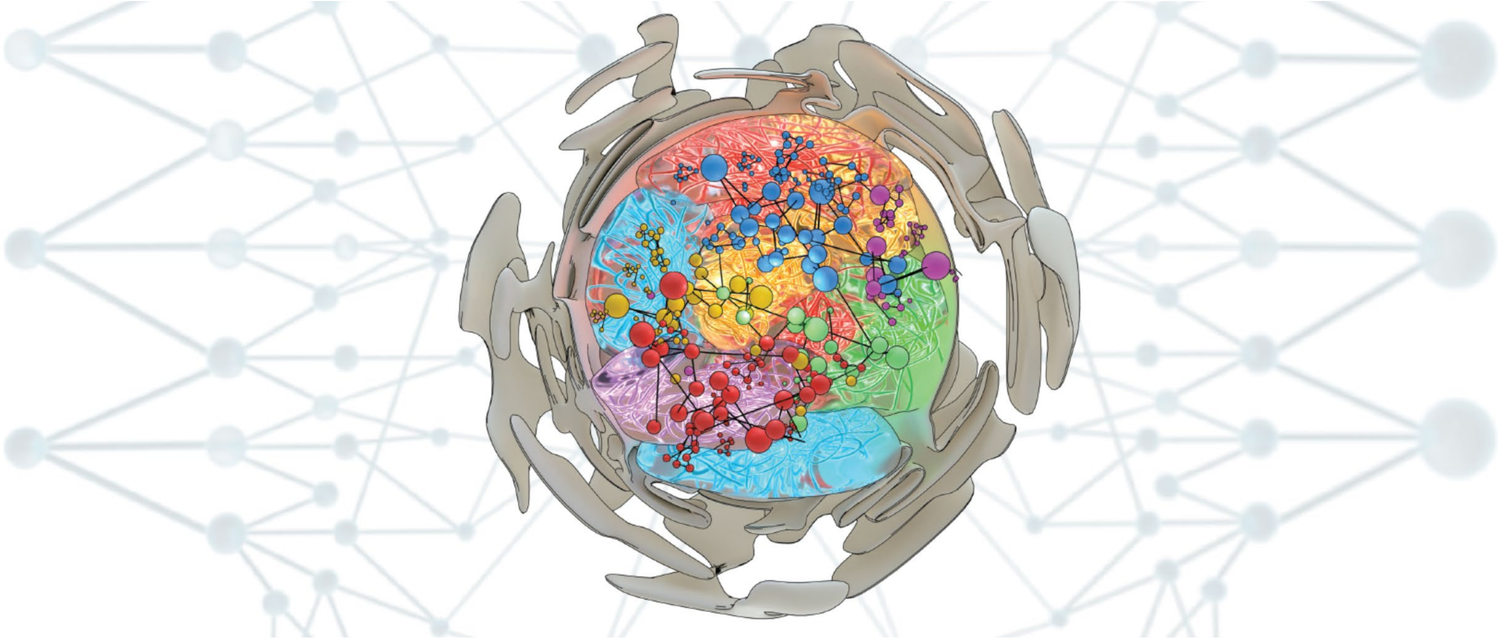
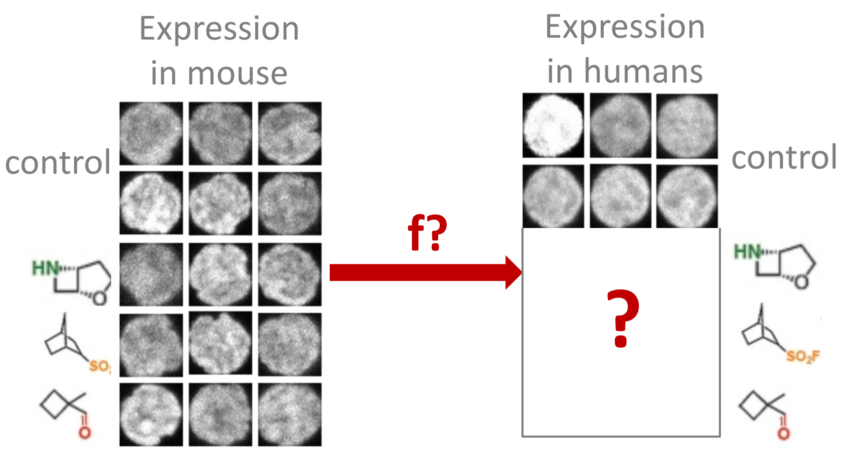
Causal Imputation

Caroline Uhler (MIT & Broad Institute)

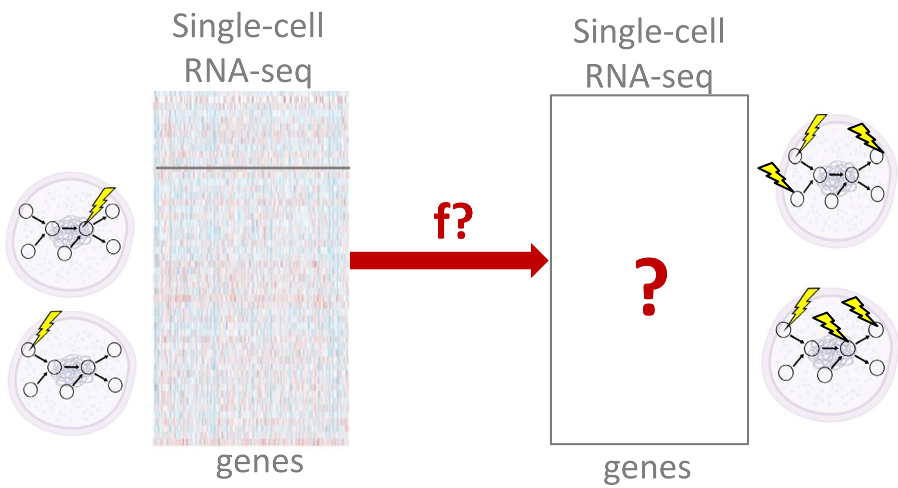


Causal imputation problems in single-cell biology

Transport between contexts (cell types, diseases, organisms)



