Some benefits of machine learning with invariances

Stefanie Jegelka
TUM and MIT

based on joint work with
Derek Lim, Joshua Robinson, Behrooz Tahmasebi,
Lingxiao Zhao, Haggai Maron, Tess Smidt, Suvrit Sra

Learning with invariances

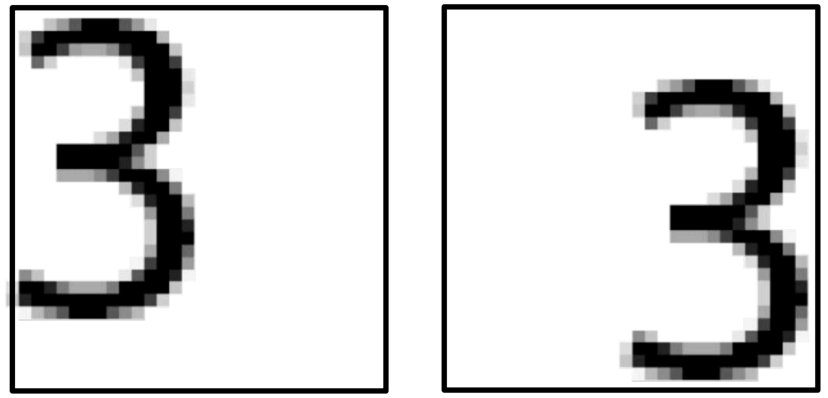
- **Standard learning setup:**
  given data $(x_1, y_1), \ldots, (x_n, y_n)$ estimate $\hat{f} \in \mathscr{F}$
  such that $\mathbb{E}[\ell(\hat{f}(X), Y)]$ is small

- **Learning with invariances:** all of the above, plus:
  select function $\hat{f} \in \mathscr{F}$ such that it is G-invariant for a given group G:

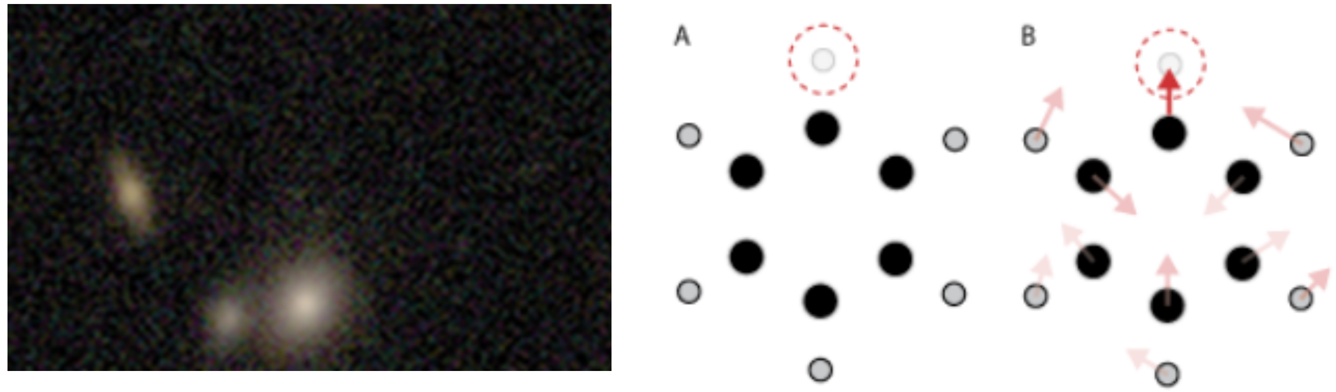
$$\hat{f}(g.x) = \hat{f}(x) \quad \forall g \in G, x \in \mathcal{X}$$

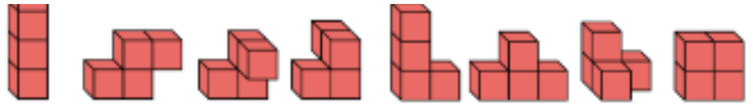
usually: $\mathscr{F}$ is a set of invariant functions

