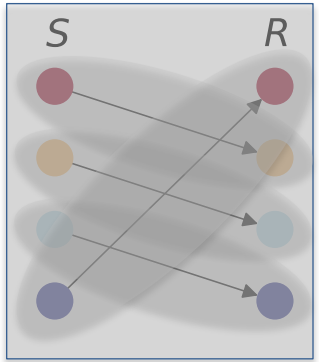
Node learning as a clustering problem

- We **observe** a set of interactions
  - $\mathcal{I} = \{(x_1, y_1), (x_2, y_2), \ldots, (x_\ell, y_\ell)\}$
  - e.g., taxi trips at geographical coordinates, blood oxygen level at voxels

- Assume interactions occur in an **unobserved** bipartite graph
  - Graph $G = (S \cup R, E)$ of senders $S = \{s_1, \ldots, s_n\}$ and receivers $R = \{r_1, \ldots, r_m\}$



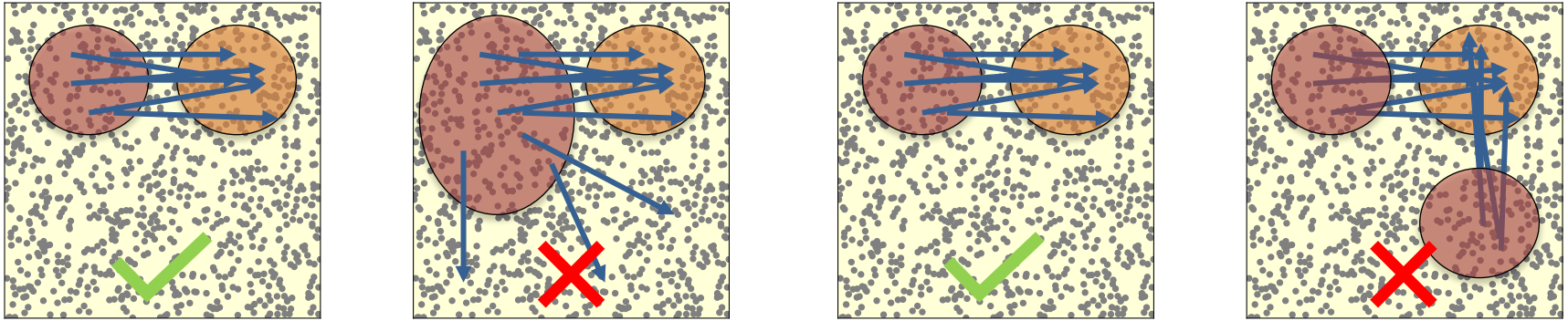
observed



hidden

Node learning as a clustering problem

- **Goal:** Given the observed interactions, we want to recover the hidden graph (i.e., senders and receivers)
- **Approach:** Cluster interactions into nodes according to two criteria:
  - **Behavioral Coherence:** Points in the same sender should all interact with the same (sparse) set of receivers
  - **Proximity Coherence:** Nodes should have contiguous shape