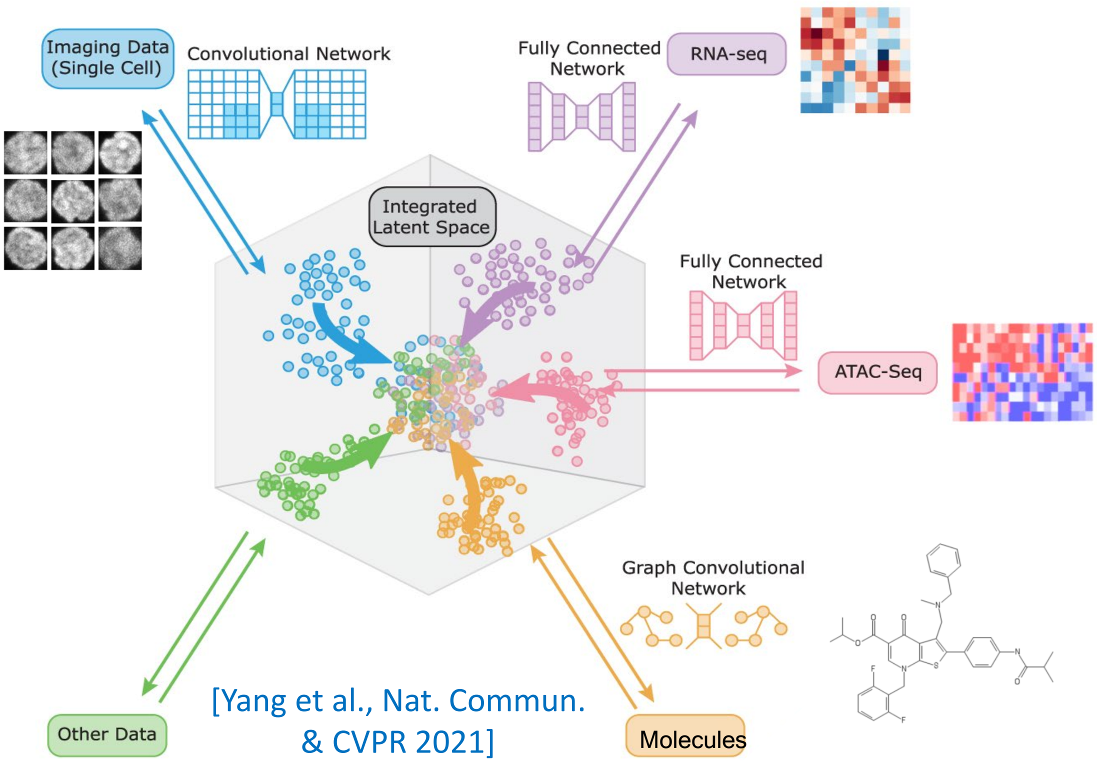
Transport between interventions (combinatorial, continuous, …)



Causal Representation Learning should not only allow for causal feature discovery, but also for predicting the effect of unseen interventions in new contexts

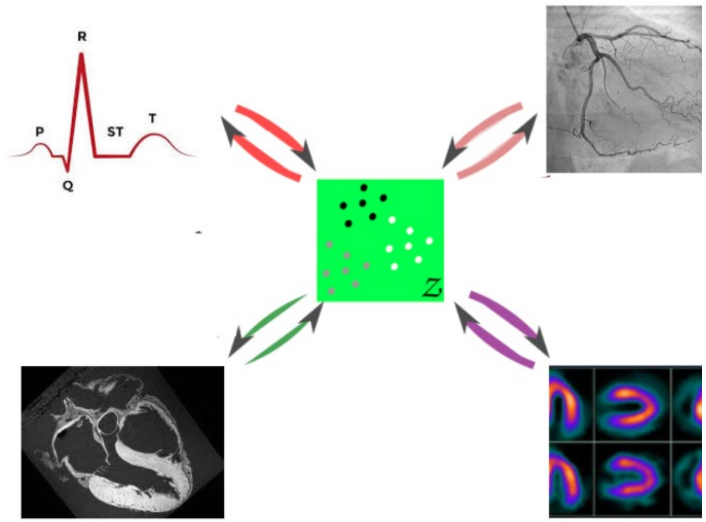
Multi-modal learning for getting at causal features

Multi-modal integration / translation



[Yang et al., Nat. Commun. & CVPR 2021]

Multi-modal integration for genetic association studies



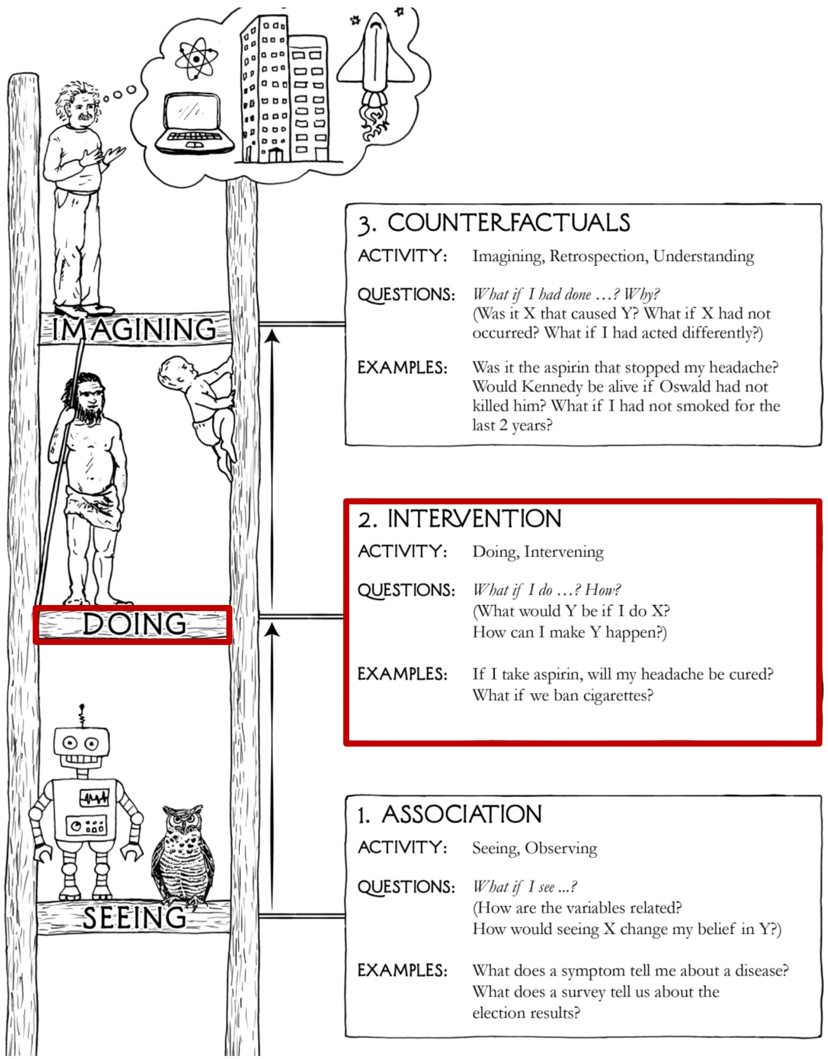
[Radhakrishnan et al., bioRxiv: 10.1101/2022.05.26.493497 2022]