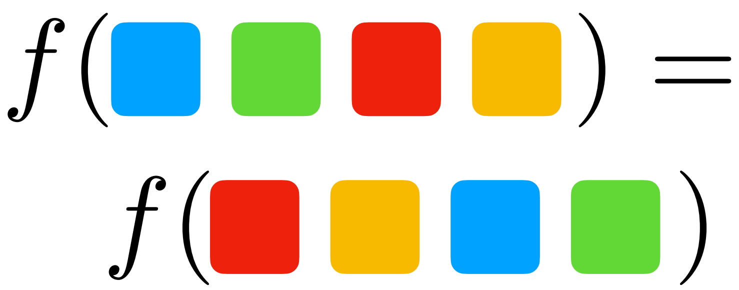
Examples



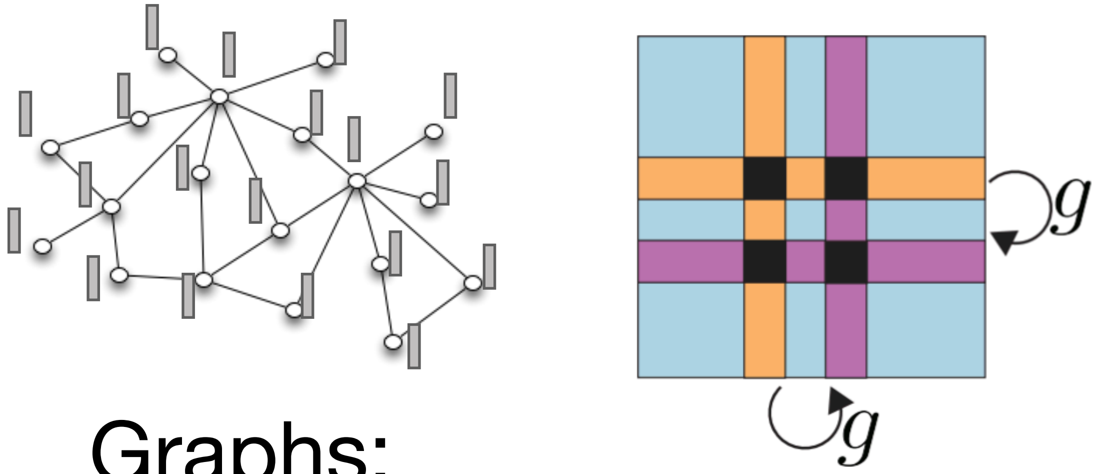
Images:
translation invariance



Point clouds: 
rotation/translation invariance



Sets:
permutation invariance



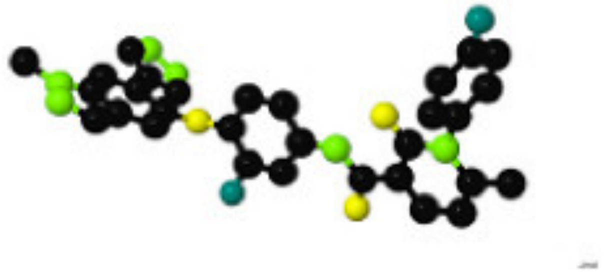
Graphs:
permutation invariance

Outline

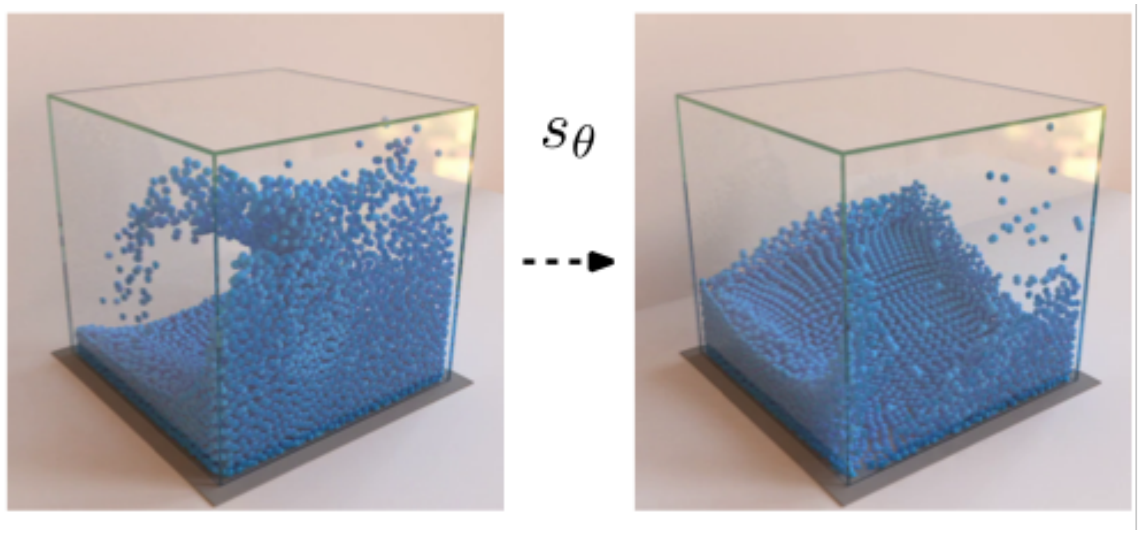
- A concrete example: (Graph) Neural Networks on eigenvectors
  D. Lim, J. Robinson, L. Zhao, T. Smidt, S. Sra, H. Maron, S. Jegelka. Sign and Basis Invariant Networks for Spectral Graph Representation Learning, ICLR 2023.
  D. Lim, J. Robinson, S. Jegelka, H. Maron. Expressive Sign Equivariant Networks for Spectral Geometric Learning. Neural Information Processing Systems (NeurIPS), 2023.

- Sample complexity gains from learning with invariances
  B. Tahmasebi, S. Jegelka. The Exact Sample Complexity Gain from Invariances for Kernel Regression on Manifolds, Neural Information Processing Systems (NeurIPS), 2023.

Machine Learning with Graph Data: Applications



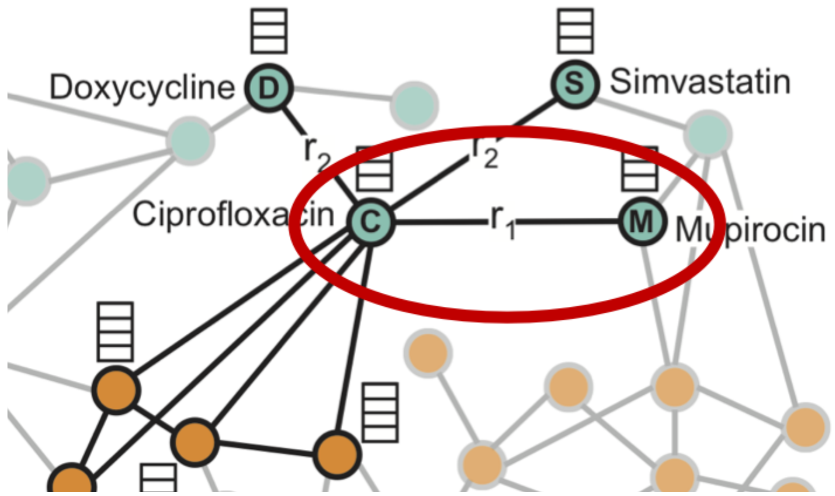
molecule property prediction
(Duvenaud et al 2015, Stokes et al 2020)



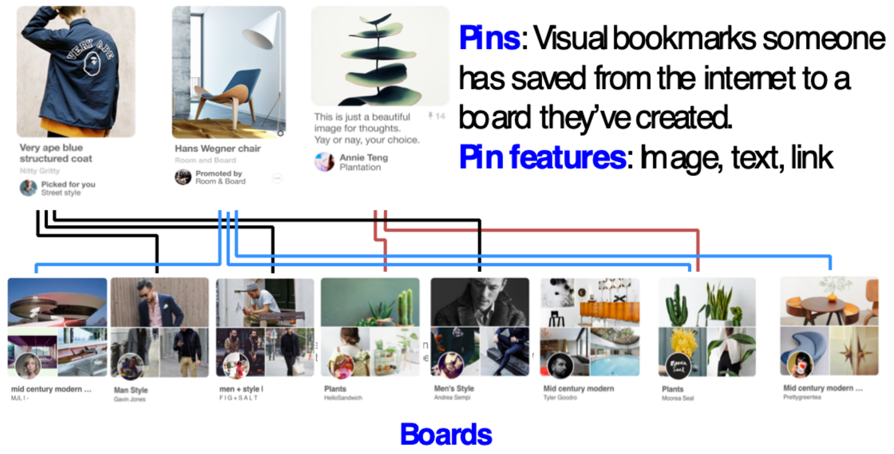
learning simulations
(Sanchez-Gonzalez et al 2020)

guiding human intuition in mathematics
(Davies et al 2021)

ETA in Google Maps
(Derrow-Pinion et al, 2021)



drug interactions
(Zitnik et al 2018)



recommender systems
(Ying et al 2018)

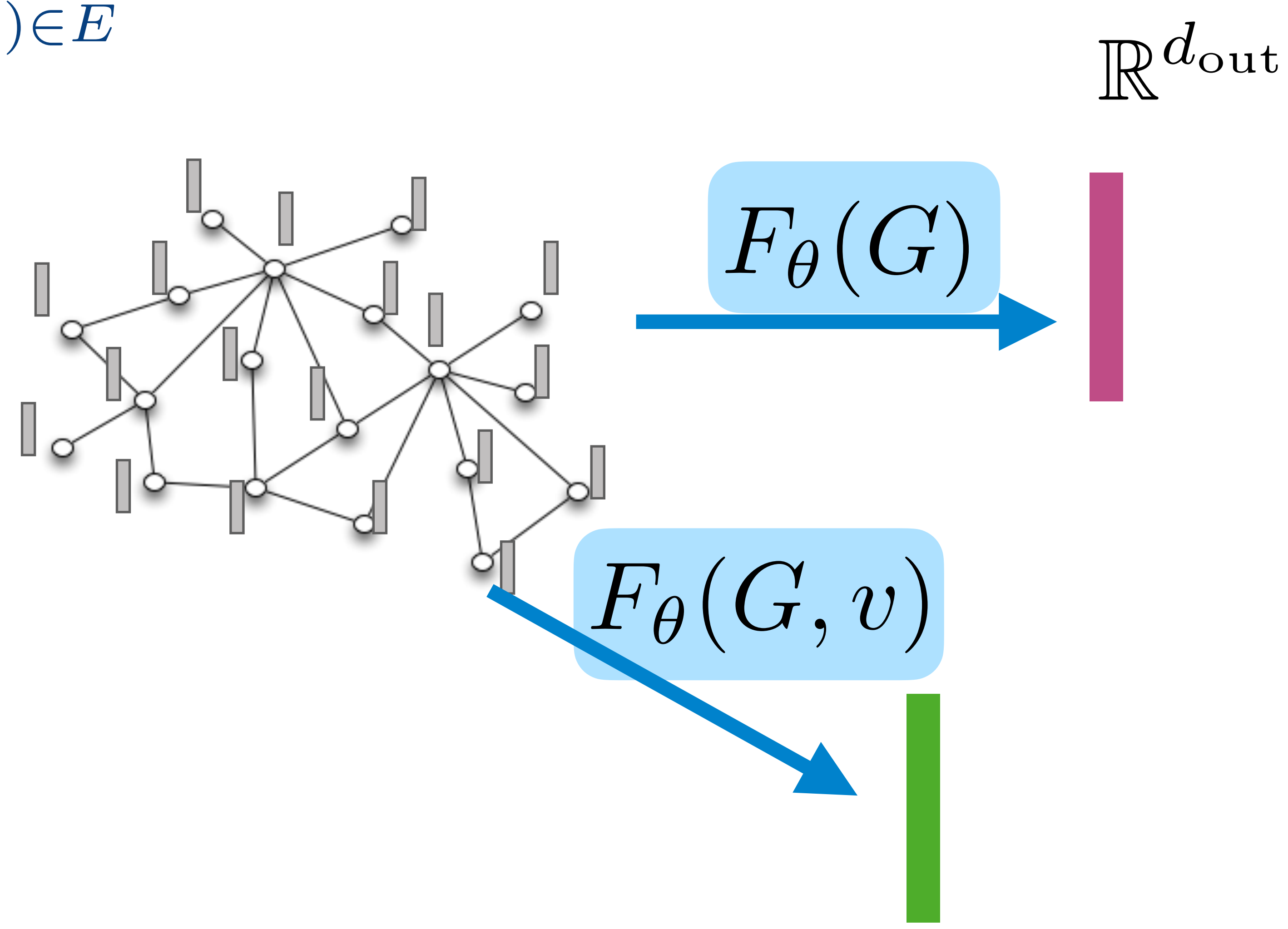
Machine Learning with Graph Data

- Data: **attributed graphs** (of bounded size)

