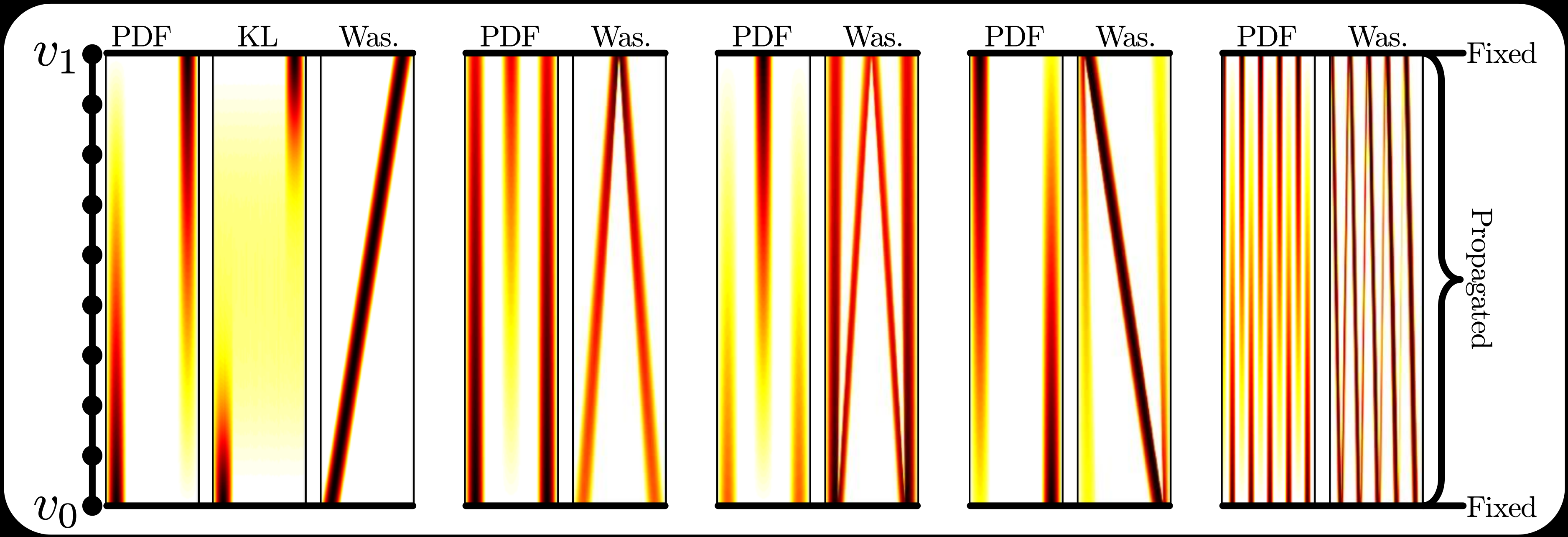



# Wasserstein Propagation for Semi-Supervised Learning

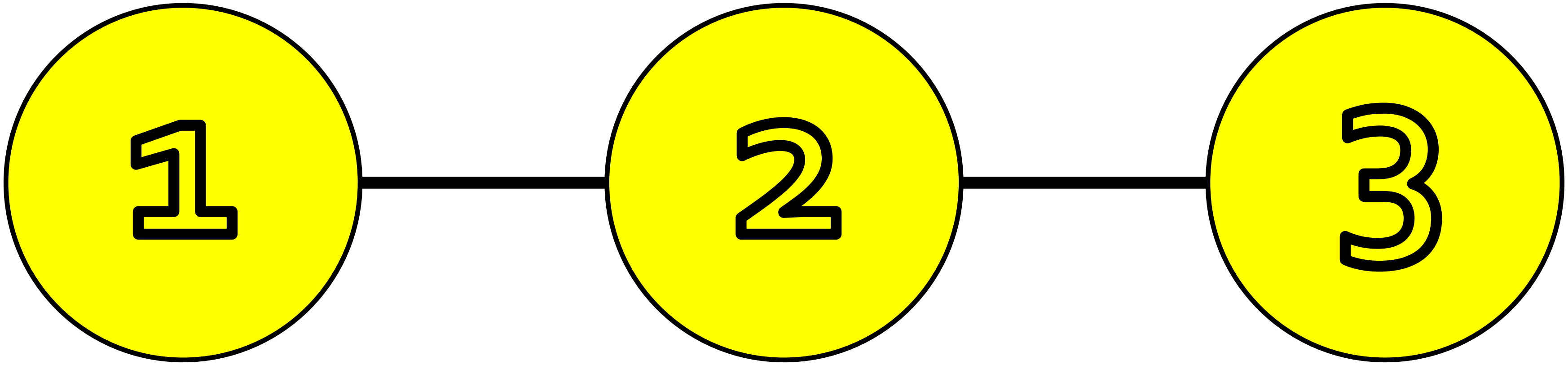
Justin Solomon,* Raif Rustamov, Leonidas Guibas
Stanford University