Multi-modal learning as a tool for causal feature discovery by learning integrated latent spaces: Causal features should be invariant to modality in which they are measured

Invariant prediction for causal inference: Peters et al., 2016 **Invariant risk minimization:** Arjovsky et al., 2019

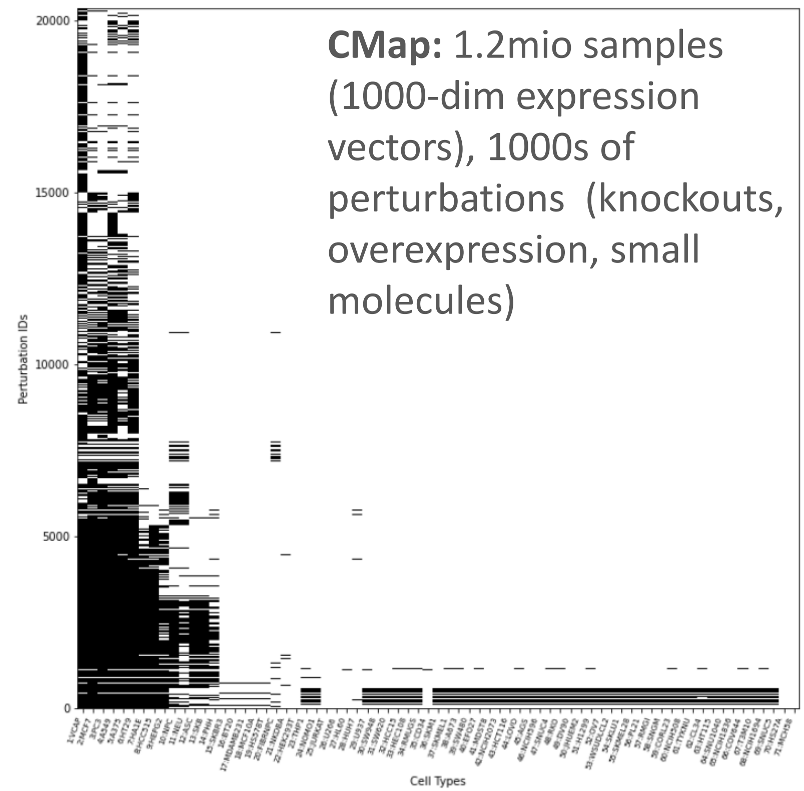
Causal inference by predicting interventions

Judea Pearl's causal hierarchy



J. Pearl, The Book of Why, 2018

Causal imputation



CMap: 1.2mio samples (1000-dim expression vectors), 1000s of perturbations (knockouts, overexpression, small molecules)

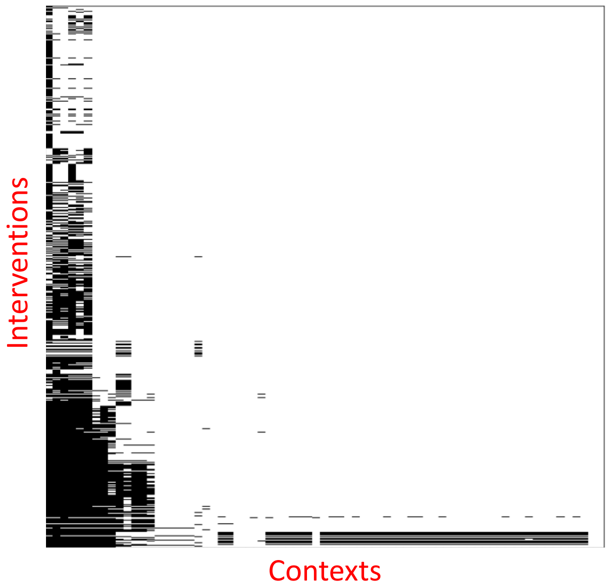
We are great at solving prediction problems:

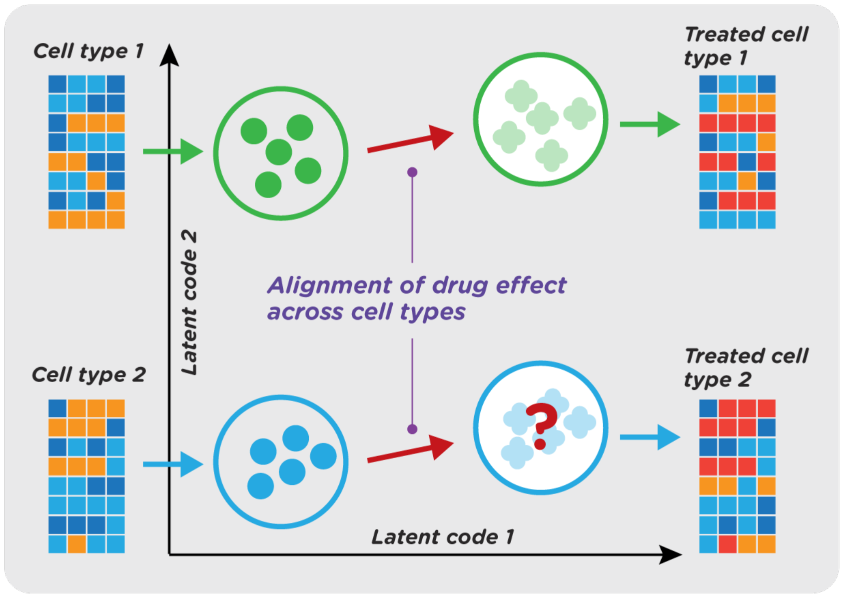
Availability of interventional data allows us to turn causal questions into prediction problems