Behavioral

Proximity

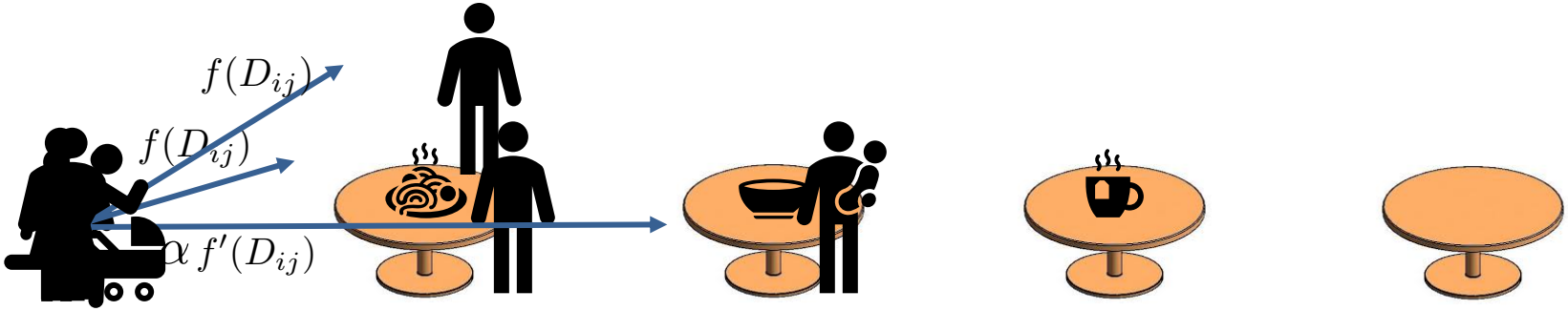
Proposed methods

- Clustering with the distance-dependent Chinese Restaurant Process (ddCRP)
  - Non-parametric generative model for cluster data
  - Constrained by distance

Probability of sitting together inversely proportional to distance

$$p(\text{PARENT}(i) = j | D, \alpha, f) \propto \begin{cases} f(D_{i,j}) & i \neq j \\ \alpha & i = j \end{cases}$$

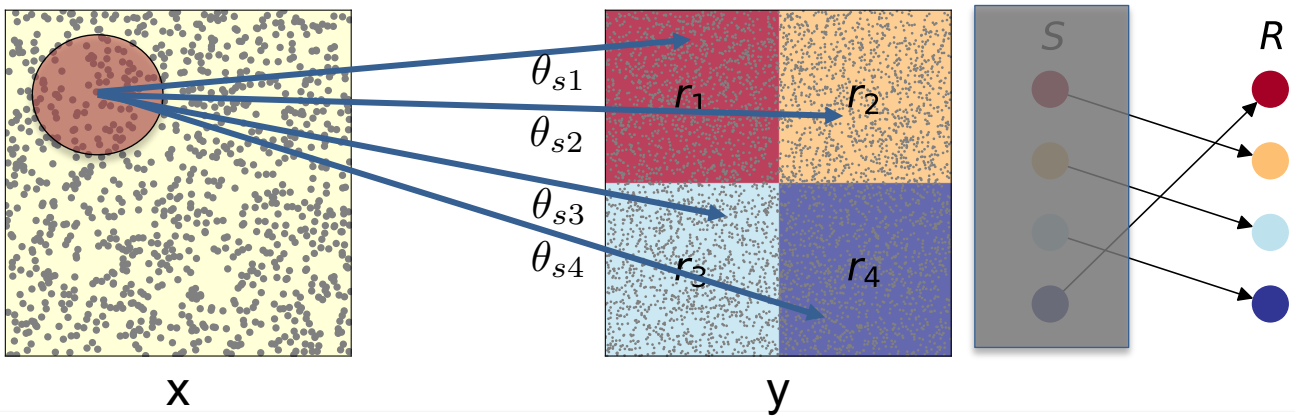
"dispersion" parameter controls probability of new table



Proposed methods

- Behavioral coherence
  - Assume we knew the receivers $R = \{r_1, \ldots, r_m\}$
  - A point x in sender s samples a receiver from a Categorical distribution