Adrian Butscher
Autodesk Research

# Motivation

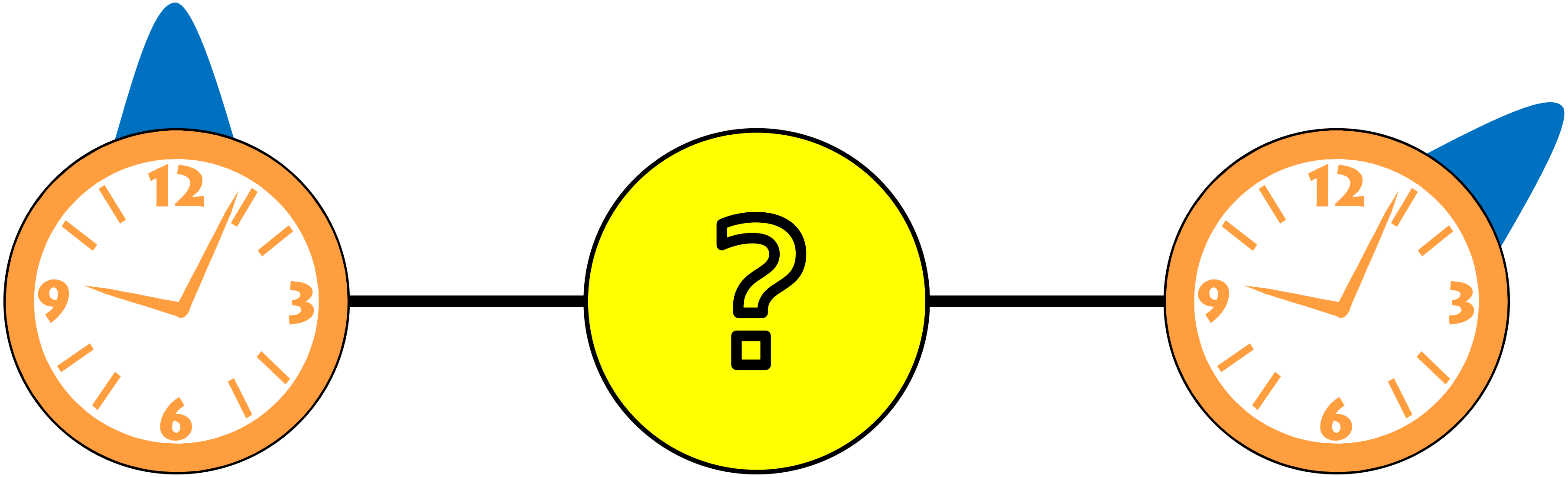

# Network of web pages

# Motivation



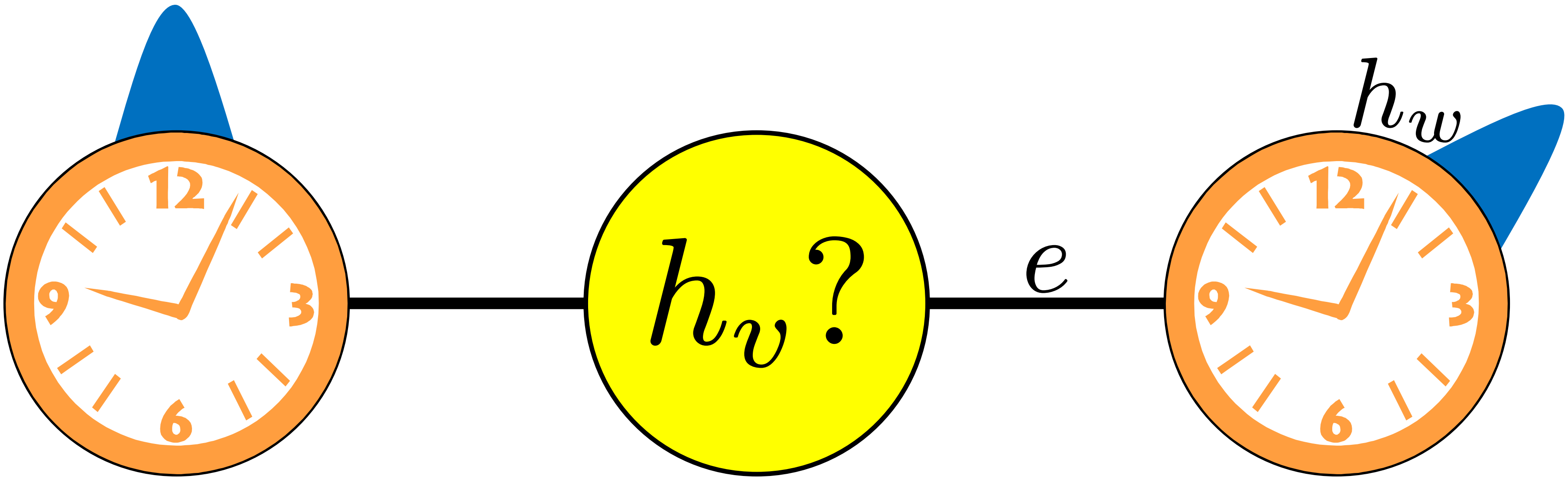

**Network of web pages**

# Histogram of Web Traffic



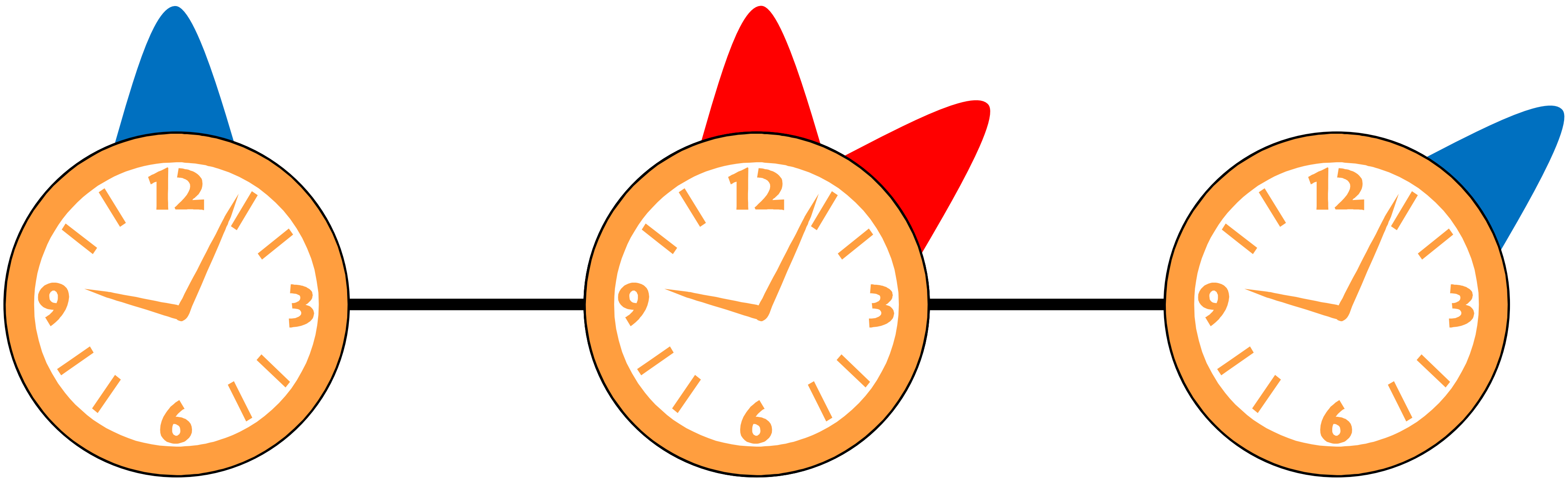

**Predict the missing histogram**

# Variational Approach



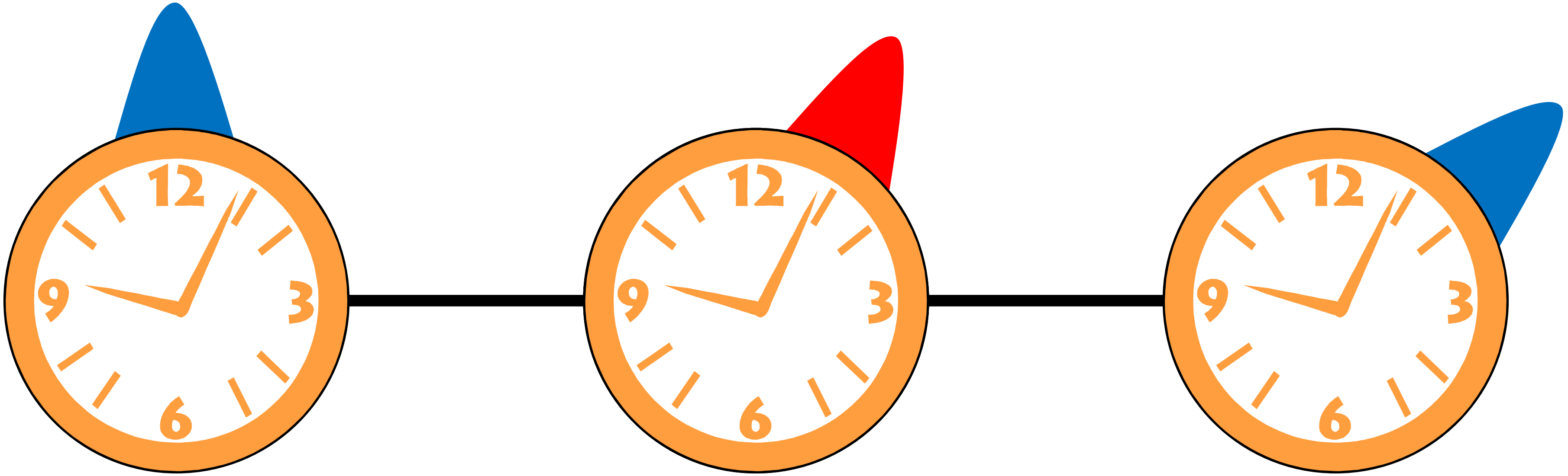

$$E \equiv \sum_{(v,w)=e} d_{\text{hist}}(h_v, h_w)$$

# KL Divergence




Bimodal result

# Desirable Output



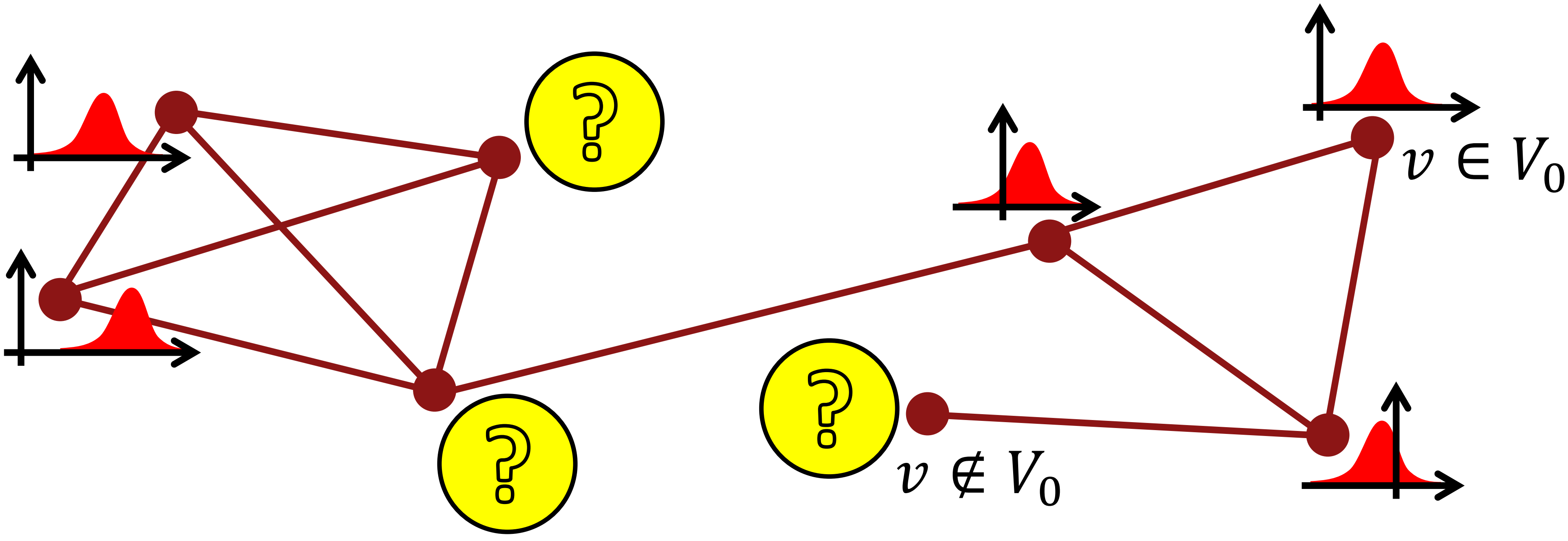

**Histograms slide along clock**

# The Punchline

**Measure of divergence matters.**

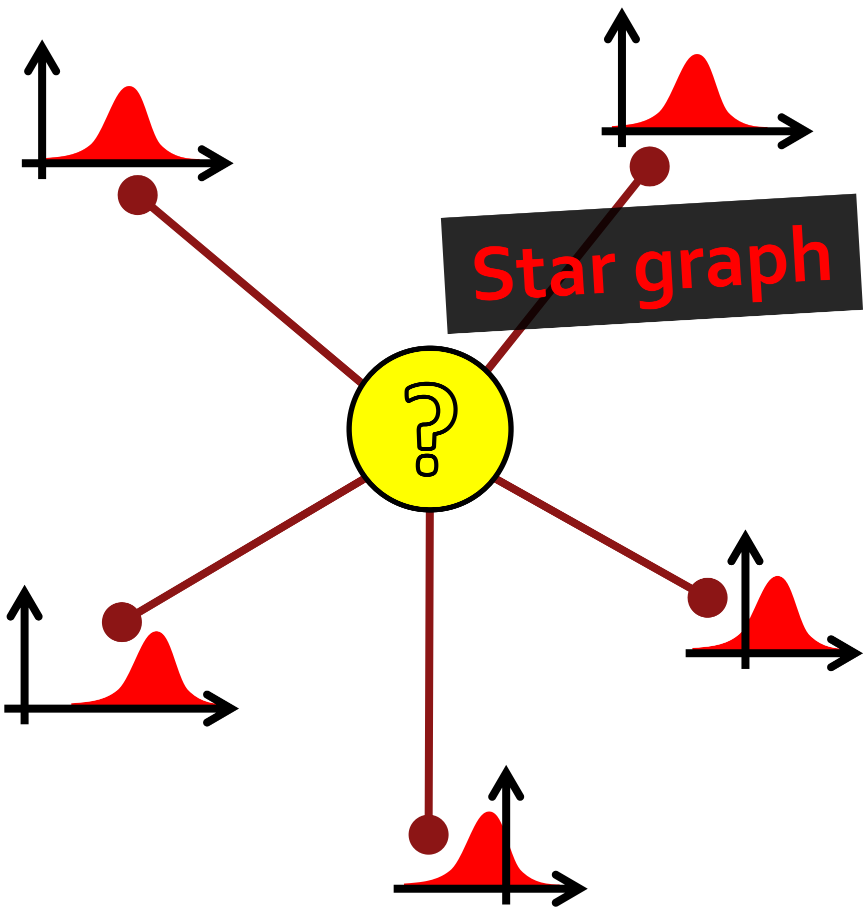
*Earth mover's/Wasserstein/Mallows Distance!*