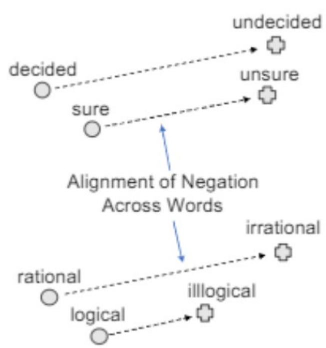
Causal imputation

3 Approaches:

1) Style transfer **2) Synthetic control/interventions** **3) NTK for matrix completion**

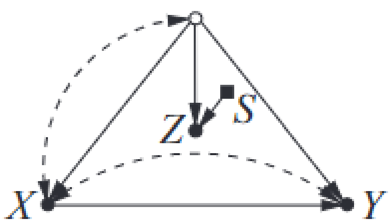




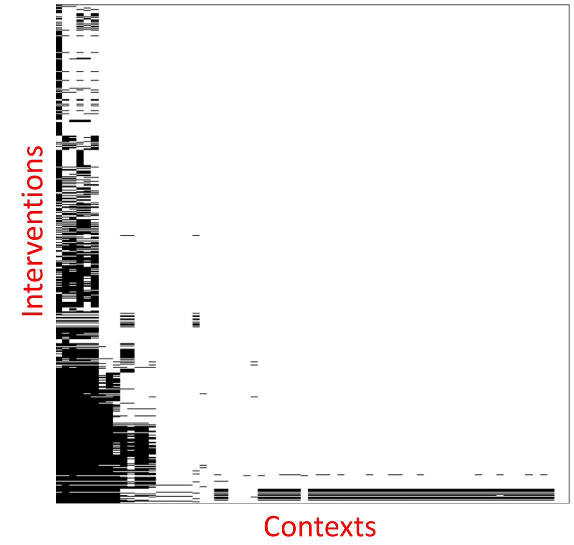


Word2Vec analogy

- Given the causal graph, then necessary and sufficient conditions for causal transportability (i.e. transport across contexts) are known [Bareinboim & Pearl, NeurIPS 2014, PNAS 2016, etc.]



Synthetic interventions & synthetic control



Synthetic control (Abadie et al., 2003 & 2010):

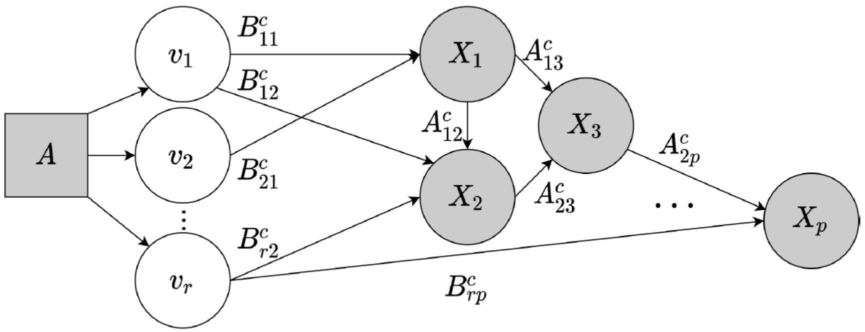
- From policy evaluation literature
- Write a unit that did get intervention as a linear combination of other units that did not get intervention to estimate what its value would have been without intervention and estimate average treatment effect

Synthetic interventions (Agarwal et al., 2020):

- Generalization to estimate value under intervention
- Testable conditions for consistency of approach (linearity and low-rank)

$$\mathbf{x}^{ca} = A^c\mathbf{x}^{ca} + B^c\mathbf{v}^a$$

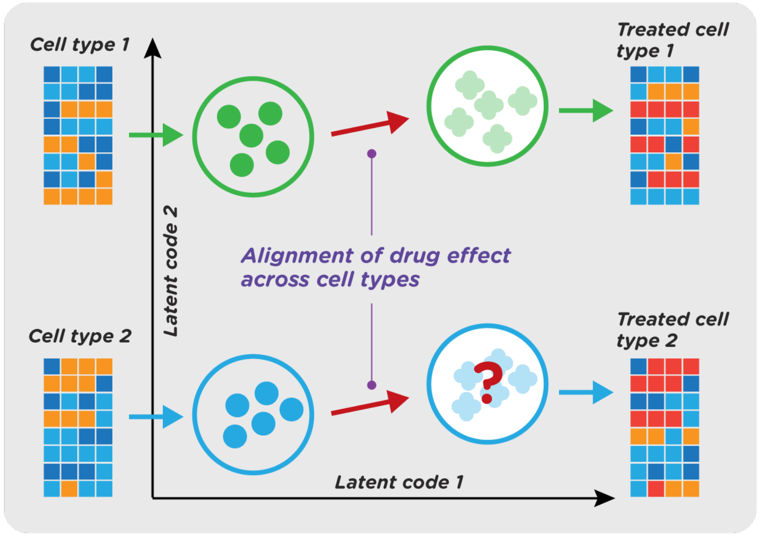
We generalized this to use relationships between actions/interventions and showed that the algorithm is consistent for the following biologically plausible model:

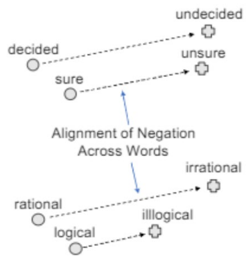


[Squires, Shen, Agarwal, Shah & Uhler: CLeaR 2022]

Over-parameterization to align causal effects

Latent spaces that align causal effects

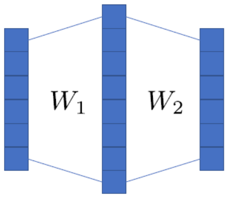


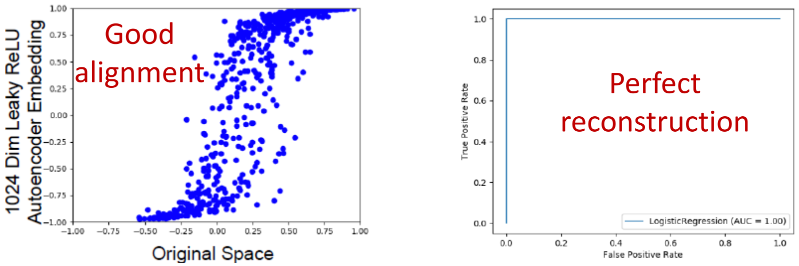


Word2Vec analogy

Over-parameterized neural nets

- Classification setting: neural nets can interpolate training data while generalizing!
- Over-parameterized autoencoders: interpolate, generalize, and align drug signatures across cell types!