$$G = (V, E, X, W) \in \mathcal{G}$$

$\{x_v\}_{v \in V}$ $\{w(u,v)\}_{(u,v) \in E}$

$x_v \in \mathbb{R}^d$



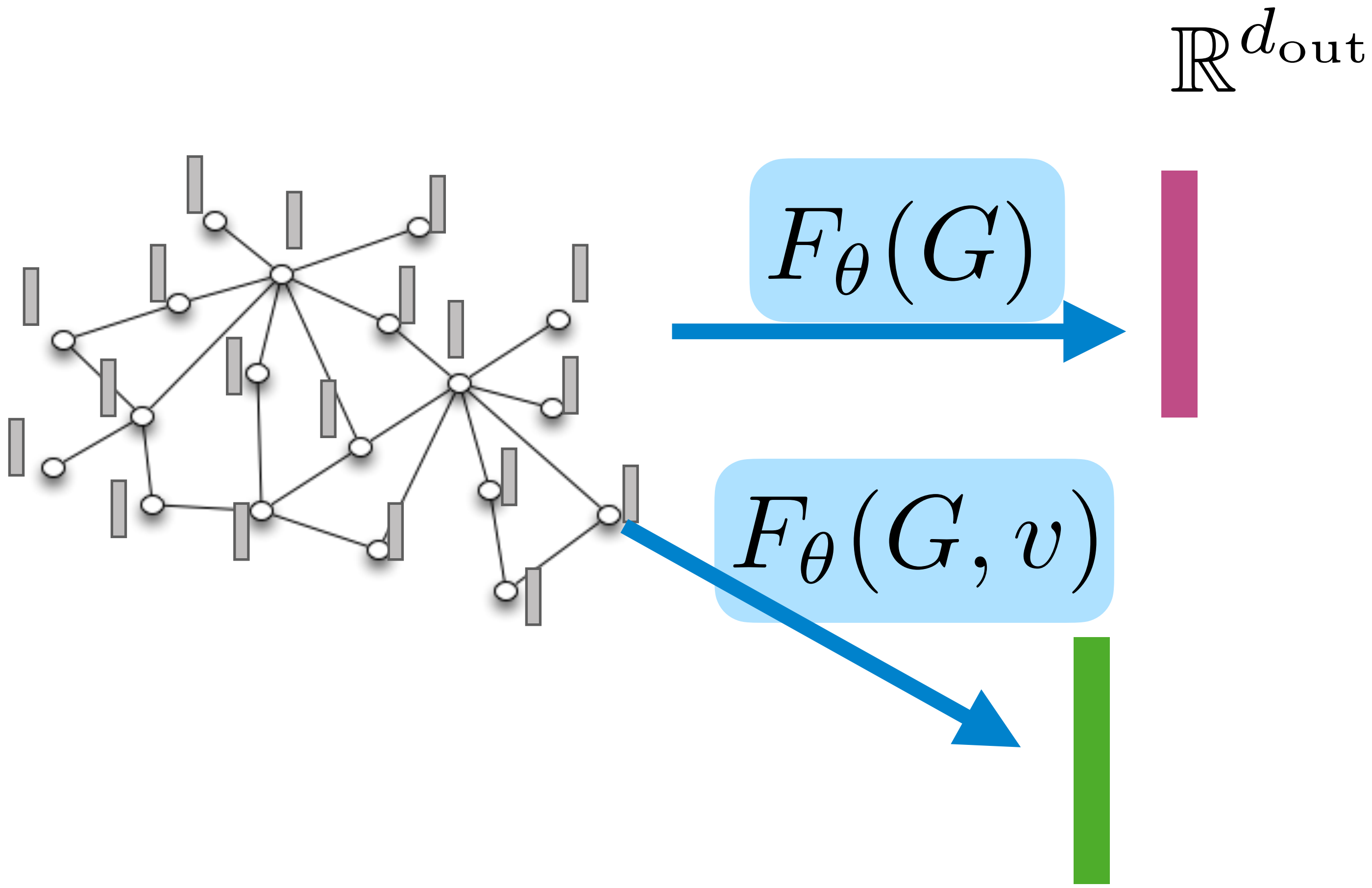
Machine Learning with Graph Data

- Data: **attributed graphs** (of bounded size)