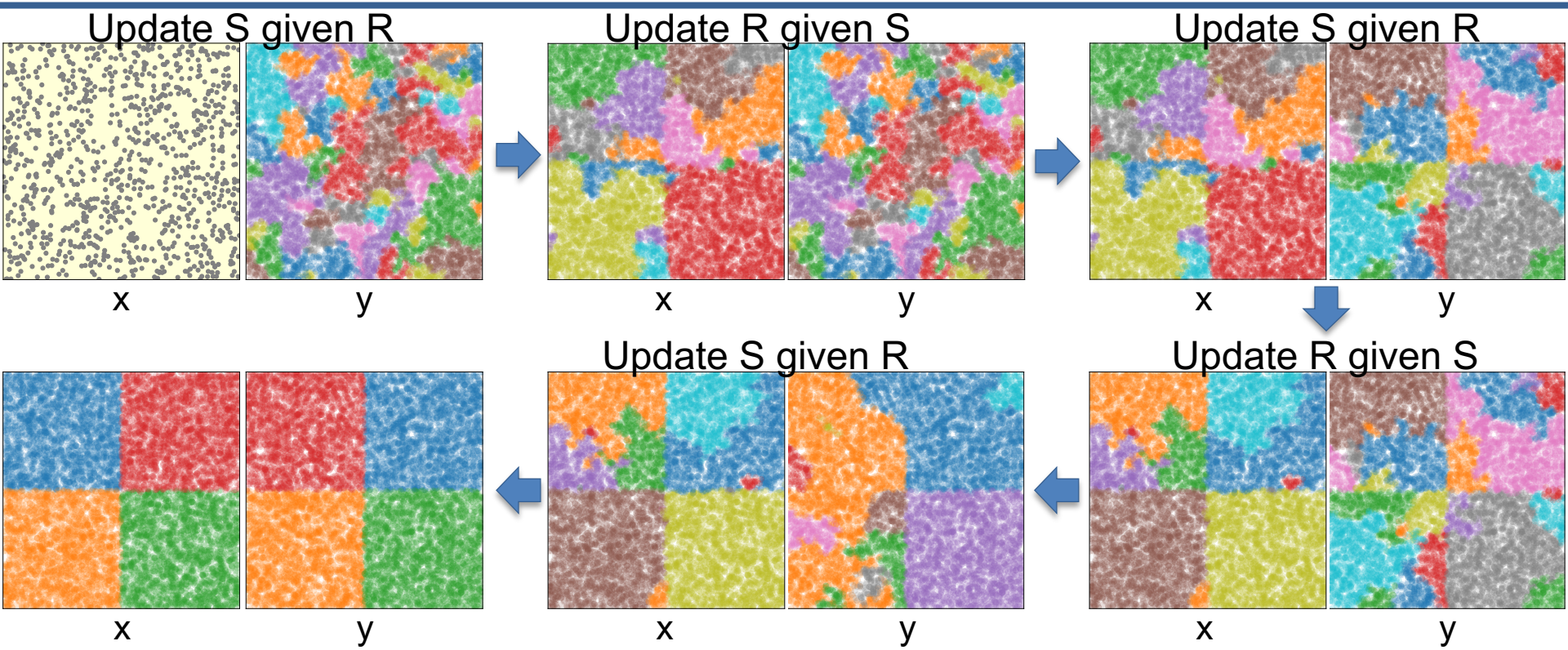
$$p(y \in r_i | \theta_s) = \theta_{si} \quad \forall (x, y), x \in s$$



Proposed methods

- If we had a set of receivers, we could cluster x's into senders using a ddCRP
  - Customers -> x's
  - Tables -> sender nodes
  - Dish -> Categorical distribution parameter
- Similarly, if we had senders, we could cluster y's into receivers
- **Main idea:** iteratively cluster x's and y's until convergence