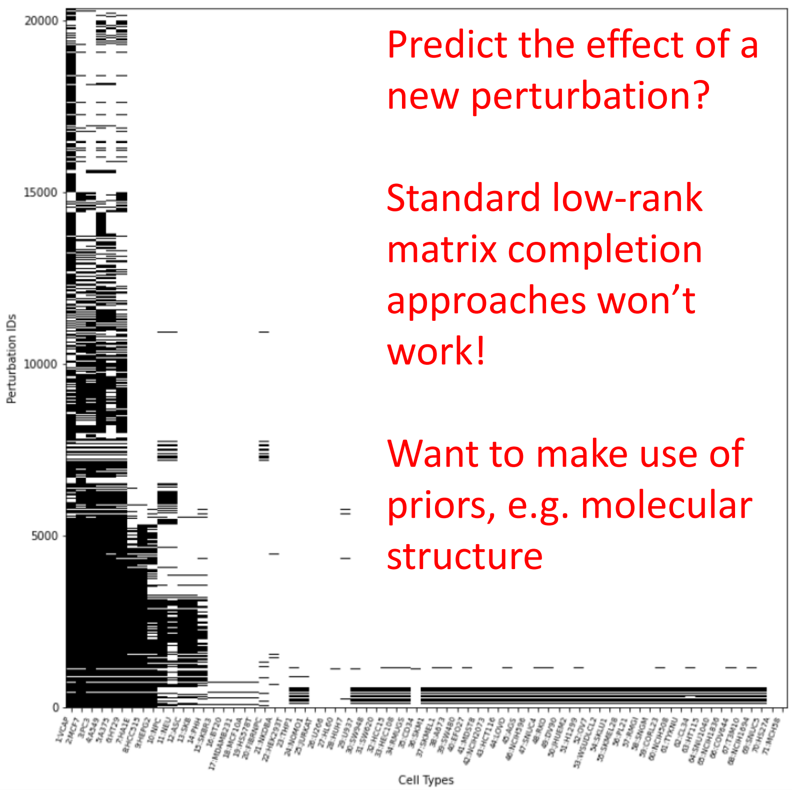




[Belyaeva et al., Nature Commun. 2021; Radhakrishnan, Belkin & Uhler, PNAS 2020]

Over-parameterized autoencoders provide more "space" to align causal effects:
We need to study their inductive biases and the interplay with causality!

Matrix completion with infinite width networks

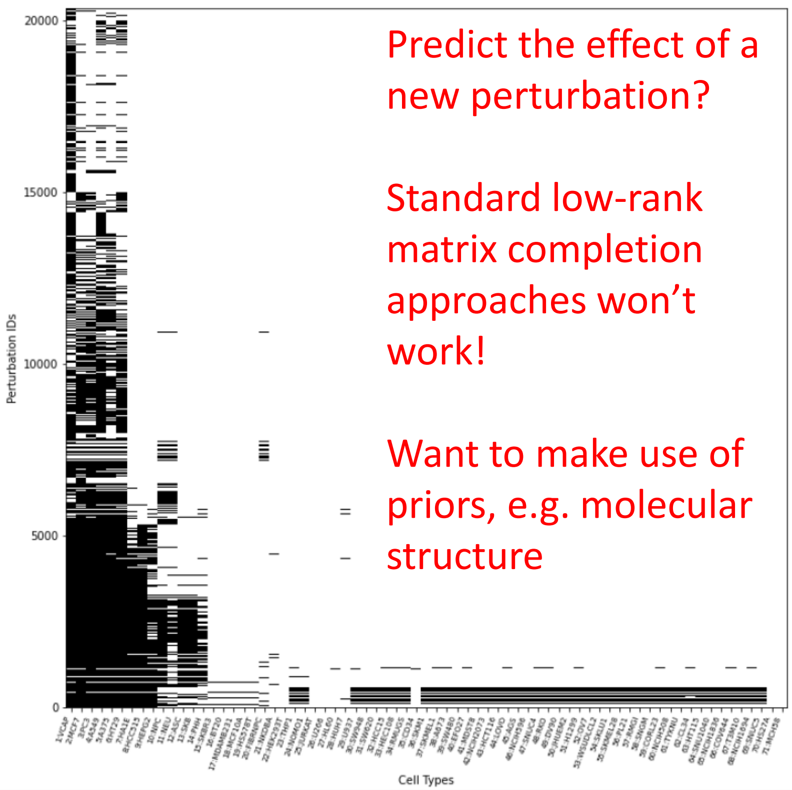


Predict the effect of a new perturbation?

Standard low-rank matrix completion approaches won't work!

Want to make use of priors, e.g. molecular structure

Matrix completion with infinite width networks



Predict the effect of a new perturbation?

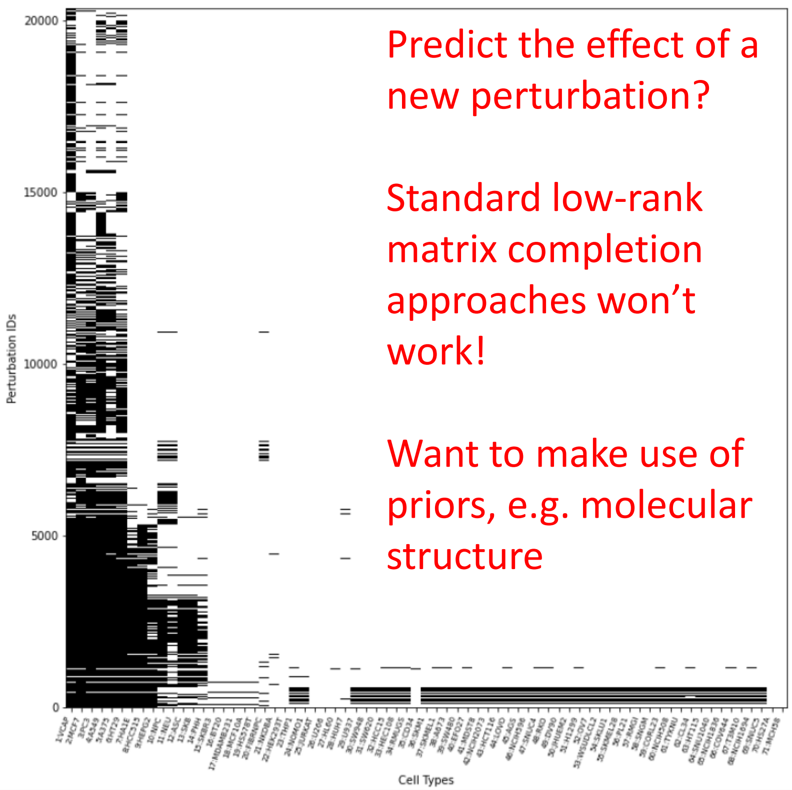
Standard low-rank matrix completion approaches won't work!