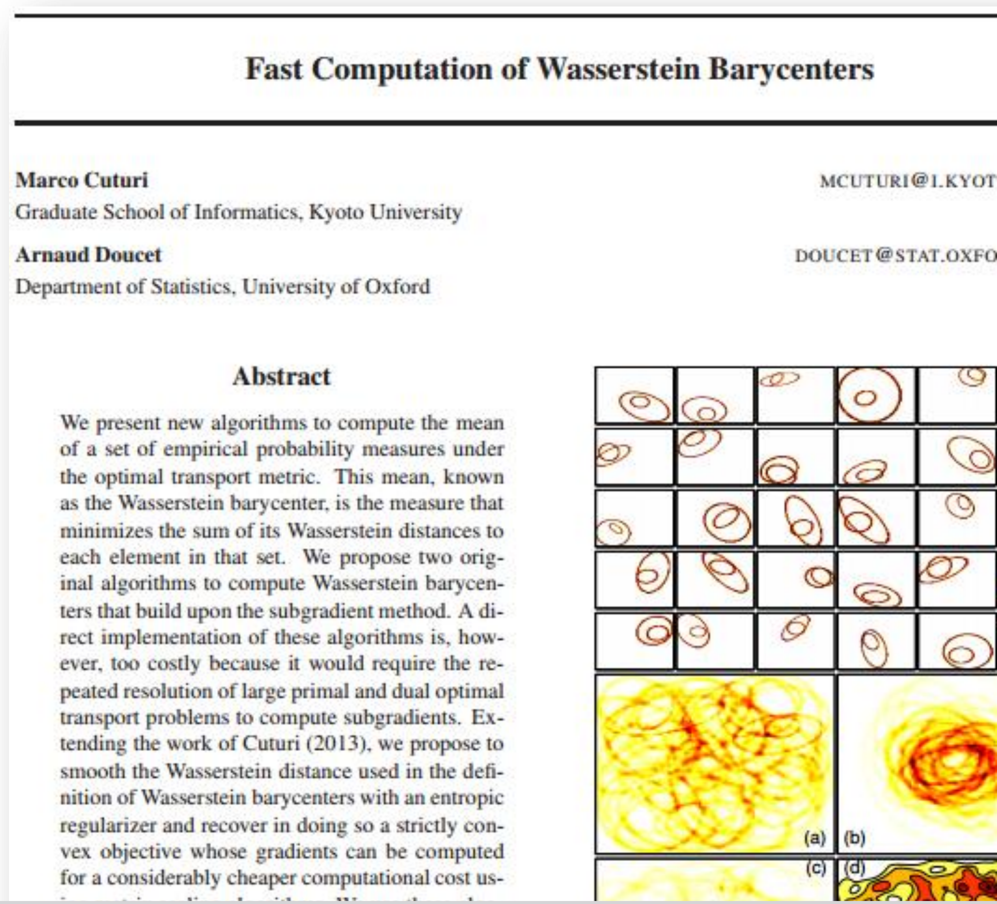
# General Problem




**Propagation of distributional labels**

# Related Problem



**Fast Computation of Wasserstein Barycenters**


**Marco Cuturi** MCUTURI@I.KYOT
Graduate School of Informatics, Kyoto University
**Arnaud Doucet** DOUCET@STAT.OXFO
Department of Statistics, University of Oxford


**Abstract**

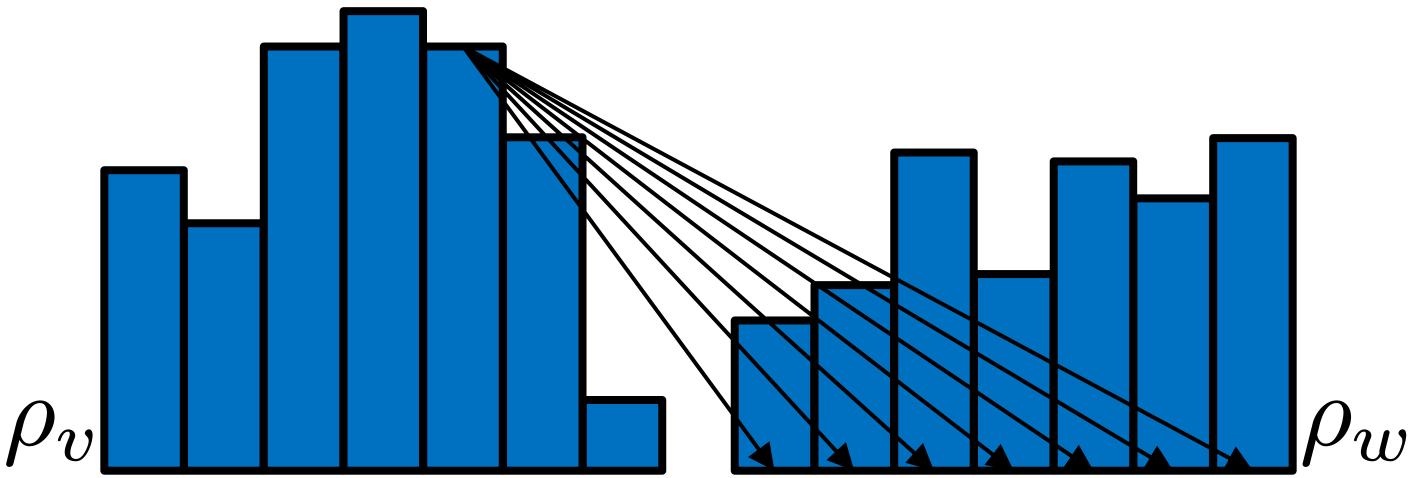

We present new algorithms to compute the mean of a set of empirical probability measures under the optimal transport metric. This mean, known as the Wasserstein barycenter, is the measure that minimizes the sum of its Wasserstein distances to each element in that set. We propose two original algorithms to compute Wasserstein barycenters that build upon the subgradient method. A direct implementation of these algorithms is, however, too costly because it would require the repeated resolution of large primal and dual optimal transport problems to compute subgradients. Extending the work of Cuturi (2013), we propose to smooth the Wasserstein distance used in the definition of Wasserstein barycenters with an entropic regularizer and recover in doing so a strictly convex objective whose gradients can be computed for a considerably cheaper computational cost us-




**Barycenter of set of distributions**

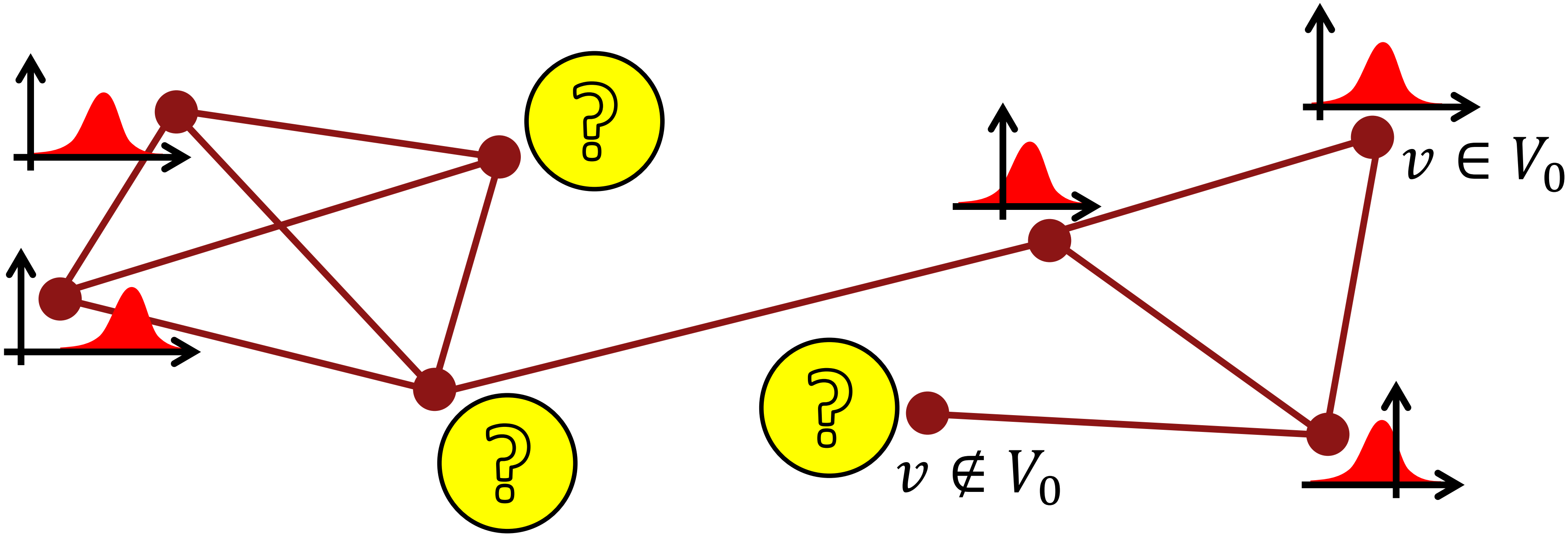
# Quadratic Wasserstein Distance

$$\mathcal{W}_2(\rho_v, \rho_w) := \inf_{\pi \in \Pi(\rho_v, \rho_w)} \left( \iint_{\mathbb{R}^2} |x-y|^2 \, d\pi(x,y) \right)^{1/2}$$



