

# Manifold learning for virtual screening and drug-target binding affinity predictions

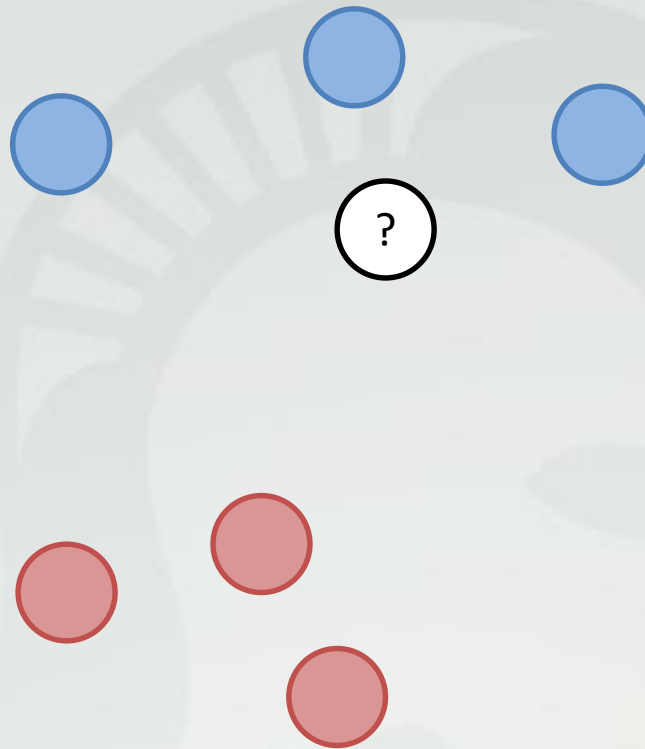
Duc Nguyen

Department of Mathematics  
Michigan State University

*In collaboration with Guo-Wei Wei*



# Data Distribution in Learning

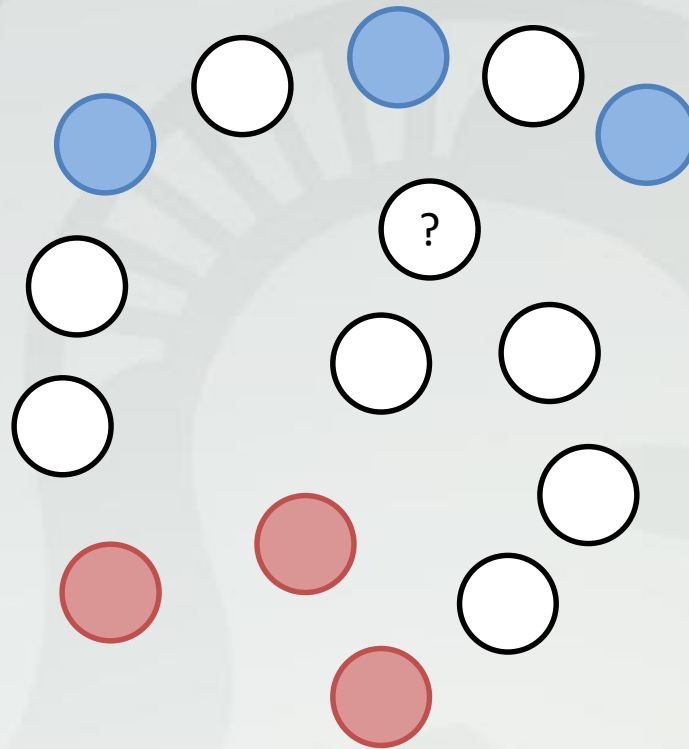


**Given:** Blue and red data points

**Task:** Predict label of

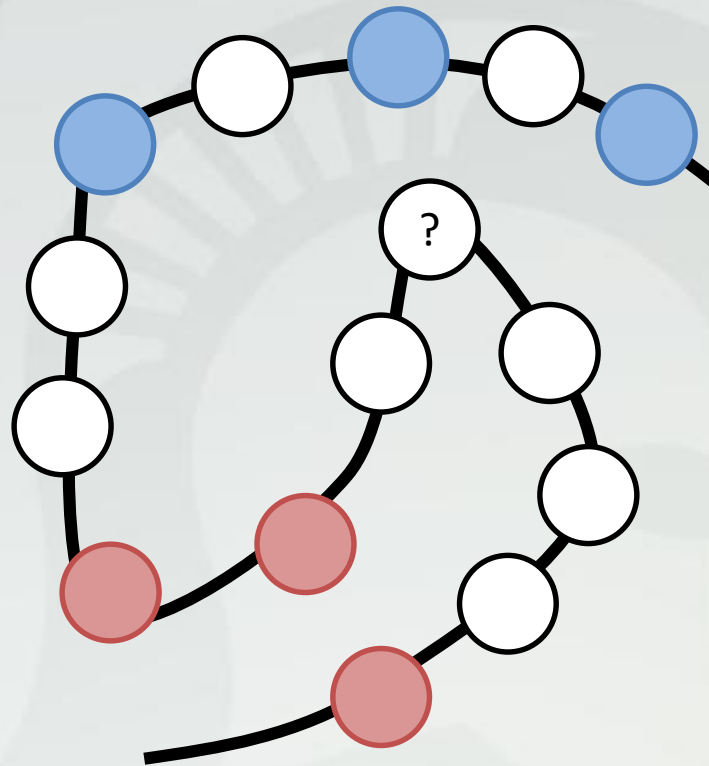


# Data Distribution in Learning



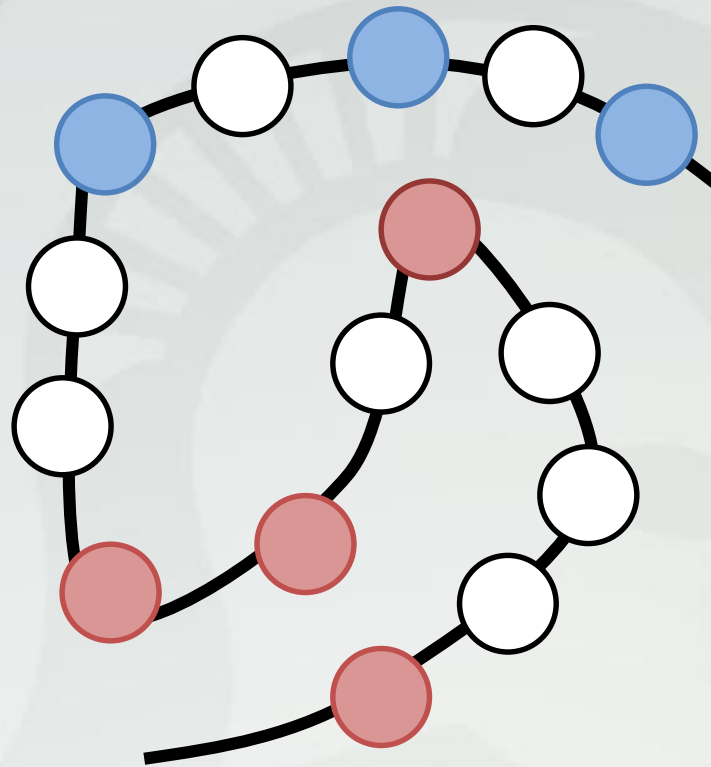