Want to make use of priors, e.g. molecular structure

Standard low-rank factorization:

$$Y = W_1 \times W_2 \times Z, \qquad Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Matrix completion with infinite width networks

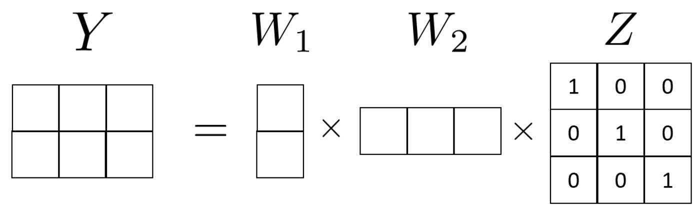


Predict the effect of a new perturbation?

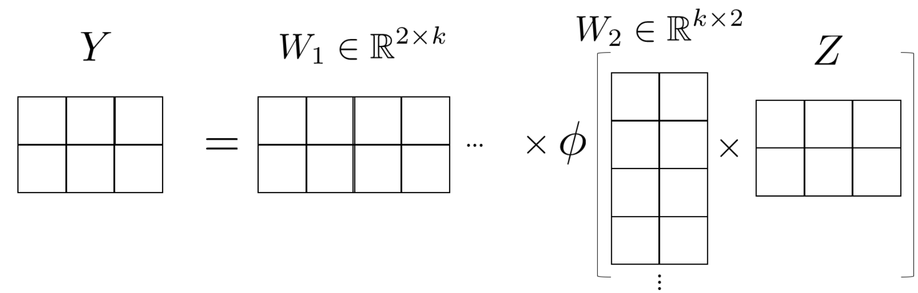
Standard low-rank matrix completion approaches won't work!

Want to make use of priors, e.g. molecular structure

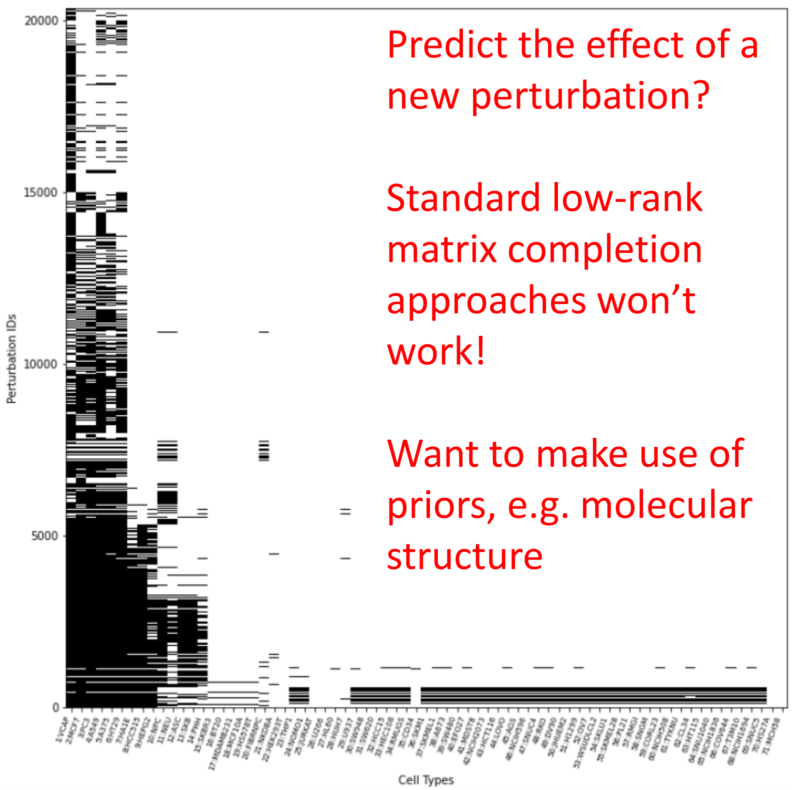
Standard low-rank factorization:



Proposed approach: k->∞ & general Z



Matrix completion with infinite width networks

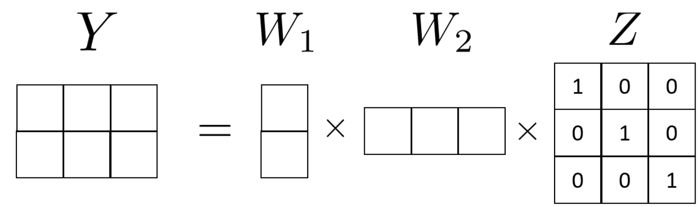


Predict the effect of a new perturbation?

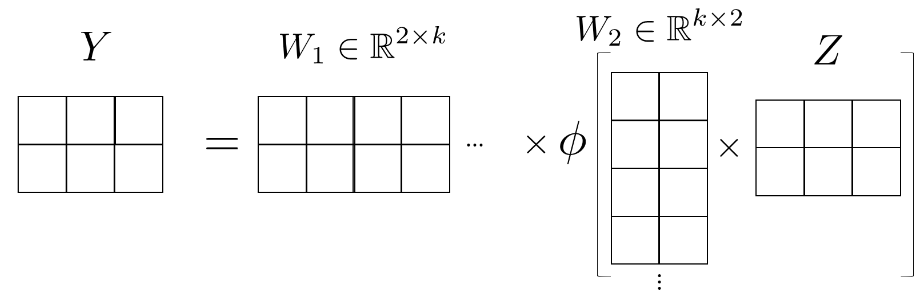
Standard low-rank matrix completion approaches won't work!