$$G = (V, E, X, W) \ \in \mathcal{G}$$

$\{x_v\}_{v \in V}$ $\{w(u,v)\}_{(u,v)\in E}$

$x_v \in \mathbb{R}^d$

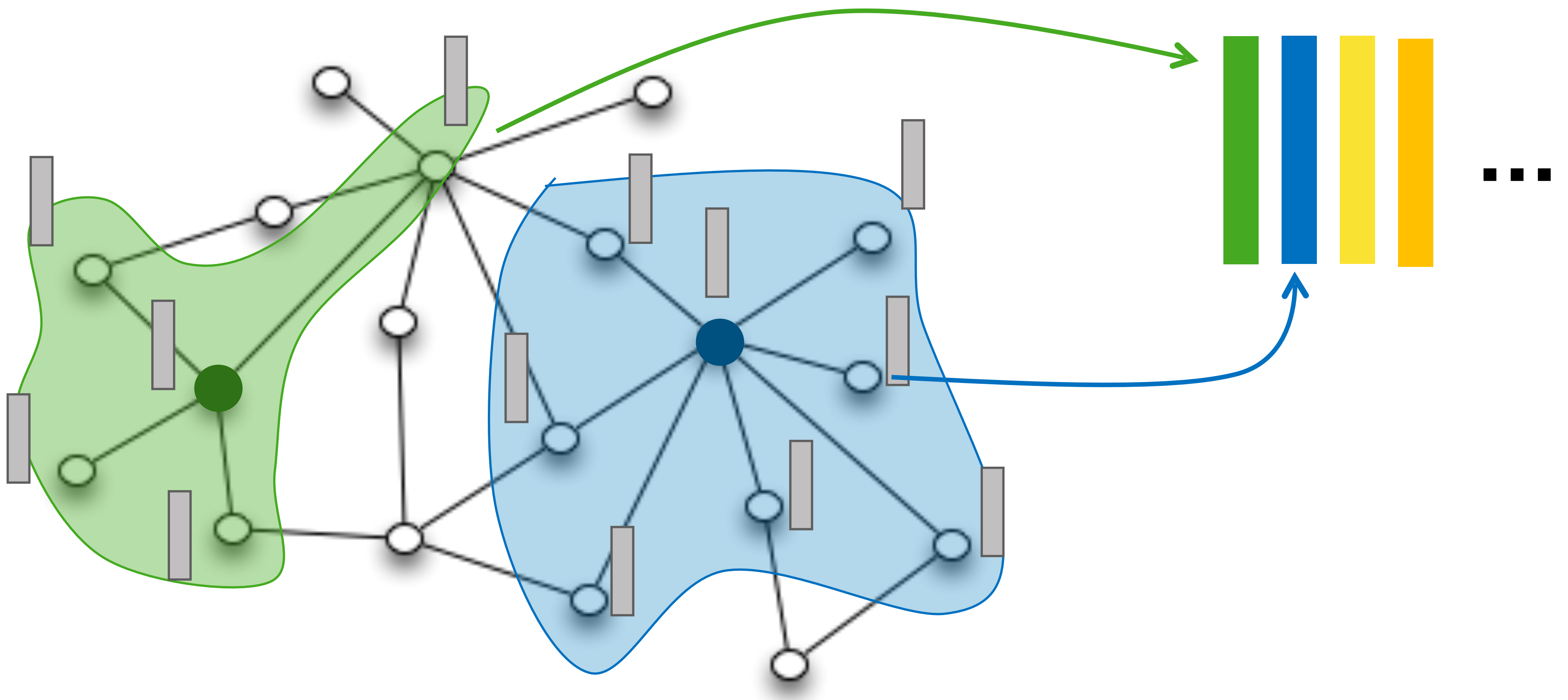


- Want: graph/node invariants

$$F_\theta(PAP^\top, PX) = F_\theta(A, X)$$ Permutation invariance

$$F_\theta(PAP^\top, PX, v) = F_\theta(A, X, v)$$ Permutation equivariance

(Message passing) Graph neural networks

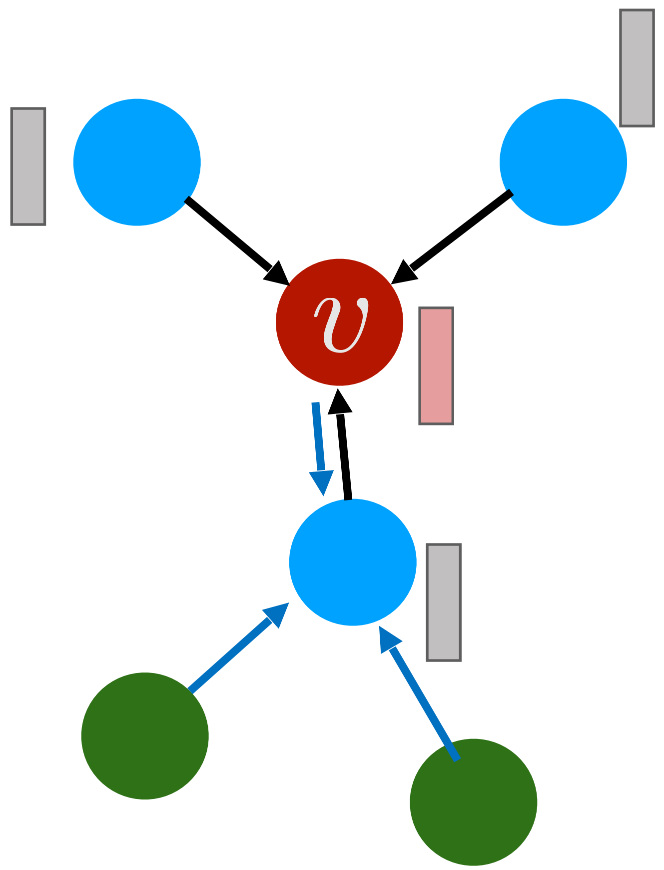


Idea:

1. Encode each node (node's neighborhood): *node embedding*
2. Aggregate *set* of node embeddings into a *graph embedding*

Merkwirth-Lengauer 05; Gori-Monfardini-Scarselli 05; Scarselli-Gori-Tsoi-Hagenbuchner-Monfardini 09; Bruna et al 14; Dai et al 16; Battaglia et al. 16; Defferrard et al. 16; Duvenaud et al. 15; Hamilton et al. 17; Kearnes et al. 16; Kipf & Welling 17; Li et al. 16; Veličković et al. 18; Verma & Zhang 18; Ying et al. 18; Zhang et al. 18; …

Node embedding: message passing



In each round *k*:

feature description of node u in round k-1

Aggregate over neighbors

$$m_{\mathcal{N}(v)}^{(k)} = \text{AGGREGATE}^{(k)}\left(\{\!\{h_u^{(k-1)} : u \in \mathcal{N}(v)\}\!\}\right)$$

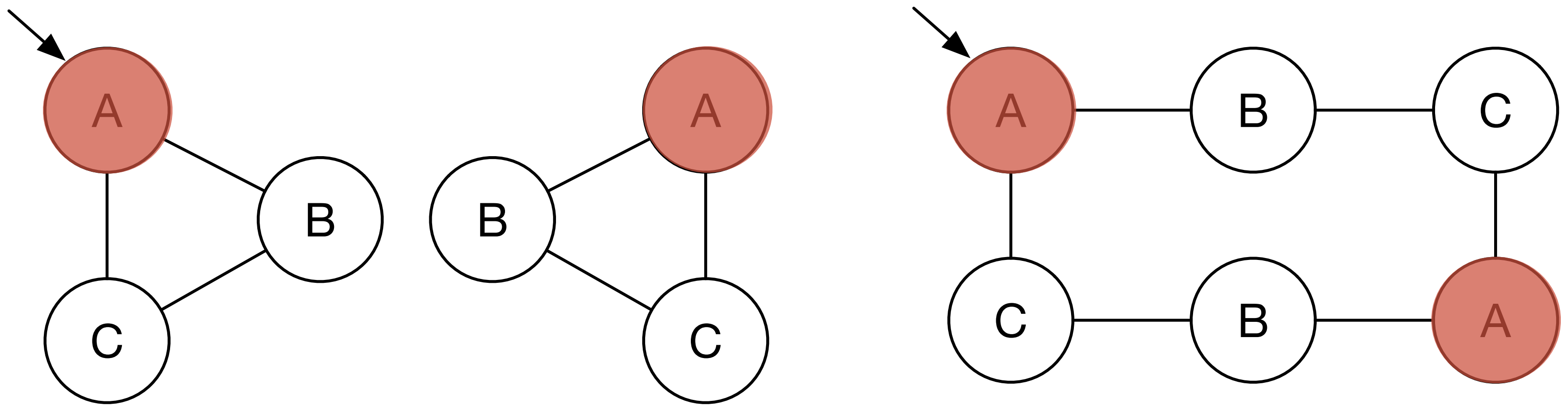
Update: Combine with current node

$$h_v^{(k)} = \text{COMBINE}^{(k)}\left(h_v^{(k-1)}, m_{\mathcal{N}(v)}^{(k)}\right)$$

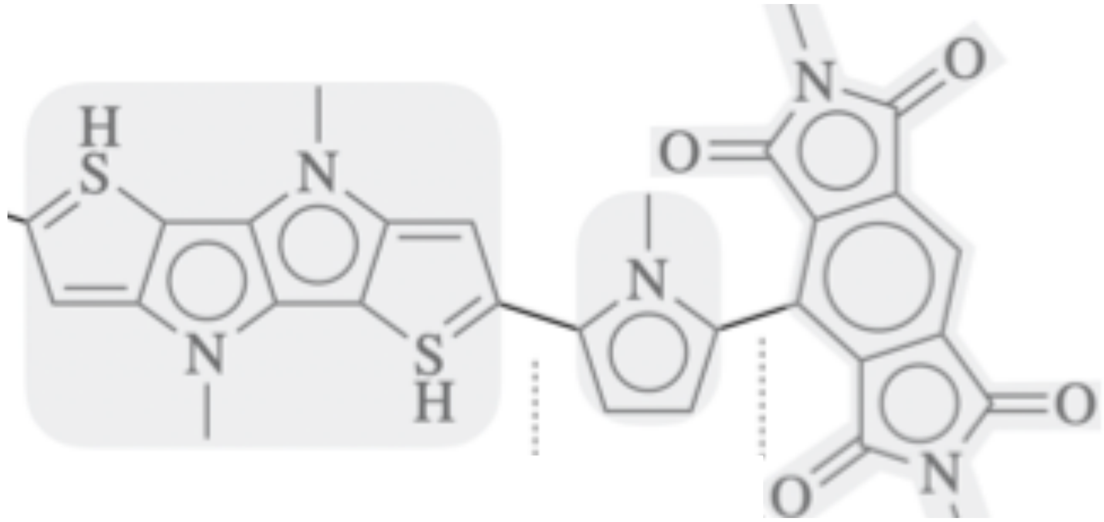
(Merkwirth & Lengauer 2005; Scarselli et al 2009; Bruna et al 2014; Dai et al 2016; Battaglia et al., 2016; Defferrard et al., 2016; Duvenaud et al., 2015; Hamilton et al., 2017; Kearnes et al., 2016; Kipf & Welling, 2017; Li et al., 2016; Velickovic et al., 2018; Verma & Zhang, 2018; Ying et al., 2018; Zhang et al., 2018; …)