Proposed methods

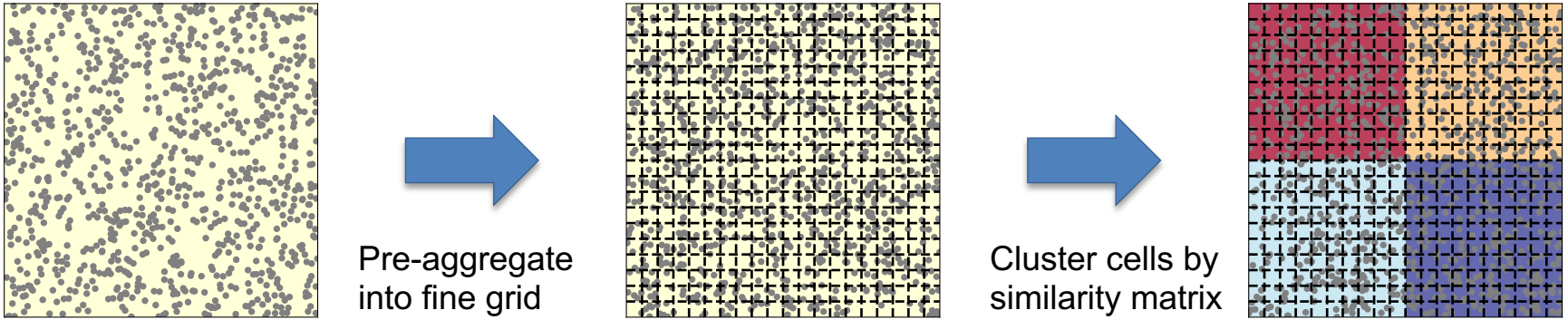


Proposed methods

- ☺ Little pre-processing needed
- ☺ Few parameters
- ☹ Inference requires sampling (slow)

Proposed methods

- Spectral clustering as a fast heuristic



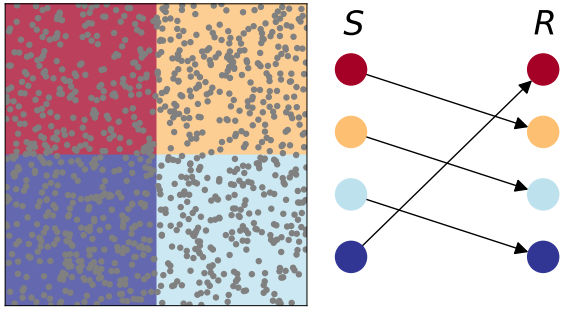
$$M = \beta(1 - D) + (1 - \beta)(1 - B)$$

Trade-off parameter

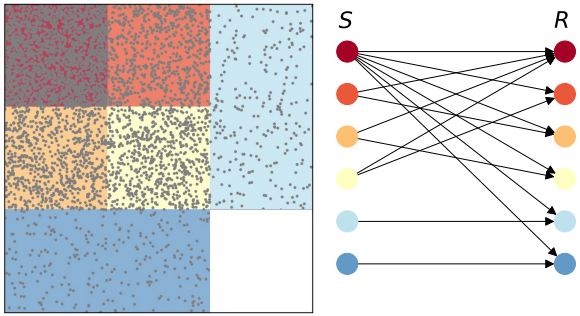
Euclidean distance

Symmetric KL divergence

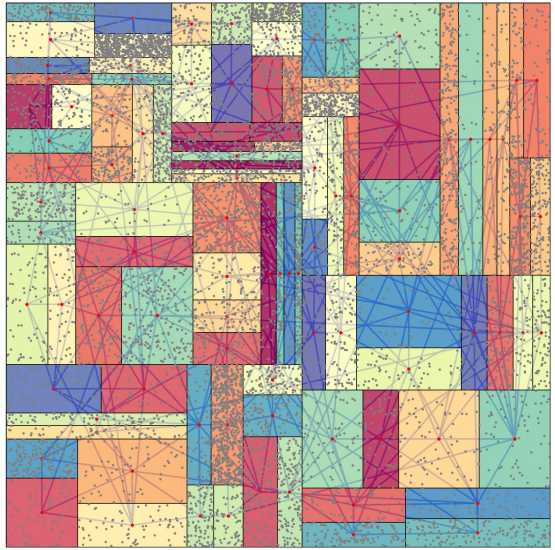
Validation on synthetic data



Uniform commute (UC)