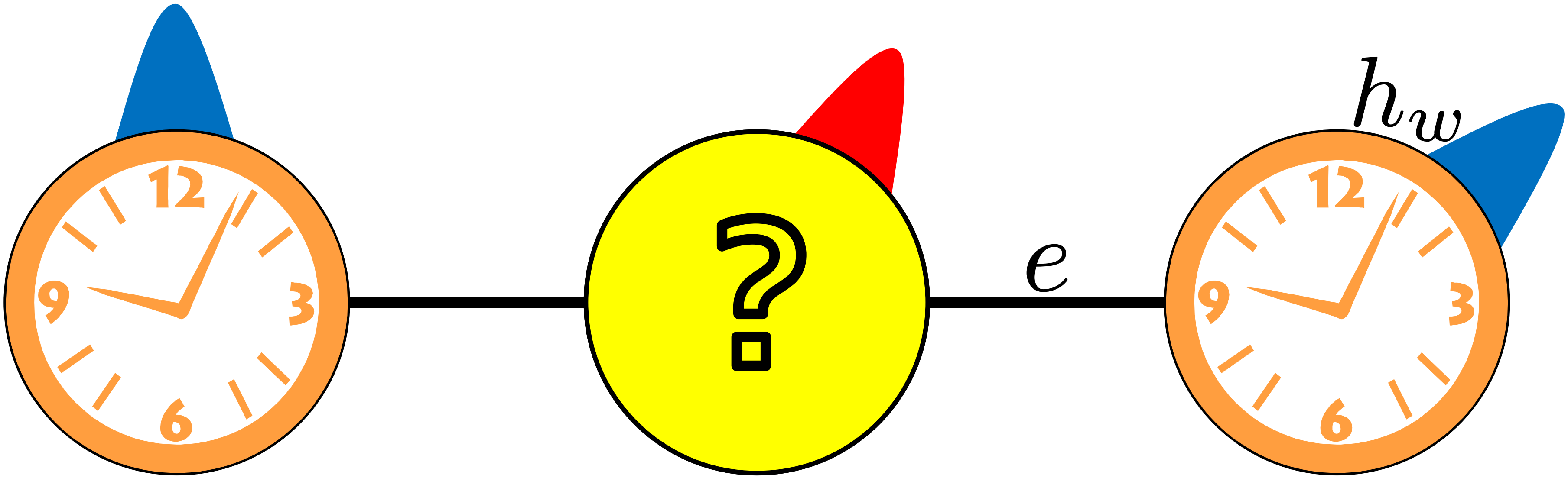

**Minimum transportation cost**

# Dirichlet Energy




$$\mathcal{E}_D[\rho] := \sum_{(v,w)\in E} \mathcal{W}_2^2(\rho_v, \rho_w)$$

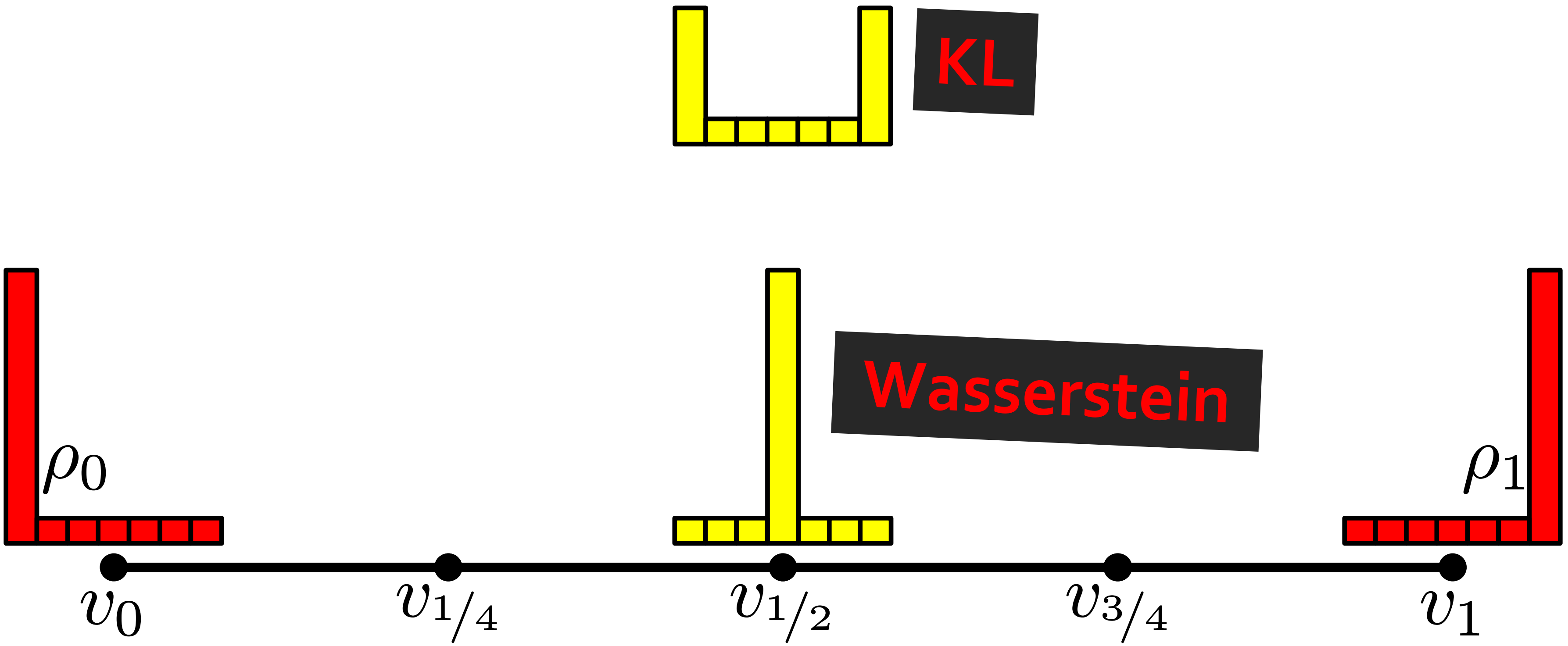
# Wasserstein Propagation



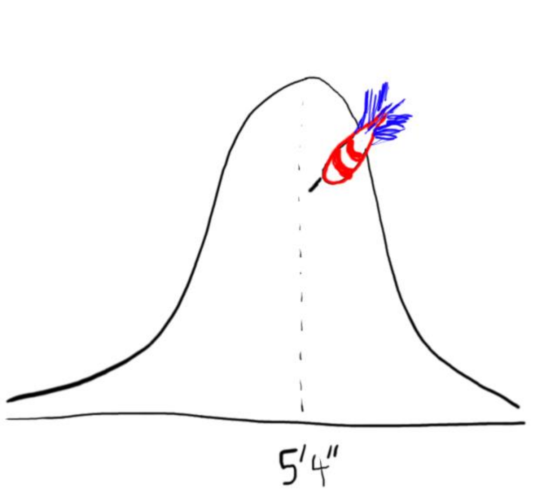

**Wasserstein Propagation**

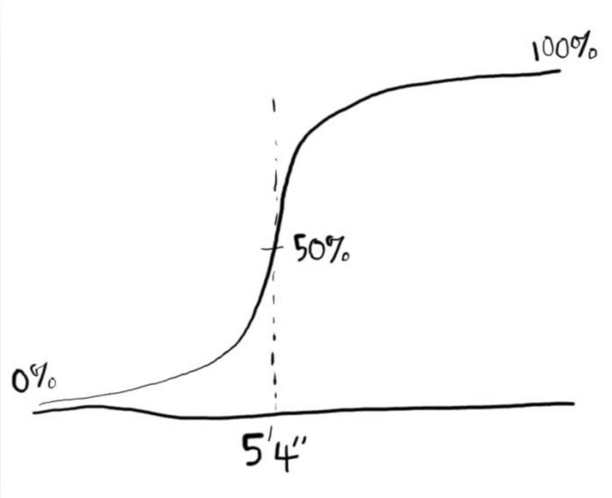
Minimize $\mathcal{E}_D[\rho]$ in the space of distribution-valued maps with prescribed distributions at all $v \in V_0$.