○ : Unlabeled data

# Data Distribution in Learning



Use unlabeled data to construct the manifold representation

# Data Distribution in Learning



Use unlabeled data to construct the manifold representation

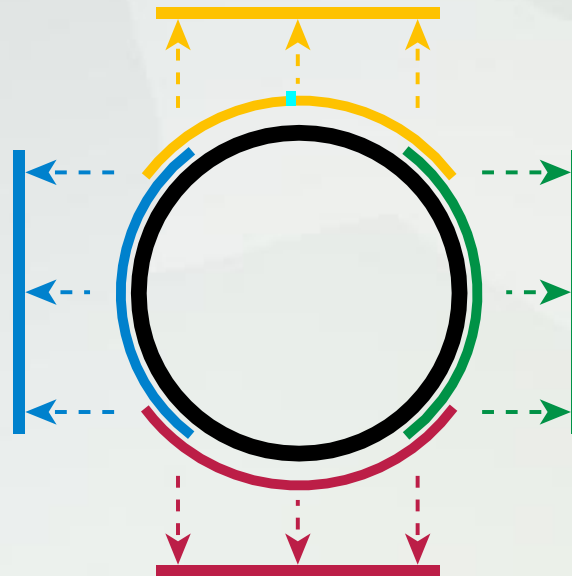
# Riemannian Manifolds

- A  $d$ -dimensional manifold

$$\mathcal{M} = \bigcup_{\alpha} U_{\alpha}$$

is a **mathematical object** in which each local patch  $U_{\alpha}$  resemble Euclidean space  $\mathbb{R}^d$