Shortcomings of message passing GNNs

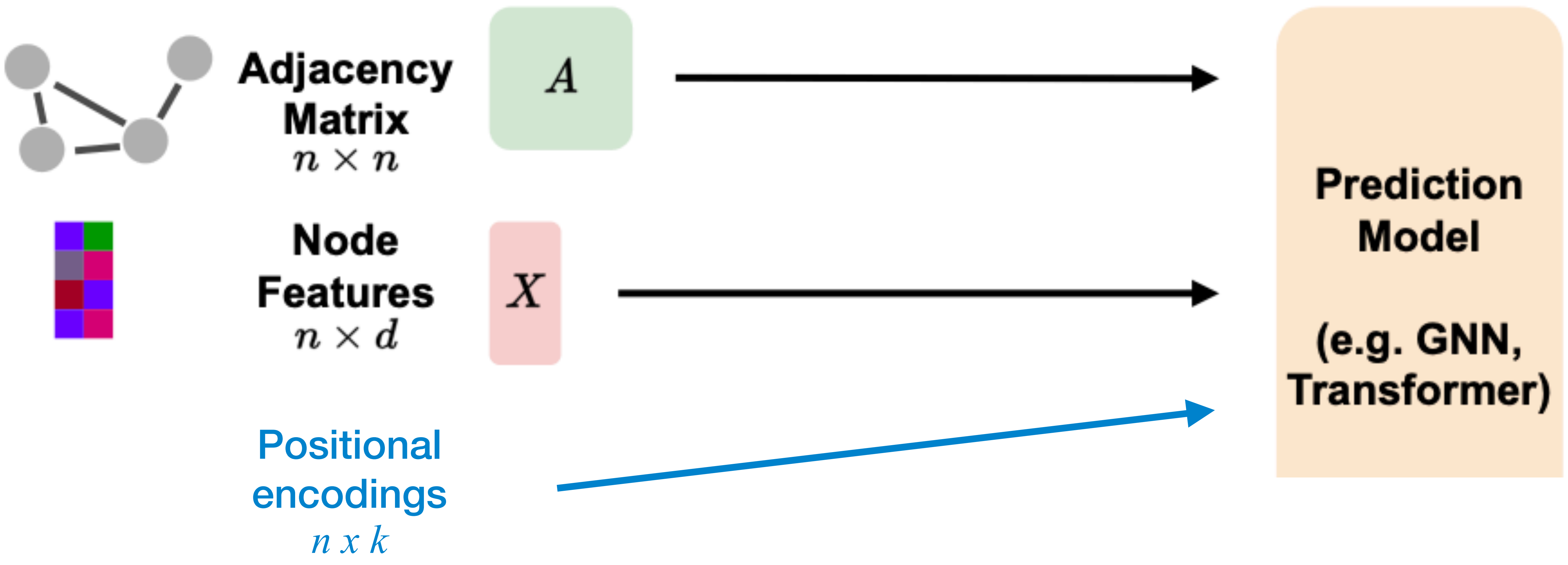
- cannot distinguish certain graphs



- => cannot express many structural functions, e.g., existence of k-cycle, diameter, etc



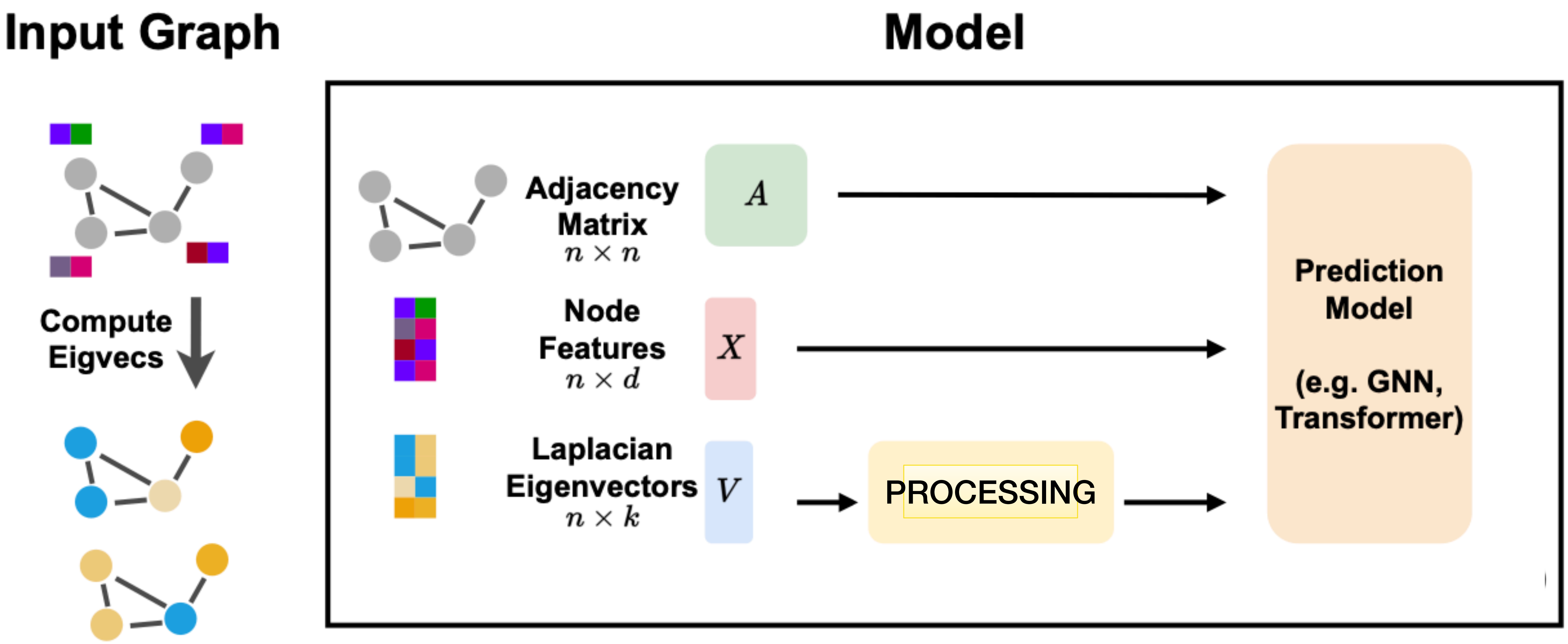
Neural Networks for Graphs



- For good performance, (often) need more information: positional encodings

(Feldman et al 2022, Dwivedi et al 2022, Kreuzer et al 2021, Dwivedi & Bresson 2021, Mialon et al 2021)