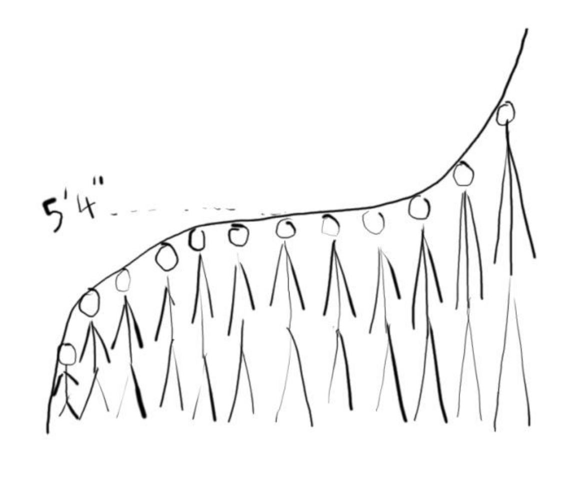
# Comparison



# Efficient Technique on a Line




**PDF** — Large linear program

**[CDF]** ··········▶

**$CDF^{-1}$** — Linear solve

http://realgl.blogspot.com/2013/01/pdf-cdf-inv-cdf.html

# Pipeline for Prob($\mathbb{R}$)

1. Transform boundary **PDFs into CDF$^{-1}$'s**
2. Perform **Dirichlet label propagation**

$$\Delta g_s = 0 \text{ with } g_s\big|_{V_0} \text{ fixed}$$

3. Transform **CDF$^{-1}$'s back to PDFs**.

# Pipeline for Prob($\mathbb{R}$)

1. Transform boundary **PDFs into CDF$^{-1}$'s**

2. Perform **Dirichlet label propagation**

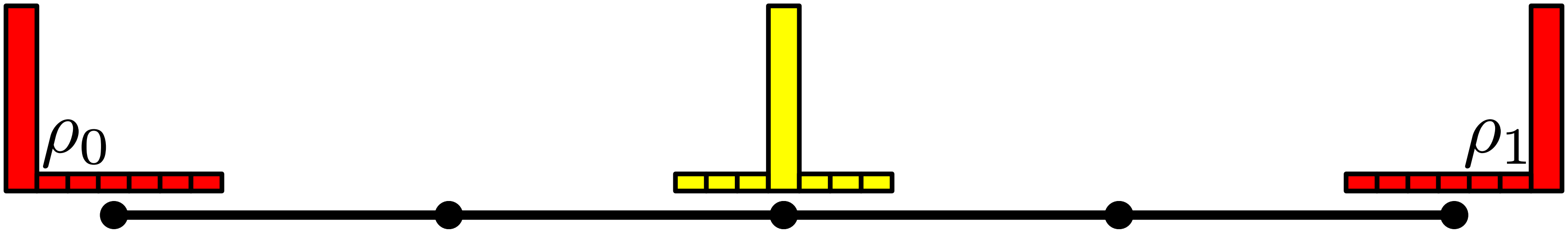
$$\Delta g_s = 0 \text{ with } g_s\big|_{V_0} \text{ fixed}$$

3. Transform **CDF$^{-1}$'s back to PDFs**.

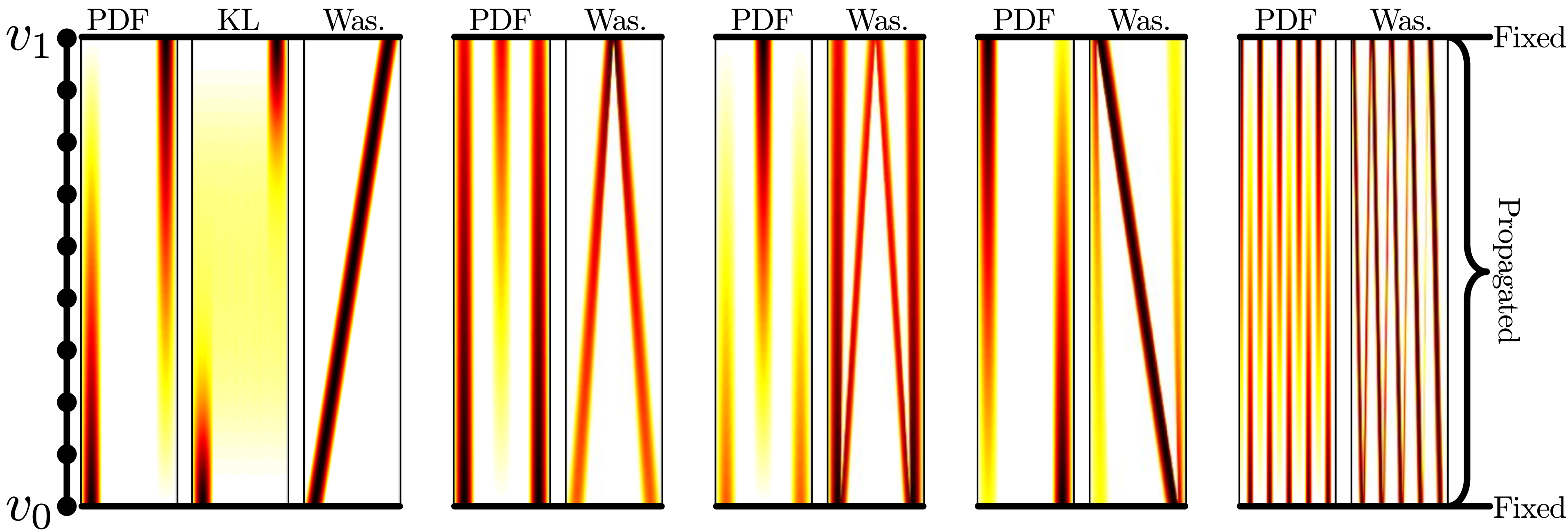
**Proposition 2: This works.**

# Theoretical Properties for Prob($\mathbb{R}$)

- **Means and variances are bounded** by those on the boundary.
- Boundary distributions are $\delta$'s $\Rightarrow$ propagated distributions are $\delta$'s.
  - Underlying map from **Dirichlet label propagation**

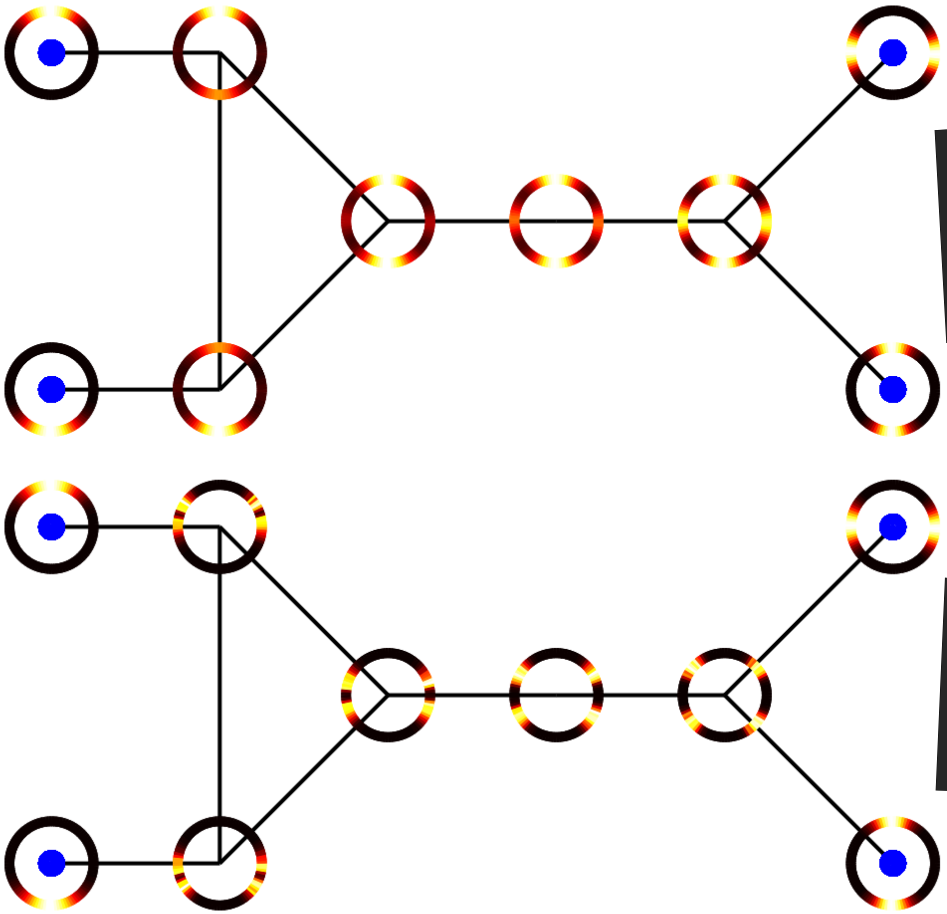


# Synthetic Experiments




# Propagation along a line

# Synthetic Experiments

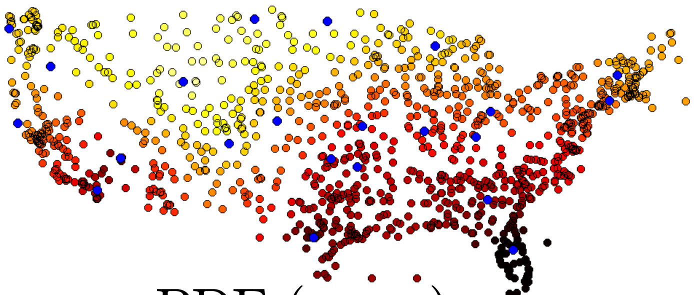


PDF propagation

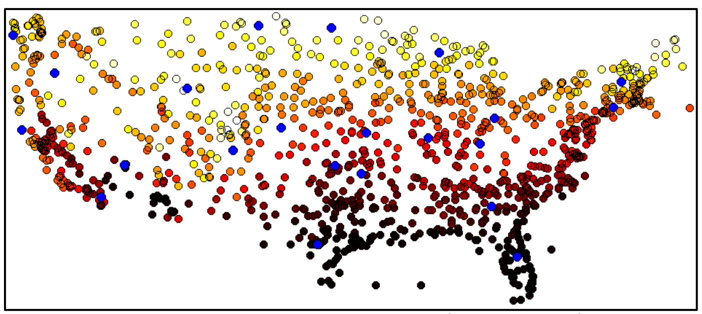
Wasserstein propagation

**Propagation of circular histograms**

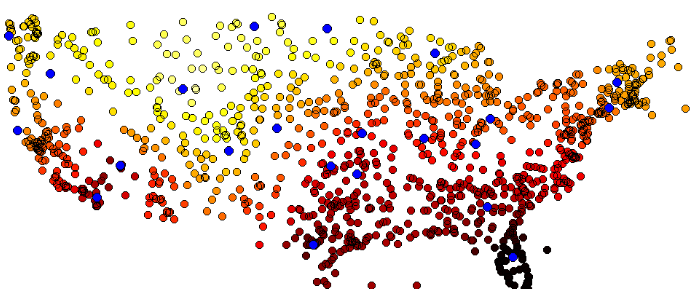
# Histograms of Temperatures



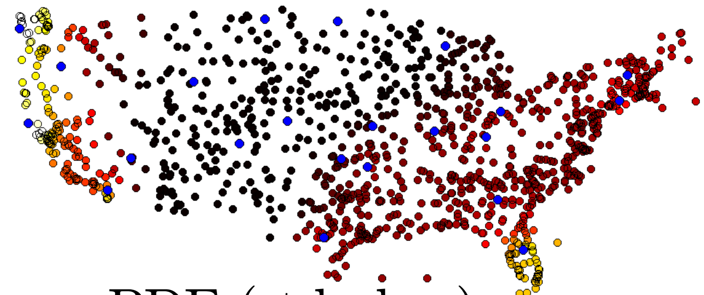
PDF (mean)