$$\varphi_{\alpha}: U_{\alpha} \rightarrow \mathbb{R}^d$$



# Riemannian Manifolds

- **Smooth manifold** in a sense that if two patches  $U_\alpha$  and  $U_\beta$  overlap then the transition function

$$\varphi_\beta \circ \varphi_\alpha^{-1}: \varphi(U_\alpha \cap U_b) \rightarrow \mathbb{R}^d$$

is smooth

- **Riemannian manifold**  $\mathcal{M}$  is a smooth manifold equipped with a Riemannian metric
- The Riemannian manifold  $\mathcal{M}$  **inherits all the geometry feature** from its local system in  $\mathbb{R}^d$

# Riemannian Manifolds

- Given  $f: \mathcal{M} \rightarrow \mathbb{R}$ , we can define gradient and Laplacian over  $\mathcal{M}$  as

Gradient:

$$\nabla f_{\mathcal{M}}(x) = \nabla f(\mathbf{x}) = \left( \frac{\partial}{\partial x_1} f(\mathbf{x}), \dots, \frac{\partial}{\partial x_d} f(\mathbf{x}) \right)$$

Laplacian:

$$\Delta f_{\mathcal{M}}(x) = \Delta f(\mathbf{x}) = - \frac{\partial^2}{\partial x_1^2} f(\mathbf{x}) - \dots - \frac{\partial^2}{\partial x_d^2} f$$



# Semi-supervised Learning

## ■ Problem setting

- Labeled data:  $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)$
- Unlabeled data:  $\mathbf{x}_{l+1}, \dots, \mathbf{x}_{l+u}$
- $\mathbf{x}_i \in X \subseteq \mathbb{R}^d, y_j \in Y \subseteq \mathbb{R}$
- Estimate a predictor  $f: X \rightarrow \mathbb{R}$

## ■ Smoothness assumption

- **Probabilistic version:**  $P$  is a probability distribution on  $X \times Y$ , then **conditional distributions**  $P(y|\mathbf{x})$  are **smooth** with respect to the **marginal**  $P(\mathbf{x})$

# Semi-supervised Learning

- Smoothness assumption

- Manifold version

- $\nabla_{\mathcal{M}} f(x)$  measures local smoothness for  $f$
- The global measure of the smoothness for  $f$  is

$$\|f\|_I^2 = \int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f(x)\|^2 dP(x)$$

- Laplace-Beltrami operator  $\mathcal{L}$  on the manifold

$$\mathcal{L}f = -\operatorname{div} \nabla_{\mathcal{M}} f(x)$$

- $\|f\|_I^2 = \int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f(x)\|^2 dP(x) = \int_{\mathcal{M}} f \mathcal{L}f dP(x)$

# Manifold Regularization

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^l V(\mathbf{x}_i, y_i, f) + \gamma_A \|f\|_K^2 + \gamma_I \|f\|_I^2$$

└──────────────────┘     └──────────┘     └──────────┘  
Loss function     Penalty term     Additional term for  
intrinsic geometry

- $\mathcal{H}$  is a Reproducing Kernel Hilbert Space (RKHS) associated with a kernel  $K$  (polynomial, radial basis functions, etc.).
- Common choice for the loss function  $V$ 
  - Squared loss  $V = (y - f(\mathbf{x}))^2$
  - Hinge loss  $V = \max[0, 1 - yf(\mathbf{x})]$

# Manifold Regularization

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^l V(\mathbf{x}_i, y_i, f) + \gamma_A \|f\|_K^2 + \gamma_I \|f\|_I^2$$

The classical Representer Theorem gives

$$f^*(\mathbf{x}) = \sum_{i=1}^l \alpha_i K(\mathbf{x}_i, \mathbf{x}) + \int_{\mathcal{M}} \alpha(\mathbf{z}) K(\mathbf{z}, \mathbf{x}) dP(\mathbf{z})$$