Positional encodings from eigenvectors



Positional Encodings from Laplacian eigenvectors

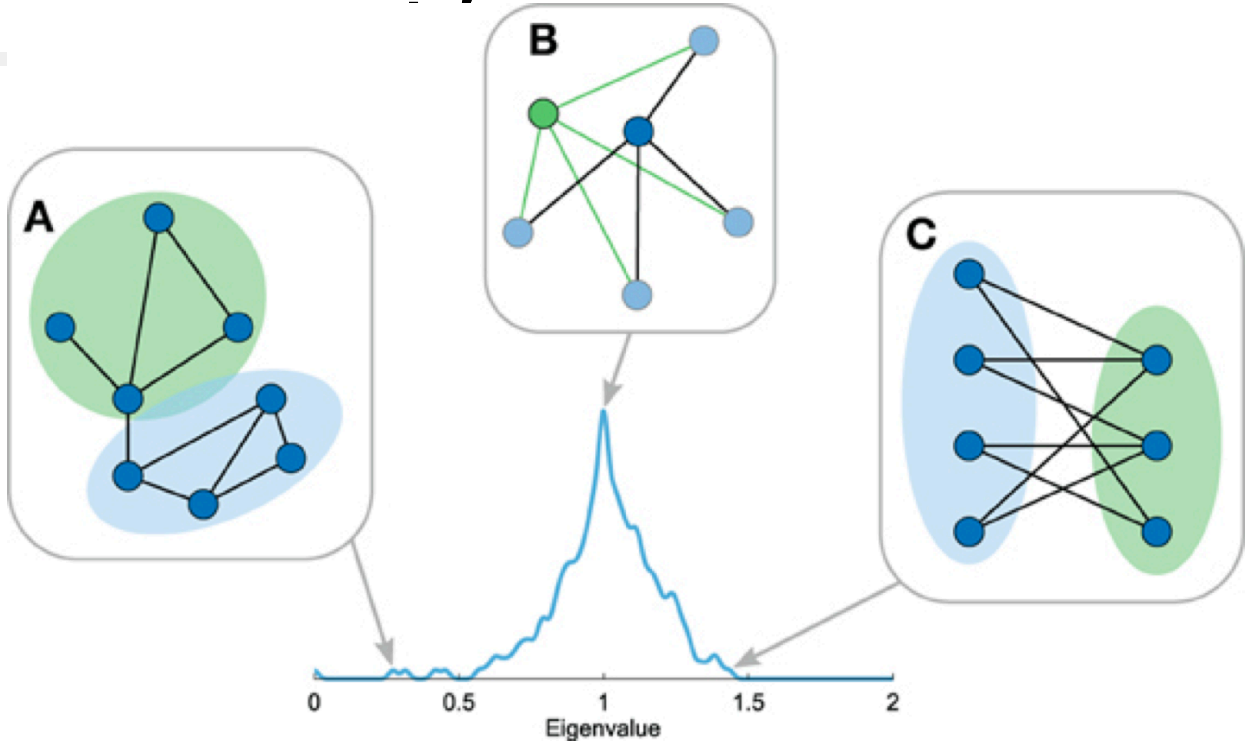
- Graph Laplacian:

$$L = I - D^{-1/2} A D^{-1/2}$$

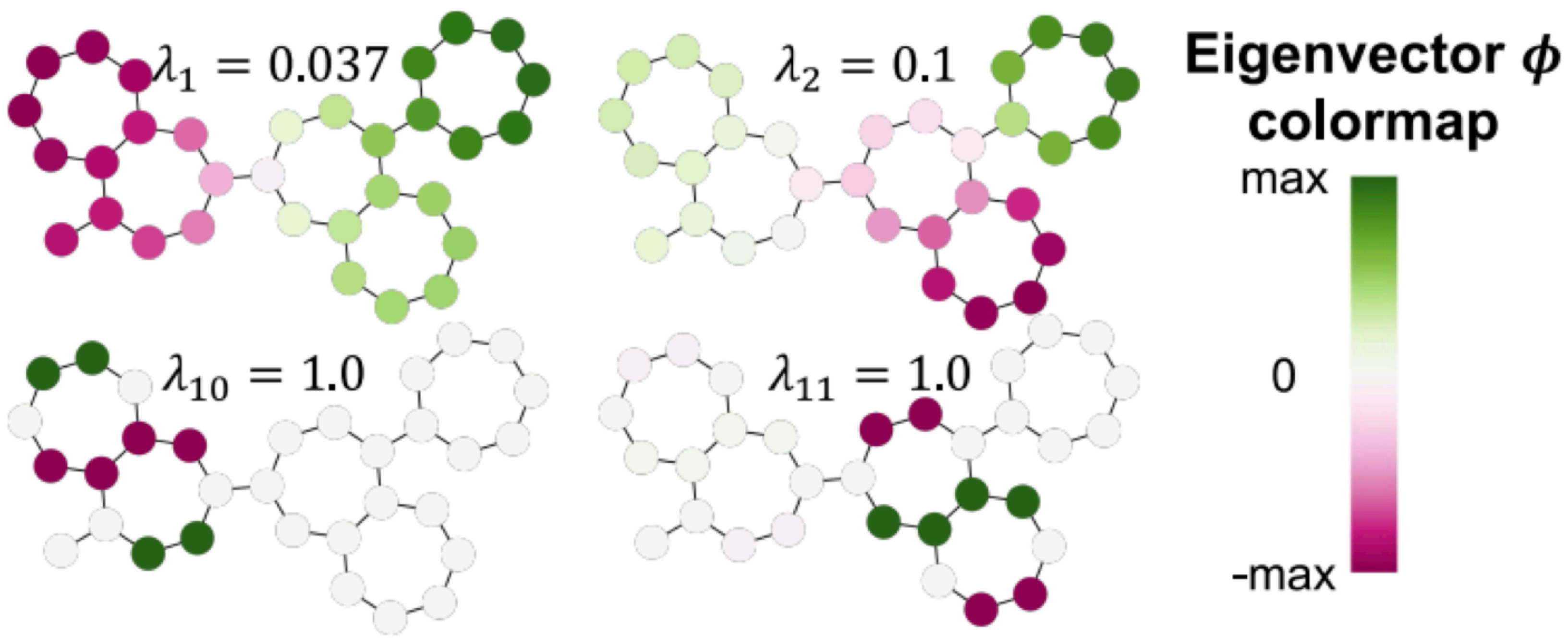
- eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$

  eigenvectors $v_1, \ldots, v_n$

- Captures distances, local structures, etc.