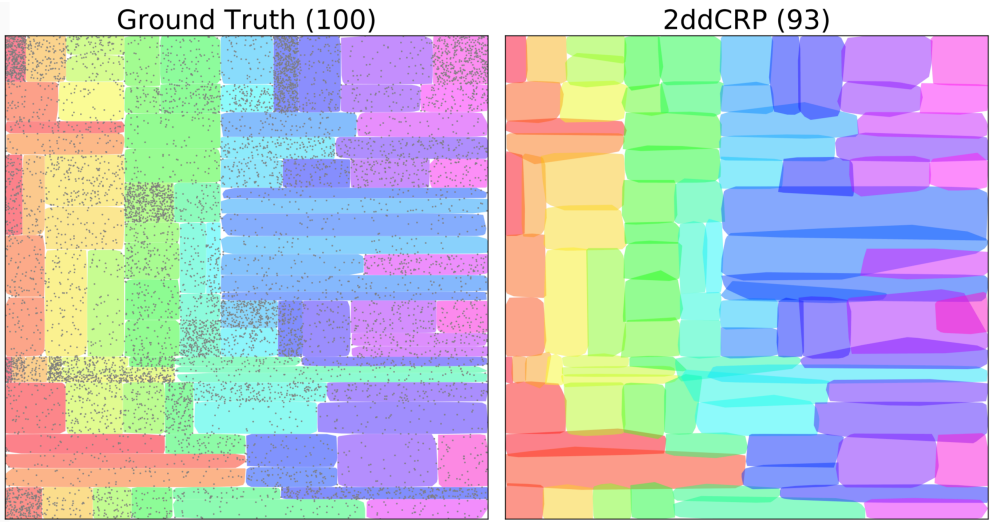
Hubs and cycles (HC)



Small world (SW)

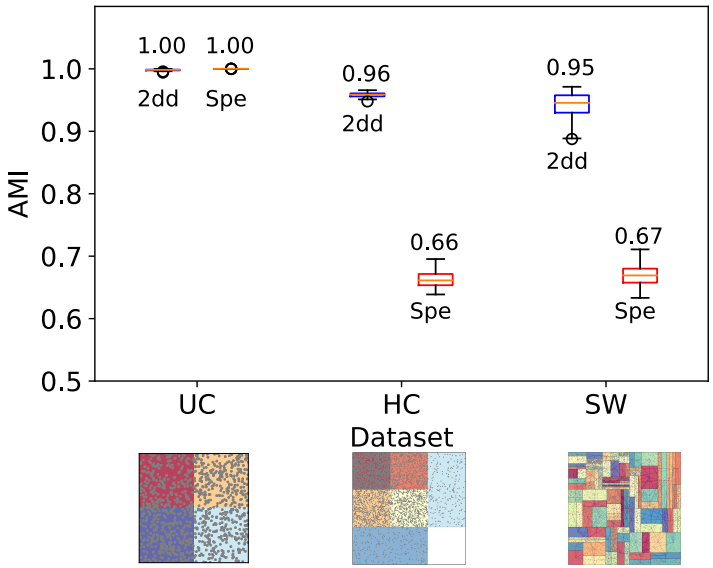
Validation on synthetic data

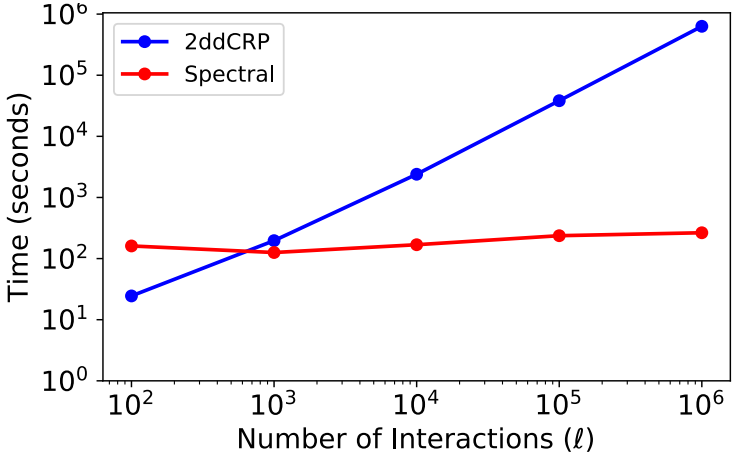
- ddCRP clustering similar to ground truth
- More accurate in regions with more observations