# Empirical Representation of Manifold

- The **intrinsic geometry term**

$$\|f\|_I^2 = \int_{\mathcal{M}} f \mathcal{L}f dP(x)$$

**cannot directly evaluated** since the marginal distribution  $P(x)$  and locally embedded manifold  $\mathcal{M}$  in  $\mathbb{R}^d$  are unknown

# Empirical Representation of Manifold



- We discretize

Manifold  $\mathcal{M} \leftrightarrow$  Graph  $\mathcal{G}(V, E)$ ,  $V = \{\mathbf{x}_i\}_{i=1}^{l+u}$ ,  $E = \{e_{ij}\}$

Laplace-Beltrami operator  $\mathcal{L} \leftrightarrow$  Laplacian operator  $L$

$$L = D - W$$

$W$ : adjacency matrix,  $D = \text{diag}\{D_{ii} = \sum_j W_{ij}\}$

$$\|f\|_l^2 \approx \frac{1}{(u+l)^2} \mathbf{f}^T L \mathbf{f}$$

$$\mathbf{f} = [f(\mathbf{x}_1), \dots, f(\mathbf{x}_{l+u})]^T$$

# Manifold Regularization (Empirical version)

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}} \frac{1}{l} \sum_{i=1}^l V(\mathbf{x}_i, y_i, f) + \gamma_A \|f\|_K^2 + \gamma_I \frac{1}{(u+l)^2} \mathbf{f}^T \mathbf{L}^p \mathbf{f}$$

- $\mathcal{H}$  is a **Reproducing Kernel Hilbert Space** (RKHS) associated with a kernel  $K$ .

The classical Representer Theorem gives

$$f^*(\mathbf{x}) = \sum_{i=1}^{l+u} \alpha_i K(\mathbf{x}_i, \mathbf{x})$$

# Data-dependent Kernel Learning

$$f^* = \operatorname{argmin}_{f \in \tilde{\mathcal{H}}} \frac{1}{l} \sum_{i=1}^l V(\mathbf{x}_i, y_i, f) + \gamma_A \|f\|_{\tilde{K}}^2$$

- The minimizer admits

$$f^*(\mathbf{x}) = \sum_{i=1}^l \alpha_i \tilde{K}(\mathbf{x}_i, \mathbf{x})$$

- Warped kernel  $\tilde{K}$  defined by

$$\tilde{K}(\mathbf{x}, \mathbf{z}) = K(\mathbf{x}, \mathbf{z}) - \mathbf{K}_{\mathbf{X}}^T (\mathbf{I} + \mathbf{M}\mathbf{K})^{-1} \mathbf{M}\mathbf{K}_{\mathbf{Z}}$$

$$\mathbf{K}_{\mathbf{X}} = [K(\mathbf{x}, \mathbf{x}_1), \dots, K(\mathbf{x}, \mathbf{x}_{l+u})]^T, \mathbf{K}_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$$

$\mathbf{M}$  is a symmetric positive semi-definite matrix



# Multiscale Manifold Learning

Manifold  $\mathcal{M} \leftrightarrow$  Multiscale Graph  $\mathcal{G}(V, E^\alpha)$ ,  $\alpha = 1, \dots, n$

Multiscale graph Laplacian

$$L = \sum_{\alpha=0}^n c_\alpha L_\alpha^{p_\alpha}$$

where  $L_\alpha = D_\alpha - W_\alpha$ ,  $W_\alpha$  is an adjacent edge matrix with