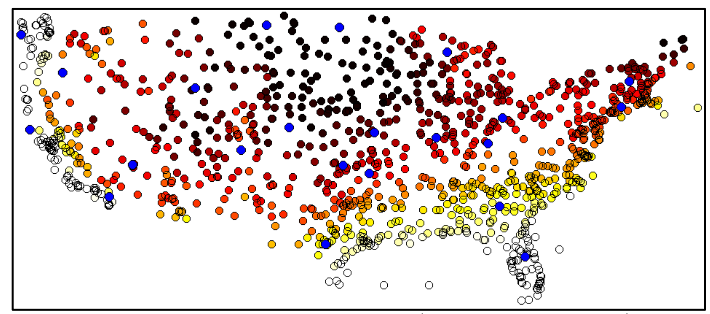
Ground truth (mean)



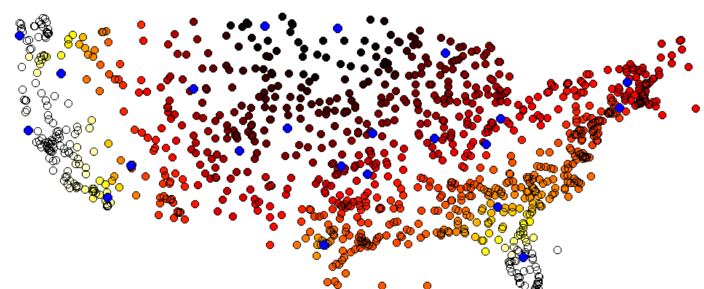
Wasserstein (mean)



PDF (std. dev.)

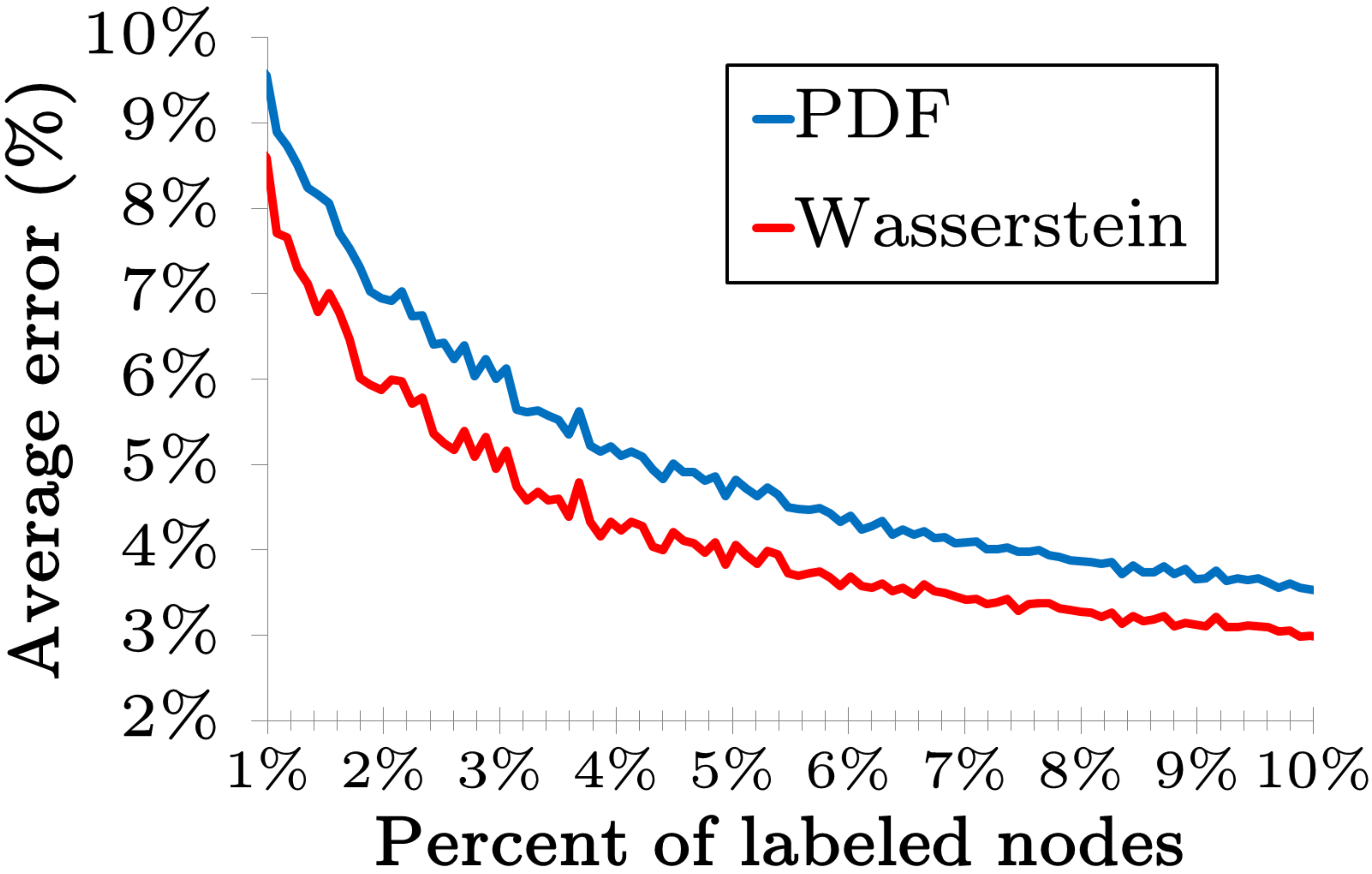


Ground truth (std. dev.)

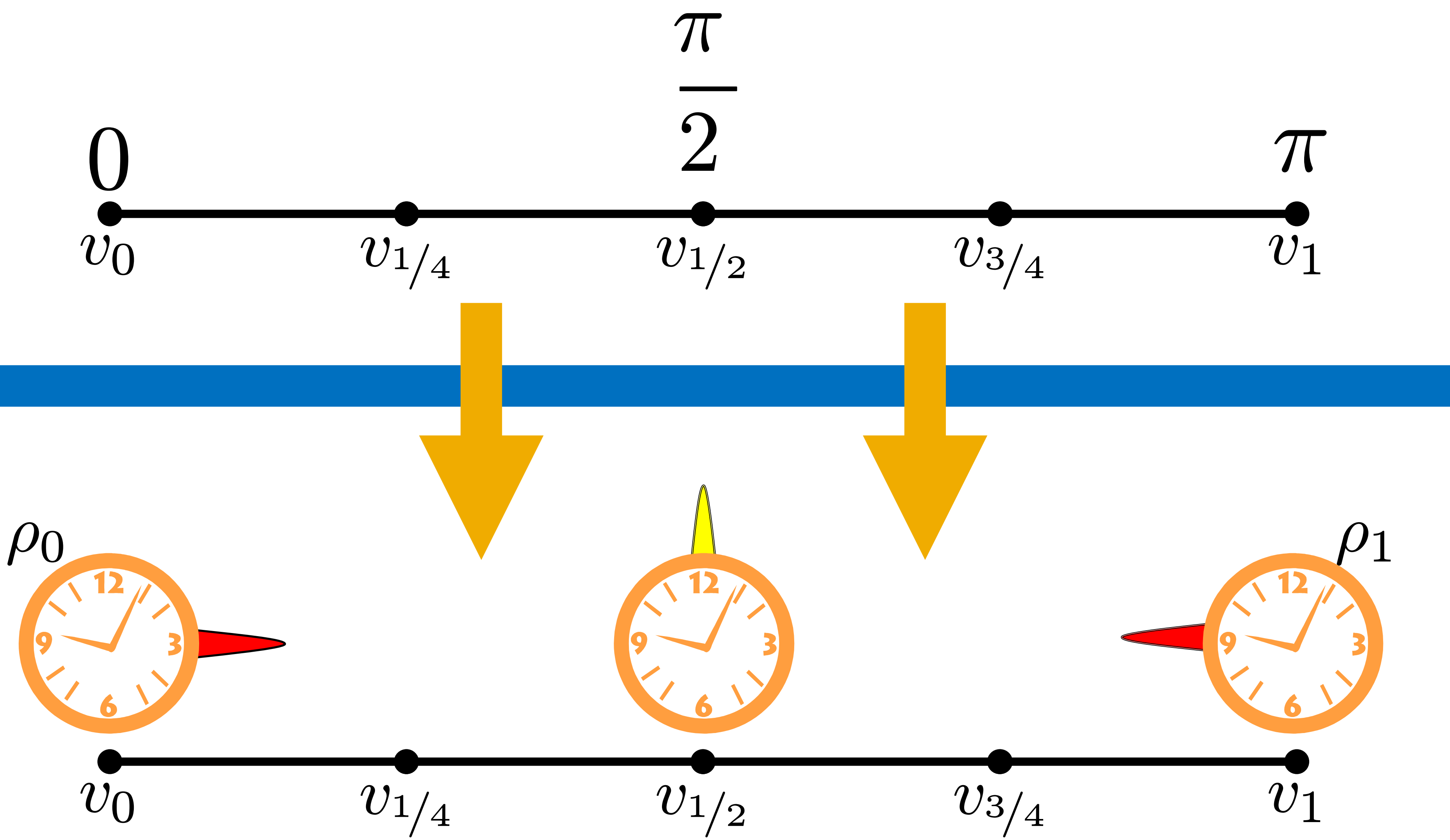


Wasserstein (std. dev.)