Want to make use of priors, e.g. molecular structure

Standard low-rank factorization:



Proposed approach: k->∞ & general Z



Neural tangent kernel (NTK) for classification setting:

[Jacot, Gabriel, Hongler; 2018]

Given a dataset $\{x^{(i)}, y^{(i)}\}_{i=1}^{n}$; as network width goes to infinity, these yield same functional solution:

$$\arg\min_{\mathbf{W}} \sum_{i=1}^{n} (y^{(i)} - f_{x^{(i)}}(\mathbf{W}))^2 \quad \longleftrightarrow \quad \arg\min_{w} \sum_{i=1}^{n} (y^{(i)} - w \nabla f_{x^{(i)}}(\mathbf{W}^{(\mathbf{0})}))^2$$

Solve using kernel regression with the NTK: $K(x, x') = \langle \nabla f_x(\mathbf{W}^{(0)}), \nabla f_{x'}(\mathbf{W}^{(0)}) \rangle$

Adit Radhakrishnan: [6.S088 Modern Machine Learning: Simple Methods that Work (mit.edu)]

NTK for matrix completion: simple, fast & flexible

Theorem 1. *Assume* $Z = \{z^{(i)}\}_{i=1}^{n} \in \mathbb{R}^{p\times n}$, *where each column is normalized with* $\|z^{(i)}\|_2 = 1$. *Let* $f_Z(\mathbf{W})$ *be a* d *layer fully connected network with nonlinearity* $\phi(x) = \max(x, 0)$
Then, as widths $k_2, k_3, \ldots, k_d \to \infty$, *the NTK for matrix completion with* $f_Z(\mathbf{W})$ *is given by*

$$K(M_{ij}, M_{i'j'}) = \begin{cases} \kappa_d({z^{(j)}}^T z^{(j')}) & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases},$$

where $\kappa_d(\xi) = \check{\phi}^{(d)}(\xi) + \kappa_{d-1}(\xi)\frac{d\check{\phi}}{d\xi}(\check{\phi}^{(d-1)}(\xi))$, *and* $\check{\phi}^{(h)}(\xi) = \check{\phi}(\check{\phi}^{(h-1)}(\xi))$ *for* $h \geq 1$ *and* $\check{\phi}^{(0)}(\xi) = \xi$.

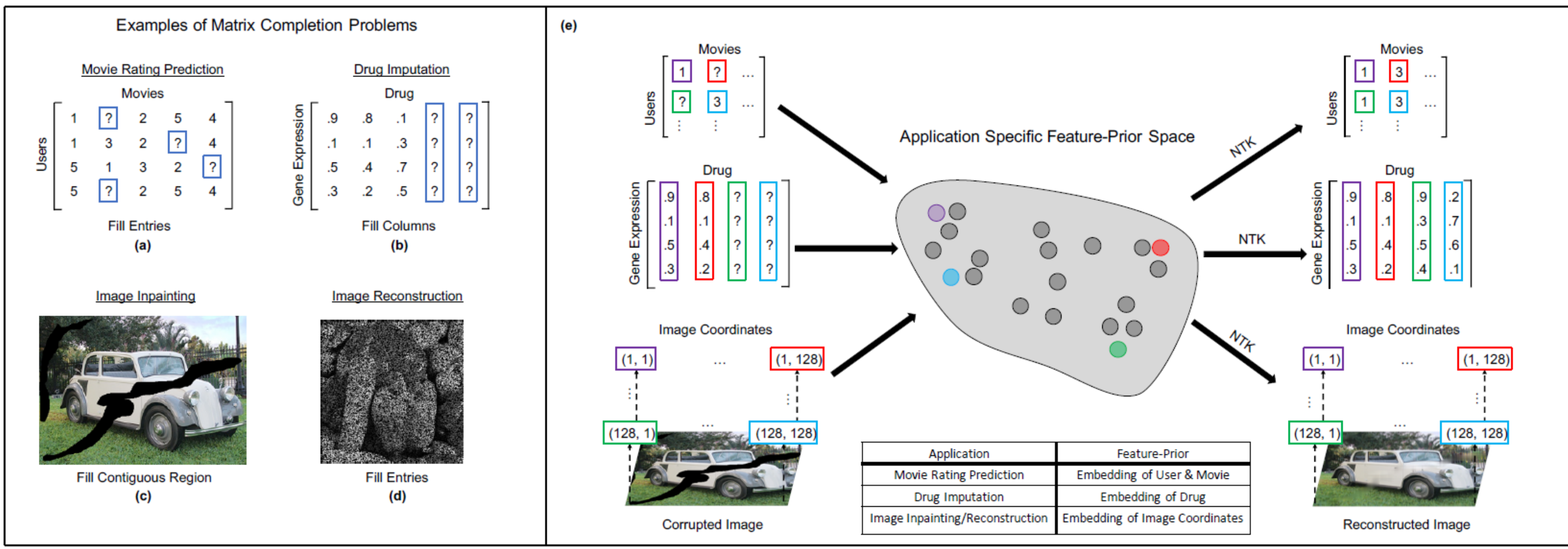
NTK for matrix completion: simple, fast & flexible

Theorem 1. *Assume $Z = \{z^{(i)}\}_{i=1}^{n} \in \mathbb{R}^{p \times n}$, where each column is normalized with $\|z^{(i)}\|_2 = 1$. Let $f_Z(\mathbf{W})$ be a d layer fully connected network with nonlinearity $\phi(x) = \max(x, 0)$ Then, as widths $k_2, k_3, \ldots, k_d \to \infty$, the NTK for matrix completion with $f_Z(\mathbf{W})$ is given by*

$$K(M_{ij}, M_{i'j'}) = \begin{cases} \kappa_d({z^{(j)}}^T z^{(j')}) & \text{if } i = i' \\ 0 & \text{if } i \neq i' \end{cases},$$

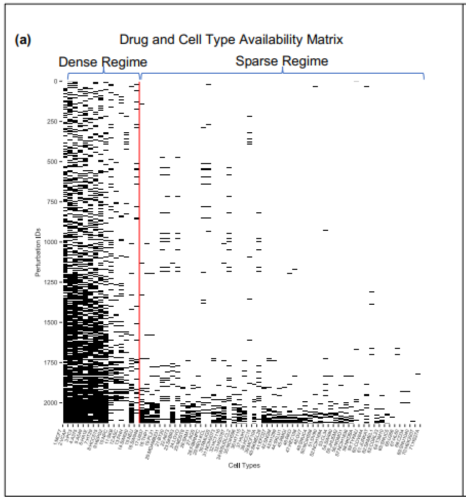
Z acts as a feature prior; akin to semisupervised learning