Validation on synthetic data

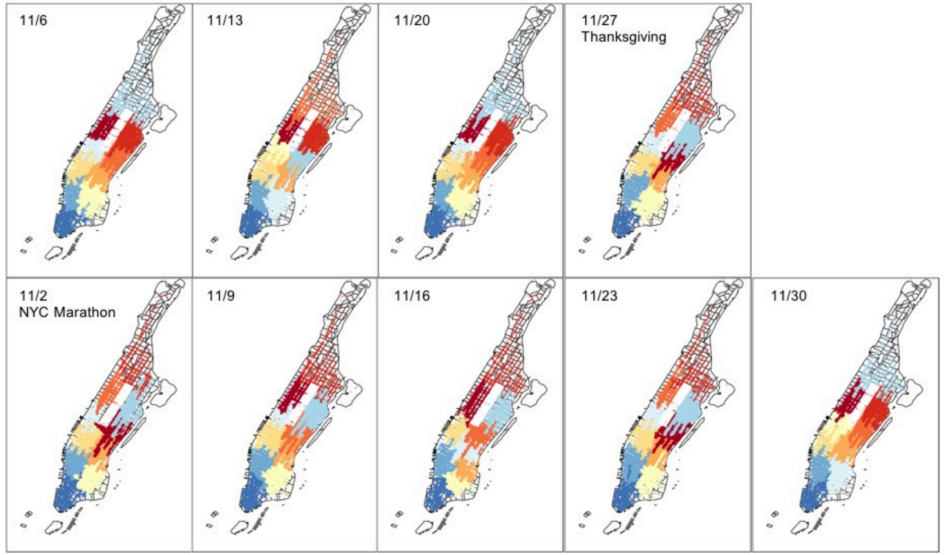
- Measured accuracy using adjusted mutual information (AMI)
  - Close to 1 if inferred clusters resemble the ground truth
- ddCRP much more accurate than spectral clustering, but much slower





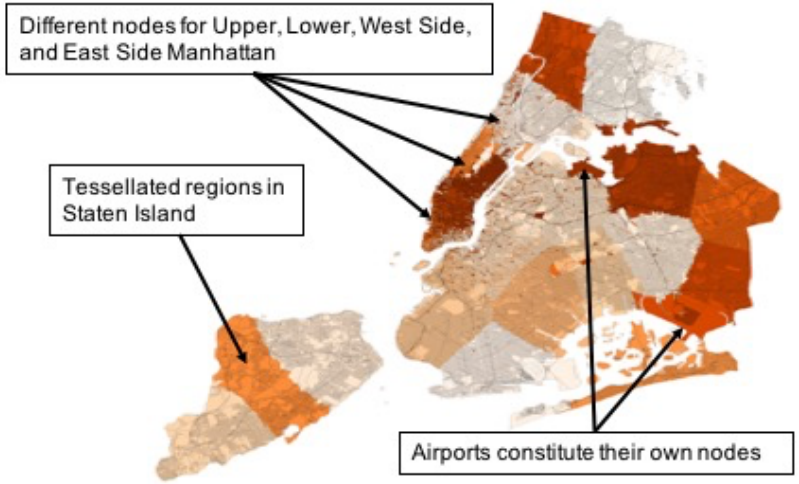
New York City Taxi Data

- ddCRP applied to Manhattan taxi trips
- Manhattan roughly divided into Upper, Lower, West Side, and East Side Manhattan.
- Nodes are consistent over different days of the month



New York City Taxi Data

- Spectral clustering applied to the entire city



Conclusions

- **Goal:** Learn nodes of a graph from “atomic” interactions

ddCRP	Spectral Clustering
• No pre-processing required • Few parameters • Infers number of nodes • Probabilistic interpretation • Slow • “Nodes” are defined as a cluster of points. How to label new observations?	• Scales to millions of interactions • Requires initial partition • Needs to specify number of nodes

Lawrence Livermore
National Laboratory