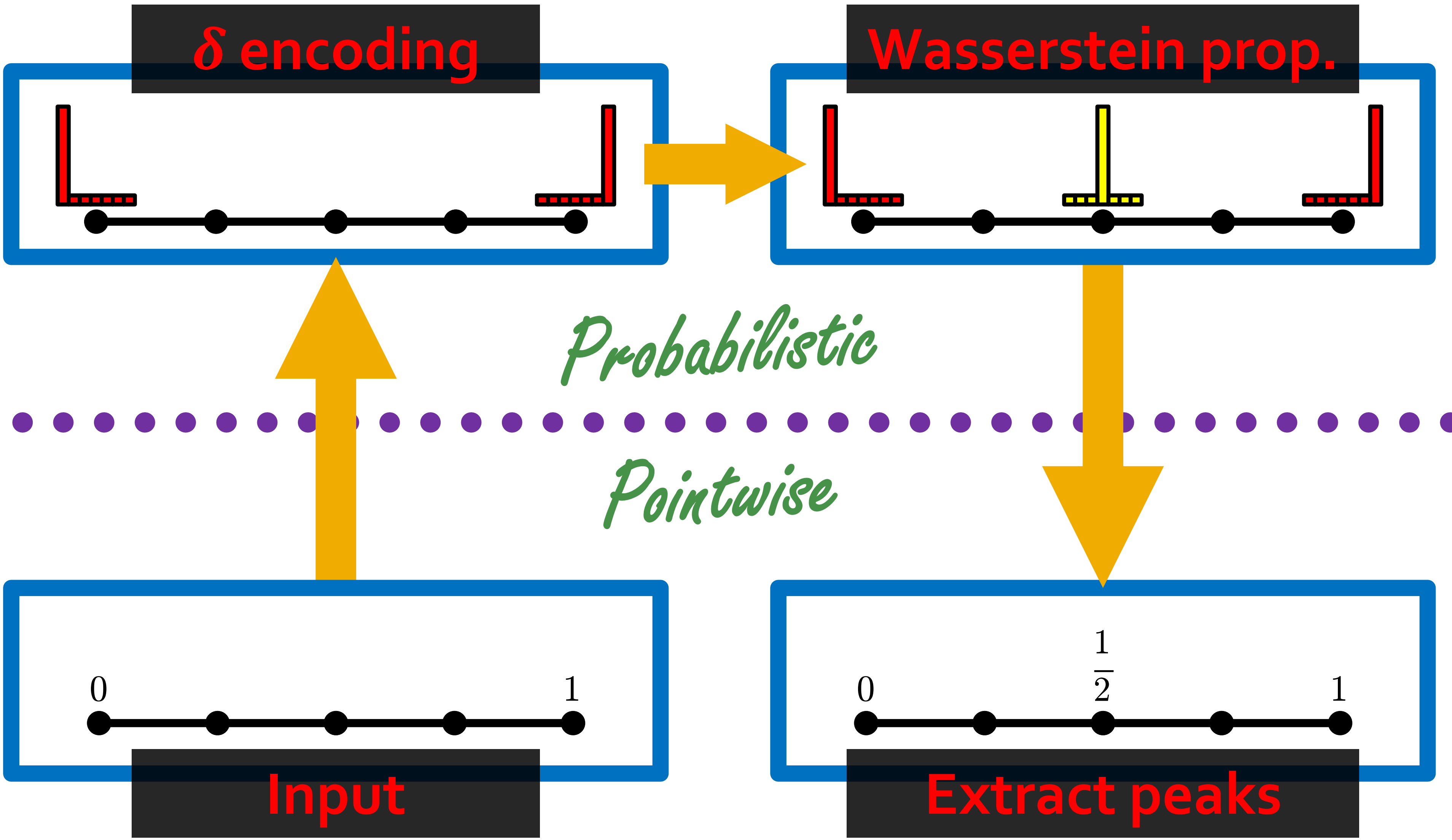
# Histograms of Temperatures



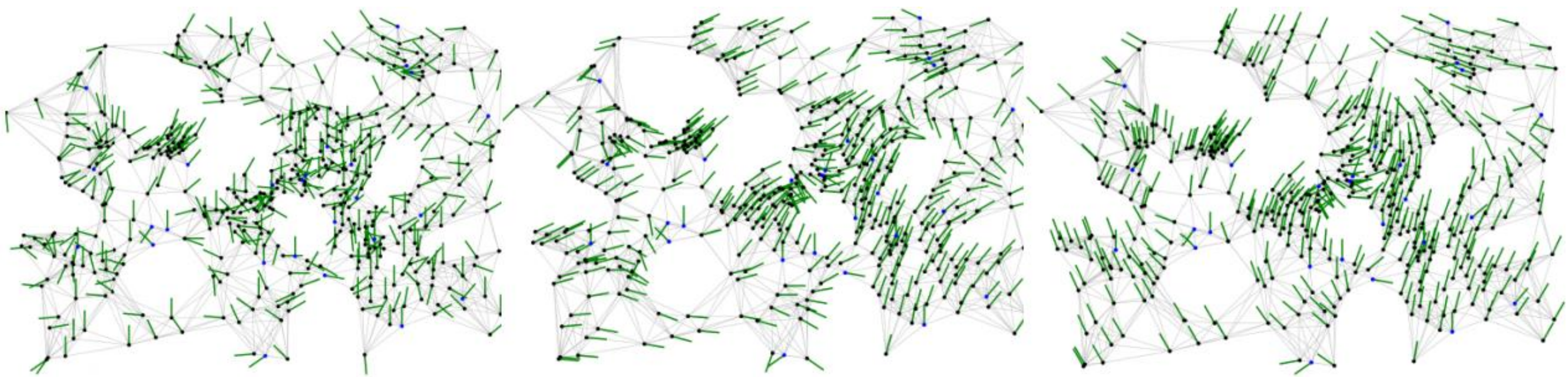
# Manifold-Valued Learning



# Manifold-Valued Learning



# Predicting Wind Directions



Ground truth PDF (19%) Wasserstein (15%)

**Labels are points on unit circle**

# Summary

- Extension of semi-supervised method to distributional labels

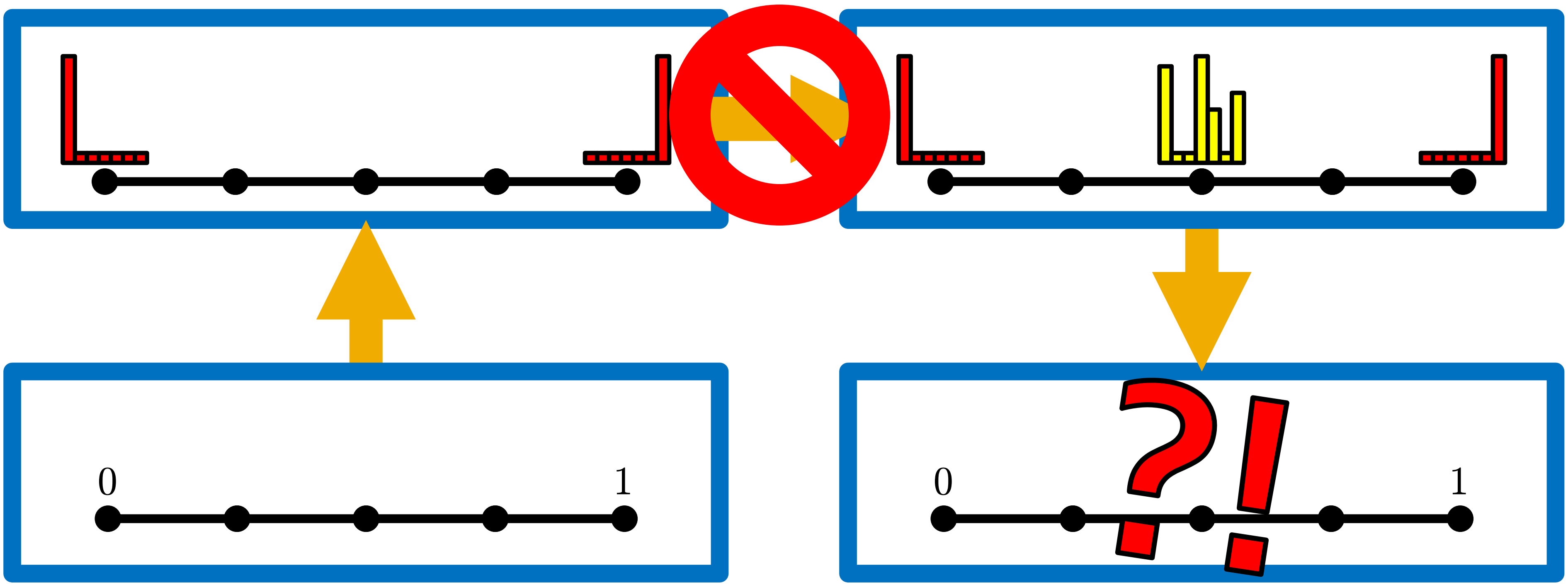
Linear program

- Fast computation and theoretical characterization for $\mathrm{Prob}(\mathbb{R})$