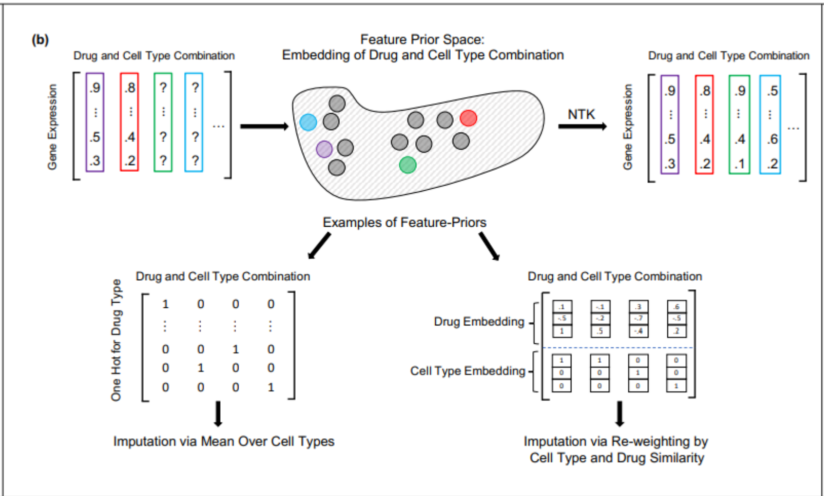
where $\kappa_d(\xi) = \check{\phi}^{(d)}(\xi) + \kappa_{d-1}(\xi) \frac{d\check{\phi}}{d\xi}(\check{\phi}^{(d-1)}(\xi))$, and $\check{\phi}^{(h)}(\xi) = \check{\phi}(\check{\phi}^{(h-1)}(\xi))$ for $h \geq 1$ and $\check{\phi}^{(0)}(\xi) = \xi$.



[Radhakrishnan, Stefanakis, Belkin & Uhler, PNAS 2022]

Applications of NTK for matrix completion





(c)

CMap
(Sparse Regime)

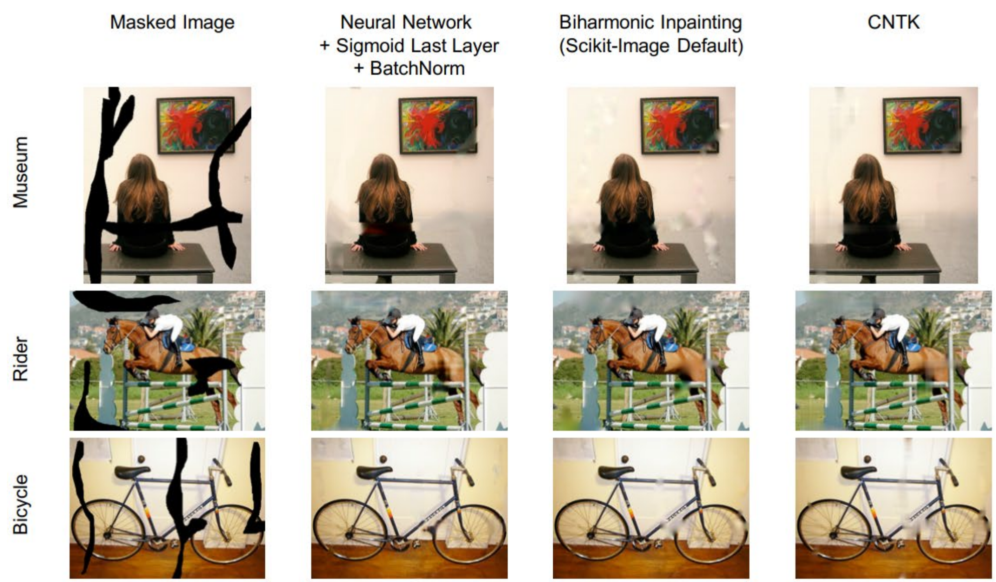
Evaluation Metric*	Mean Over Cell Type (Naïve Baseline)	FaLRTC (Liu et al. 2013)	DNPP (Hodos et al. 2018)	NTK (Ours)
Pearson r	0.450	0.544	0.538	**0.573**
Mean R^2	0.197	0.285	0.278	**0.324**
Mean Cosine Similarity	0.448	0.536	0.532	**0.565**

(d)

CMap
(Full Dataset)

Evaluation Metric*	Mean Over Cell Type (Naïve Baseline)	FaLRTC (Liu et al. 2013)	DNPP (Hodos et al. 2018)	NTK (Ours)
Pearson r	0.374 ± 0.0004	0.545 ± 0.0003	0.556 ± 0.0003	**0.572 ± 0.0002**
Mean R^2	0.134 ± 10^(-5)	0.286 ± 0.0003	0.296 ± 0.0004	**0.320 ± 0.0002**
Mean Cosine Similarity	0.371 ± 10^(-5)	0.536 ± 0.0004	0.541 ± 0.0004	**0.554 ± 0.0002**

*Higher is better, with a maximum of 1.



(a)

Image	CNTK (PSNR)	Neural Network + Sigmoid Last Layer + BatchNorm (PSNR)	Biharmonic (PSNR)
Museum	**31.90**	30.69	30.03
White Car	28.66	**28.73**	26.20
Bicycle	27.67	28.57	**28.67**
Chair	29.87	**29.88**	27.81
Car Field	28.91	**29.94**	27.67
Rider	29.20	27.76	**29.38**
Library	**21.73**	20.76	17.71
Vase	**31.75**	31.51	28.96
Pool	34.51	33.08	**34.62**
Average	**29.36**	28.99	27.89

(b)

[Radhakrishnan, Stefanakis, Belkin & Uhler, PNAS 2022]

Acknowledgments

PhD students:
- Louis Cammarata
- Parmida Davarmanesh
- Daniel Paysan
- Adityanarayanan Radhakrishnan
- Chandler Squires
- Yitong Tseo
- Karren Yang
- Jiaqi Zhang
- Xinyi Zhang

MSc/undergraduate students
- Cathy Cai
- Emily Liu
- Max Ruiz Luyten
- Nten Nyiam
- Ishika Shah

Postdocs:
- Salil Bhate
- Wengong Jin
- Neriman Tokcan

Collaborators
- **Mikhail Belkin**
- **GV Shivashankar**

Funding:

ERIC AND WENDY SCHMIDT CENTER AT BROAD INSTITUTE

DEPARTMENT OF THE NAVY ONR Science & Technology

MIT

SIMONS FOUNDATION

IBM

NSF