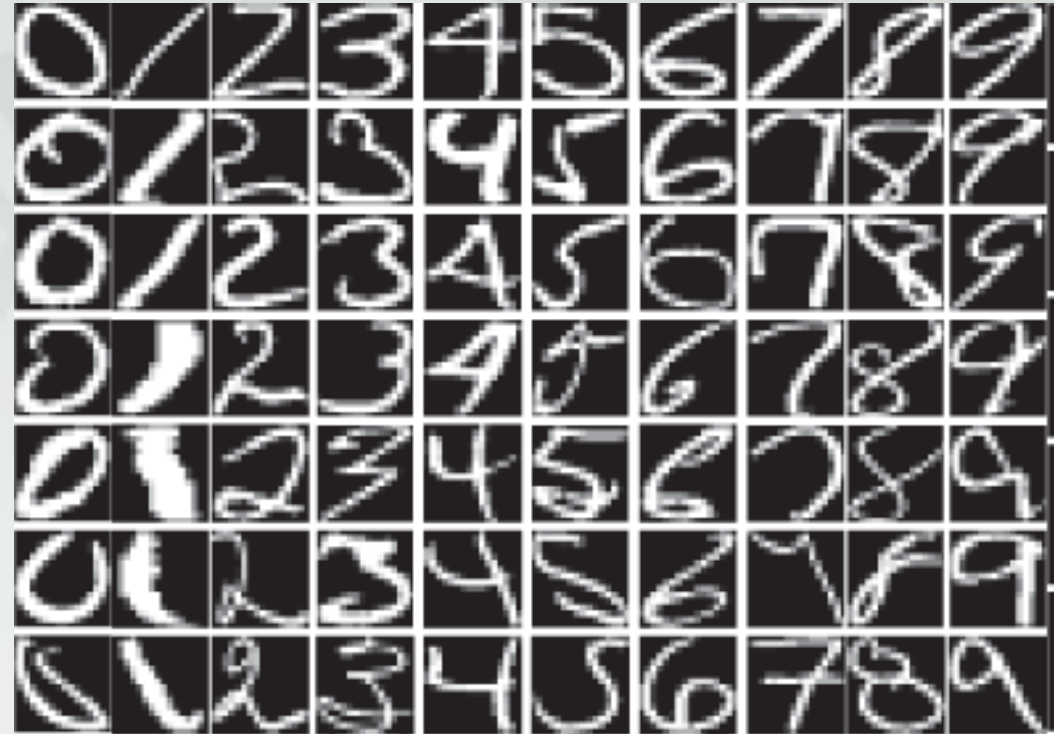
$$[W_\alpha]_{ij} = \frac{1}{\sqrt{\sigma_\alpha}} \psi_\alpha \left( \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{\sigma_\alpha} \right) e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma_\alpha^2}}$$

$\psi_\alpha$  is an  $\alpha^{th}$  order normalized Hermite polynomial,  
and  $[D_\alpha]_{ii} = \sum_j [W_\alpha]_{ij}$

# USPS Handwriting Data Set

(Chappele, Zien, *AI & Stat. 2005*; Sindhvani et. al., *ICM 2005*; Nguyen, Wei 2018)

Dataset → Algorithm ↓	Uspst
Graph-Trans	21.3
TSVM	26.5
Graph-density	16.9
∇TSVM	17.6
LDS	15.8
LapSVM	12.7
LapRLS	12.7
M-LapSVM (1 ker)	13.89
M-LapRLS (1 ker)	13.89
M-LapSVM (2 kers)	<b>9.43</b>
M-LapRLS (2 kers)	<b>9.43</b>
M-LapSVM (2 kers)	9.52
M-LapRLS (2 kers)	9.52



(Image courtesy of Wang et. al., 2013)

Dataset	No. of Classes	Sample dim.	No. of data	No. of labeled data
Uspst	10	256	2007	50

# Manifold-based Molecular Representation

- Use proxy of manifold, **multiscale weighted color subgraph**  $\mathcal{G}(V^d, E)$ , to describe the molecular structure

$$V^d = \left\{ (\mathbf{r}_i, \alpha_i, \beta_i) \mid \mathbf{r}_i \in \mathbb{R}^3, \alpha_i \in \mathcal{C}, \beta_i \in \mathcal{S}, \right. \\ \left. \exists j \|\mathbf{r}_i - \mathbf{r}_j\| < d, \beta_i + \beta_j = 1 \right\}$$