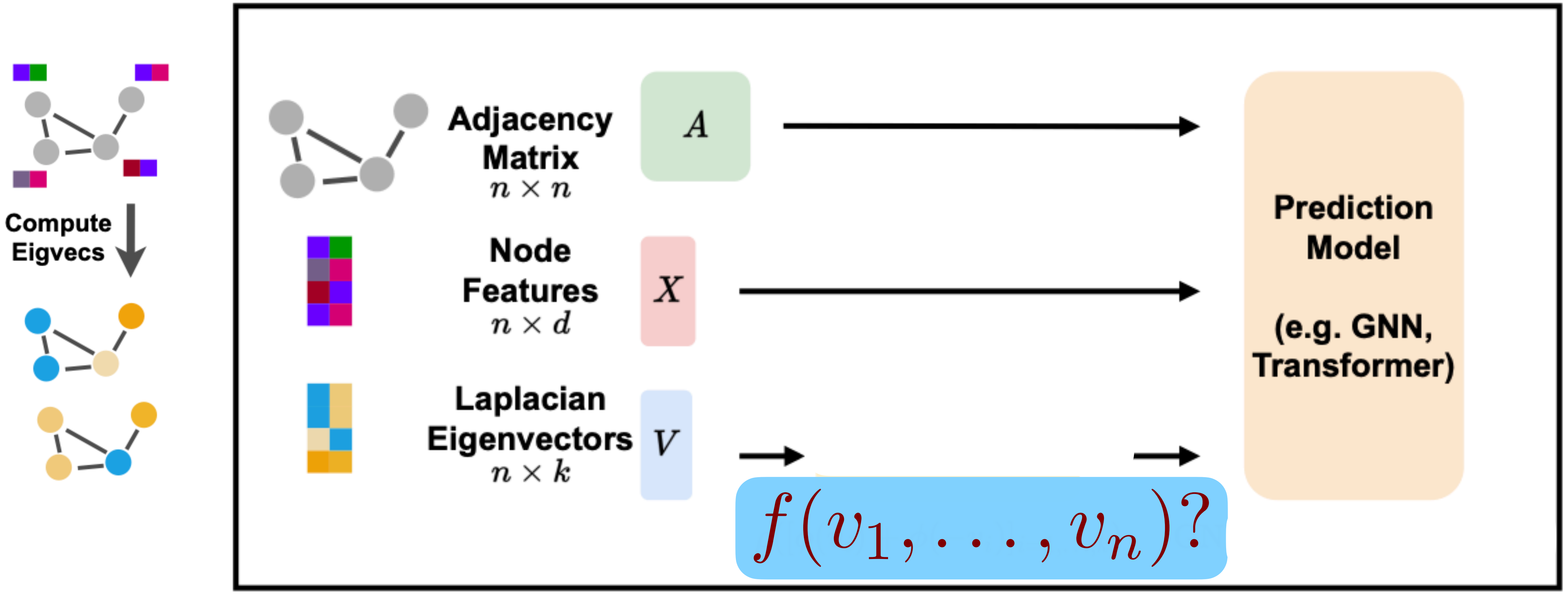


(de Lange, de Reus, van den Heuvel 2014)



(Kreuzer, Beaini, Hamilton, Létourneau, Tossou 2021)

Functions on eigenvectors



Can we learn an arbitrary function on a set of eigenvectors (and eigenvalues)?

What invariances must f have?

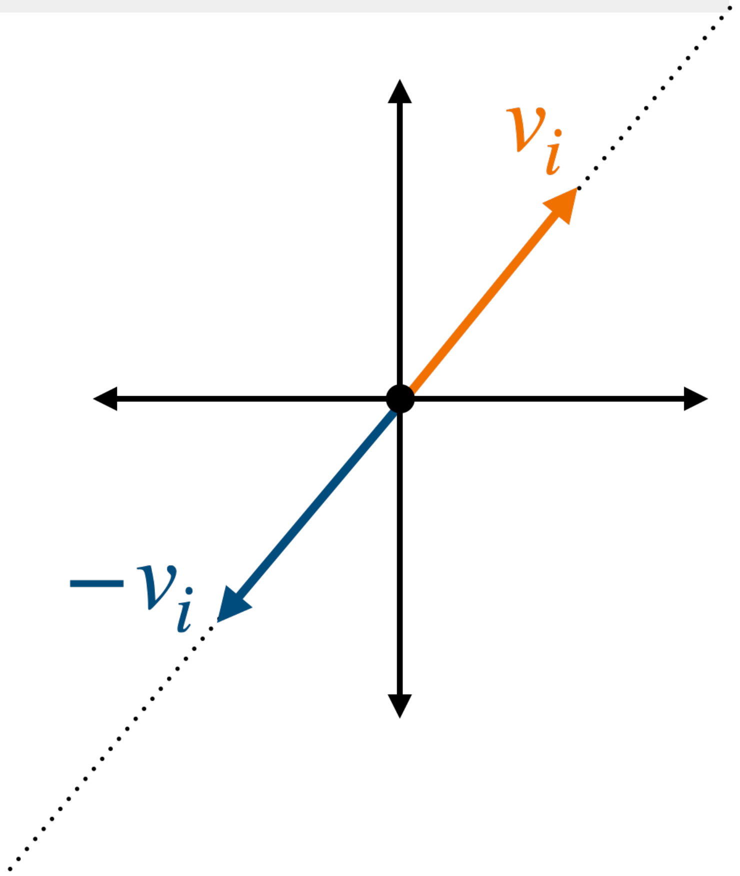
How parametrize architecture to approximate any such f?

Necessary invariances

- **Sign invariance:** with all distinct eigenvalues, $\lambda_i \neq \lambda_j, \forall i, j$

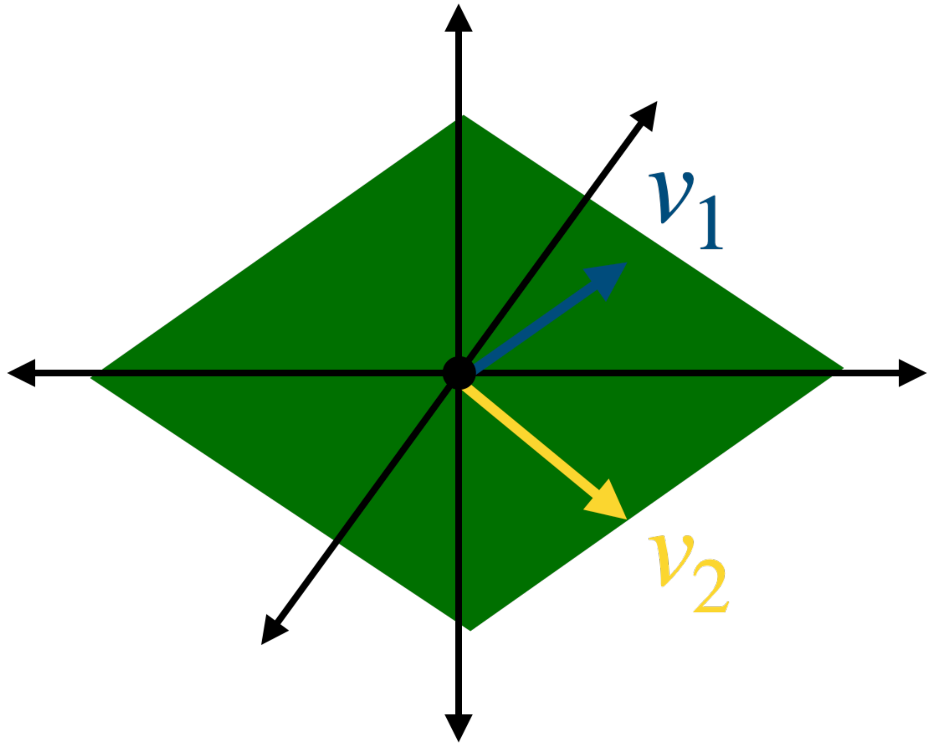
  Solver may return v_i or $-v_i$.

  $$f(v) = f(-v)$$



- **Eigenspaces:** eigenvalue multiplicities $\lambda_{i_1} = \ldots = \lambda_{i_d}$

  Any basis for *d*-dimensional eigenspace is valid.
  Multiplicities are frequent in real data!



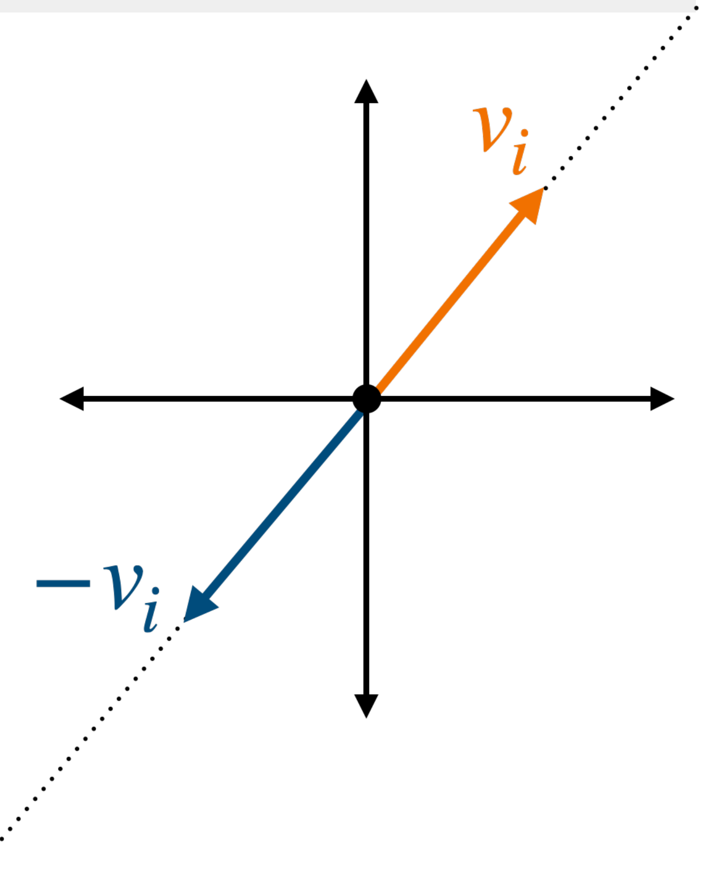
Invariances needed for generalization

Necessary invariances

- **Sign invariance:** with all distinct eigenvalues, $\lambda_i \neq \lambda_j, \forall i, j$