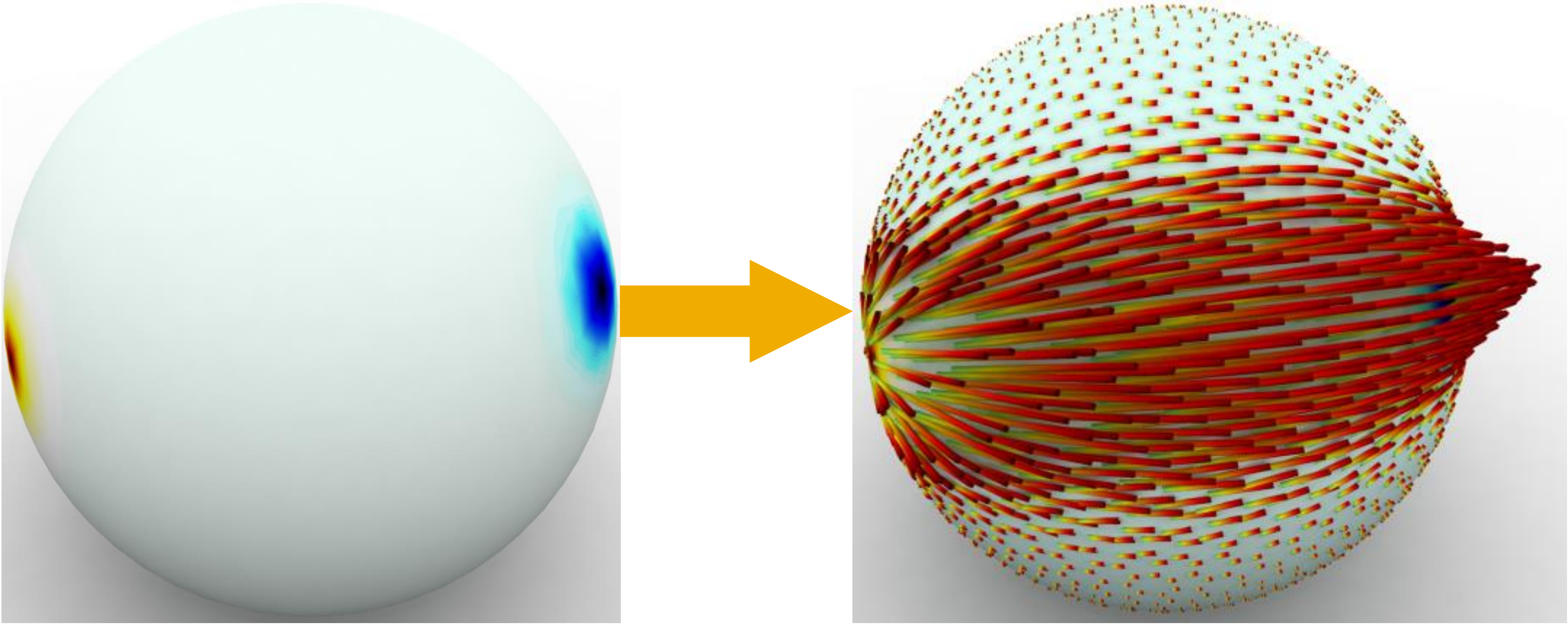
Linear solve

# What's Next: Theory

**Do peaked distributions propagate to peaked distributions?**



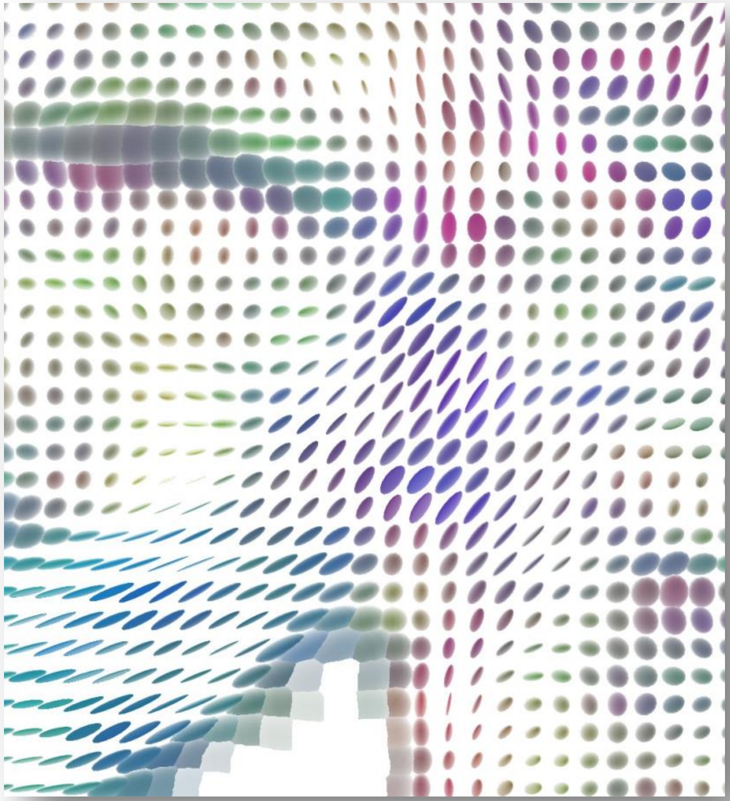
# What's Next:  Computation

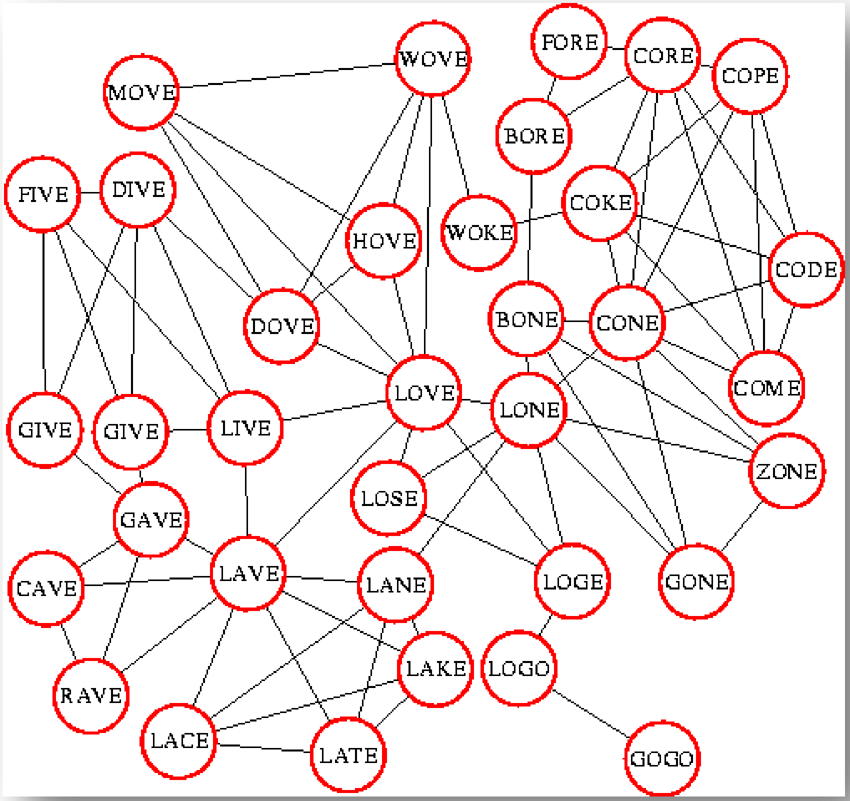


Optimization for optimal transportation

*SIGGRAPH 2014!*

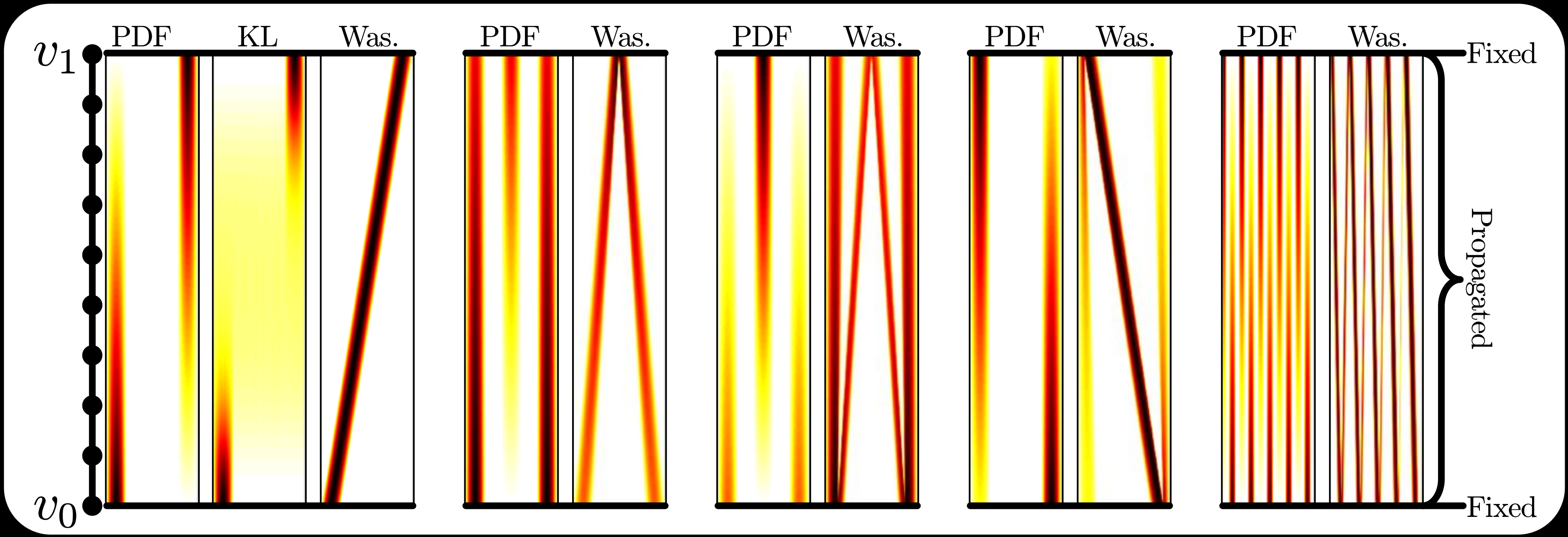
# What's Next: Applications






http://en.wikipedia.org/wiki/Diffusion_MRI • http://www.leda-tutorial.org/en/unofficial/ch03s02.html

**Learning with geometric labels**




# Wasserstein Propagation for Semi-Supervised Learning

## Questions?