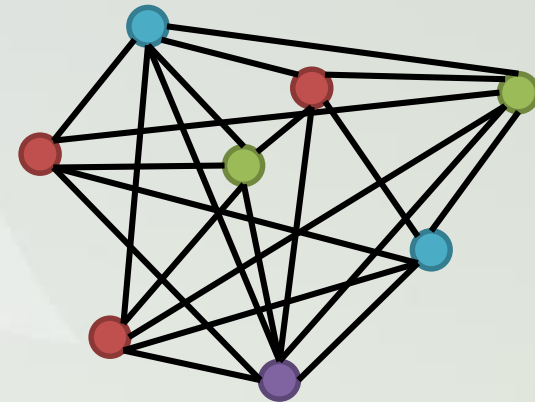
$$\mathcal{C} = \{\text{C, N, O, S, P, F, Cl, Br, I}\}$$

$$\mathcal{S} = \{0, 1\}$$

$$\text{Ordered colored edge set } \mathcal{P} = \{(\alpha, 0), (\alpha', 1)\}$$

Edges describe potential pairwise atomic interactions

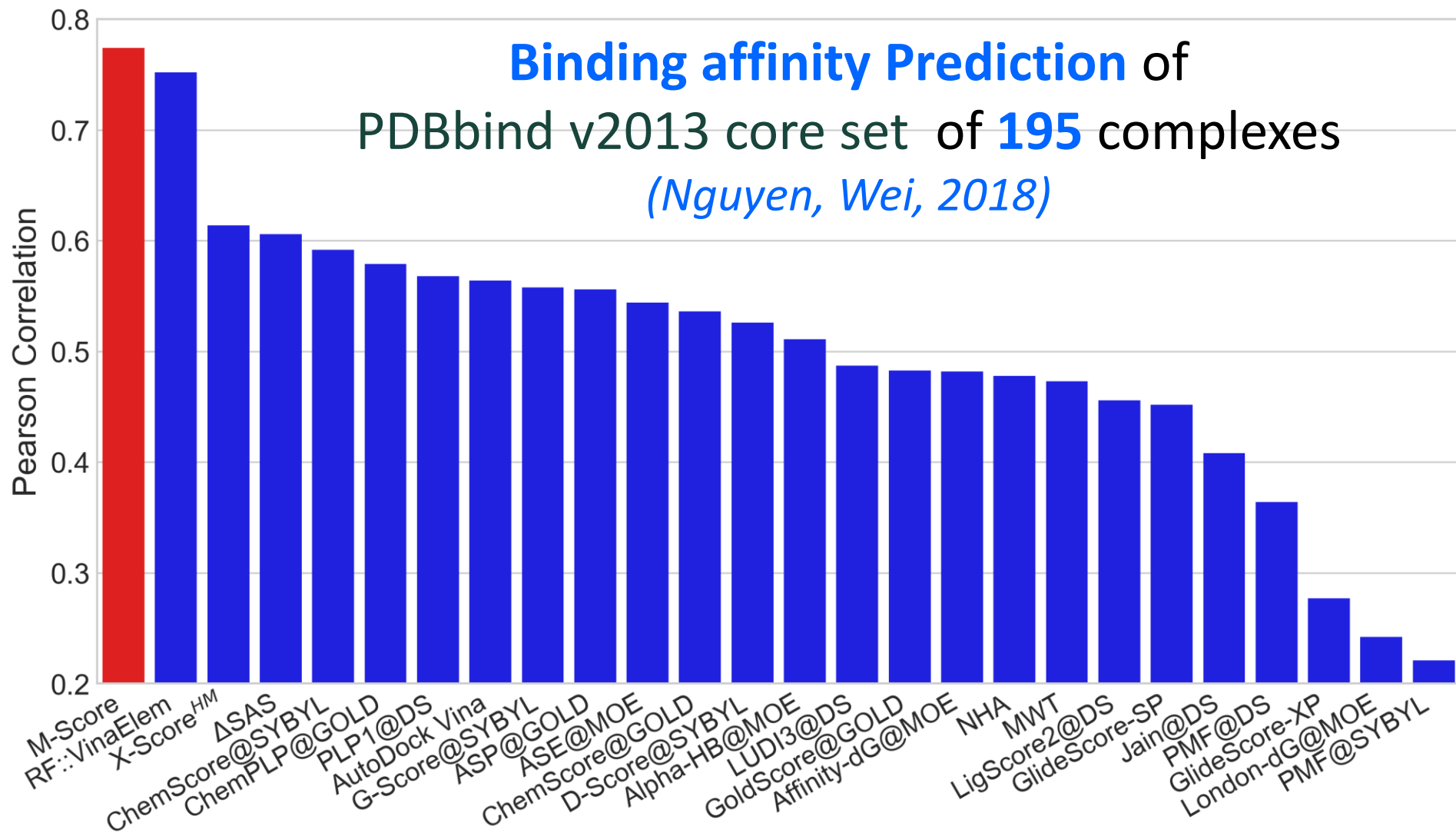
$$E_{\mathcal{P}} = \left\{ K(\mathbf{r}_i, \mathbf{r}_j) \mid \left( (\alpha_i, \beta_i), (\alpha_j, \beta_j) \right) \in \mathcal{P} \right\}$$



**DUD:** Classification of ligands and decoys  
**128,374** protein-ligand/decoy pairs

Method	AUC	Ref.
<b>M-VS</b>	<b>0.81</b>	<i>Nguyen, Wei 2018</i>
ICM	0.79	<i>Neves et. al., J Comput Aided Mol Des 2012</i>
Glide SP	0.77	<i>Cross JB et. al., J Chem Inf Model. 2009</i>
Surflex	0.72	<i>Cross JB et. al., J Chem Inf Model. 2009</i>
Rosetta Ligand	0.65	<i>Armstrong et. al., J Comput Aided Mol Des 2010</i>
AutoDock Vina	0.64	<i>Armstrong et. Al., J Comput Aided Mol Des 2010</i>
FlexX	0.61	<i>Cross JB et. al., J Chem Inf Model. 2009</i>

# Binding affinity Prediction of PDBbind v2013 core set of **195** complexes *(Nguyen, Wei, 2018)*

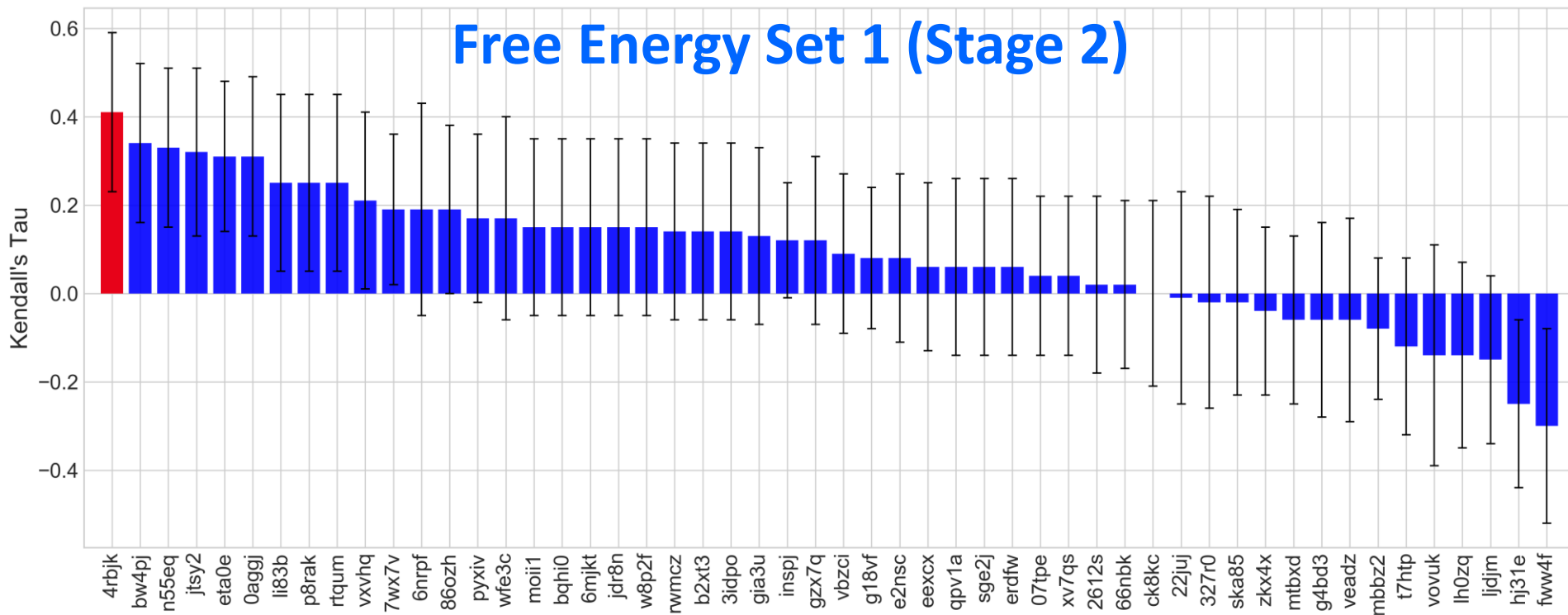


# Drug Design and Discovery Resource (D3R) Grand Challenge 2

**Given:** Farnesoid X receptor (FXR) and 102 ligands

**Challenges:** Dock 102 ligands, predict poses, compute binding free energies and affinity ranking

**Results: 1<sup>st</sup> place** in Free energy ranking in Free energy Set 1 (stage2)



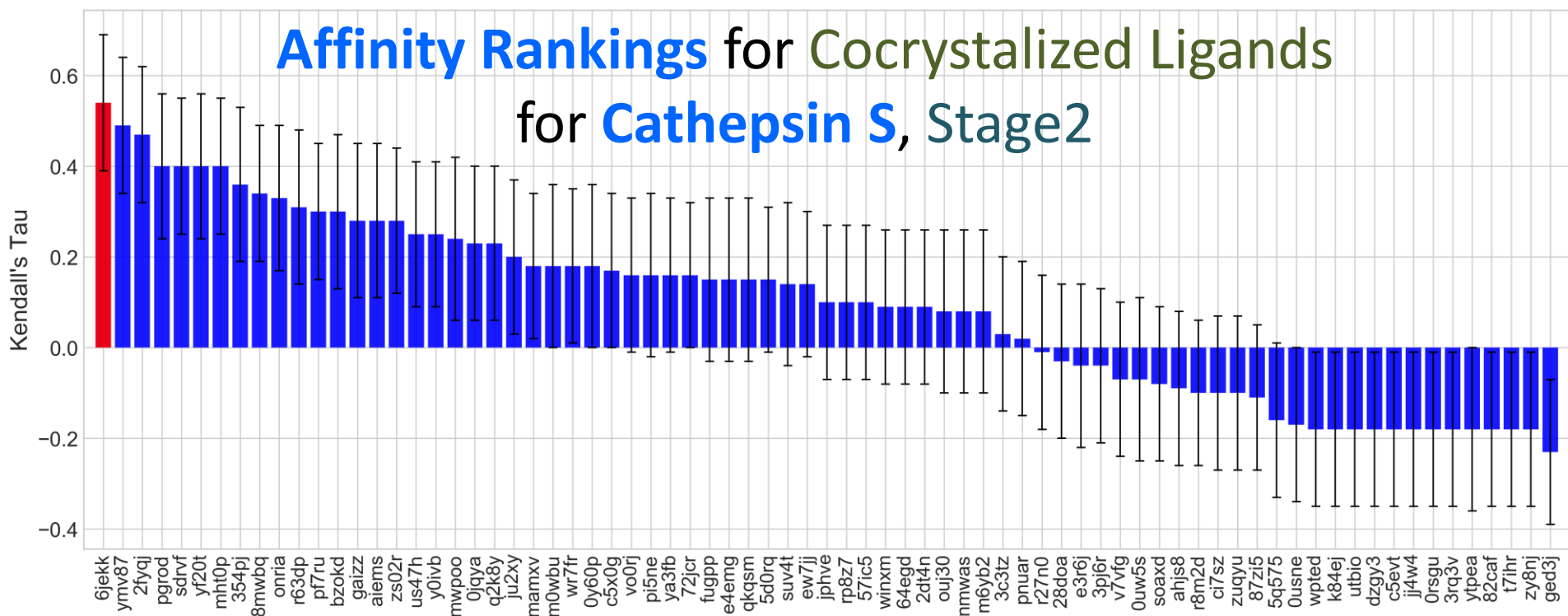
# Drug Design and Discovery Resource (D3R) Grand Challenge 3

**Given:** 6 protein targets: Cathepsin S (CatS), Vascular endothelial growth factor receptor 2 (VEGFR2), Janus Kinase 2 (JAK2), p38- $\alpha$ , Angiopoietin-1 receptor (TIE2), ABL1

**Challenges:** docking, binding free energy predictions, affinity ranking

**Results:** Our predictions were ranked **1<sup>st</sup> in 10 out of 26** predictive tasks

# Drug Design and Discovery Resource (D3R) Grand Challenge 3







Thank you!