  Solver may return v_i or $-v_i$.

  $$f(v) = f(-v)$$

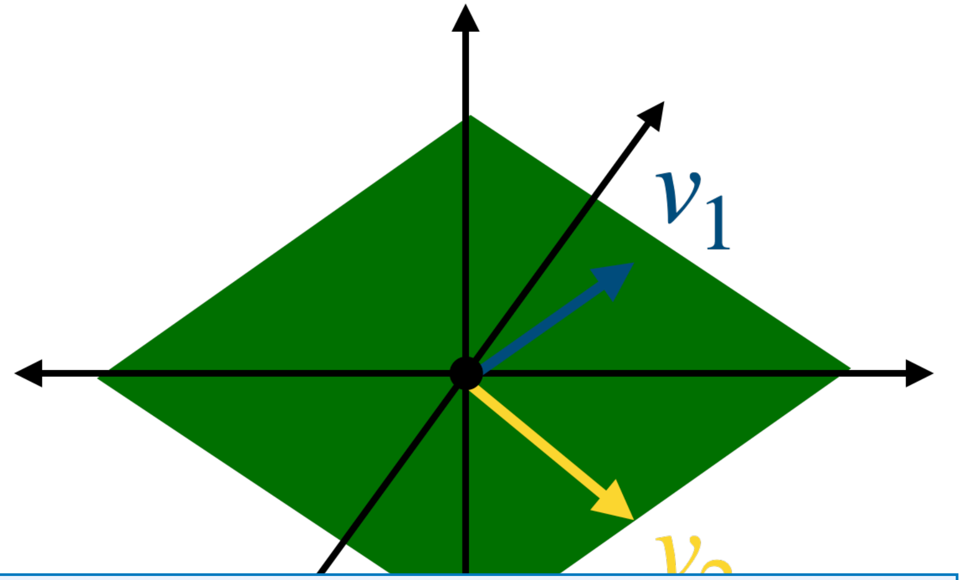


- **Eigenspaces:** eigenvalue multiplicities $\lambda_{i_1} = \ldots = \lambda_{i_d}$

  Solver may return any basis for *d*-dimensional eigenspace.

  $$f(V) = f(VQ)$$ for all Q in orthogonal group *O(d)*

  $V = [v_{i_1}, \ldots, v_{i_d}]$



+ multiple subspaces
+ permutation equivariance …

One subspace: sign invariance

Proposition

$f : \mathbb{R}^n \to \mathbb{R}$ is continuous and sign invariant if and only if $f(v) = \phi(v) + \phi(-v)$ for some continuous ϕ.

If f is also permutation equivariant, then so is ϕ .

Universal Architecture:

General f : $\phi = \mathrm{MLP}$

Permutation equivariant f: $\phi = \mathrm{DeepSets}$ *(Zaheer et al 2017, Lee et al 2019)*

One subspace: basis invariance

Proposition

If $f : \mathbb{R}^{n\times d} \to \mathbb{R}$ is continuous and $f(VQ) = f(V),\ \forall Q \in O(d)$, then $f(V) = \phi(VV^\top)$ for some continuous ϕ.

$(VQ)(VQ)^\top = V(QQ^\top)V^\top = VV^\top$

One subspace: basis invariance

Proposition

If $f : \mathbb{R}^{n \times d} \to \mathbb{R}$ is continuous and $f(VQ) = f(V), \forall Q \in O(d)$, then $f(V) = \phi(VV^\top)$ for some continuous ϕ.

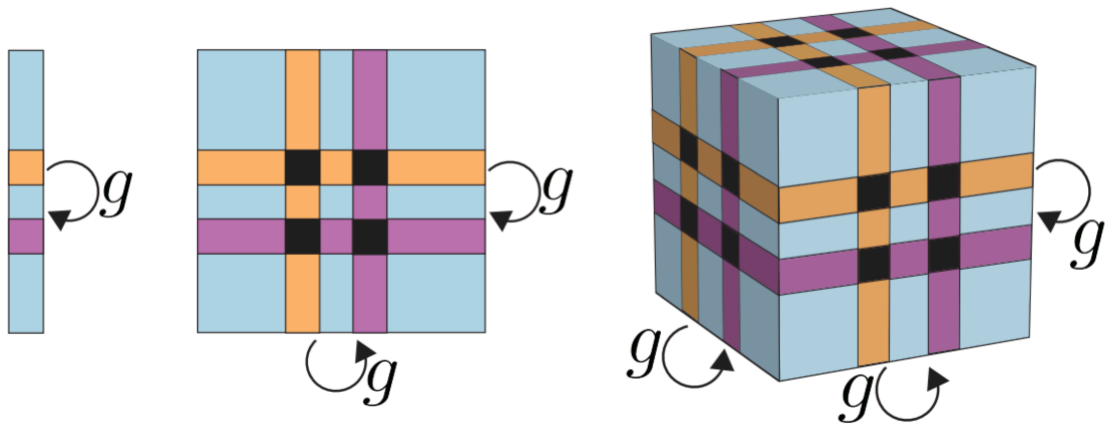
If f is also permutation equivariant, then $\phi : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ is permutation equivariant from matrices to vectors.

Universal approximation of basis-invariant functions

General f: $\phi = \mathrm{MLP}$

Permutation equivariant f: $\phi = \mathrm{IGN}$

Invariant Graph Network (Maron et al 2018)



Multiple subspaces: group invariance

- $V_1, \ldots, V_\ell$ bases of eigenspaces, $\dim V_i = d_i$

- Invariance to change of basis in each eigenspace:

$$f(V_1Q_1, \ldots V_\ell Q_\ell) = f(V_1, \ldots, V_\ell), \quad Q_i \in O(d_i)$$

  invariant to $G = O(d_1) \times \ldots \times O(d_\ell)$

- Sign invariance: $f(\pm v_1, \ldots, \pm v_\ell) = f(v_1, \ldots, v_\ell)$

Multiple subspaces: representation

$$f(V_1Q_1, \ldots V_\ell Q_\ell) = f(V_1, \ldots, V_\ell), \quad Q_i \in O(d_i)$$

Decomposition Theorem

Under mild assumptions, every continuous f that is invariant to $G_1 \times \ldots \times G_\ell$ can be written as:

$$f(x_1, \ldots, x_\ell) = \rho\left(\phi_1(x_1), \ldots, \phi_\ell(x_\ell)\right)$$

1. ϕ_i is G_i-invariant
2. If $d_i = d_j$ then can take $\phi_i = \phi_j$.

- Only need to do G_i invariance for $G_1 \times \ldots \times G_l$ invariance!!
- => Universal Approximation of invariant continuous functions.

Practical instantiations

$$f(x_1, \ldots, x_\ell) = \rho\left(\phi_1(x_1), \ldots, \phi_\ell(x_\ell)\right)$$

function on sequence / set / vector

function on set / graph nodes / vector / matrix

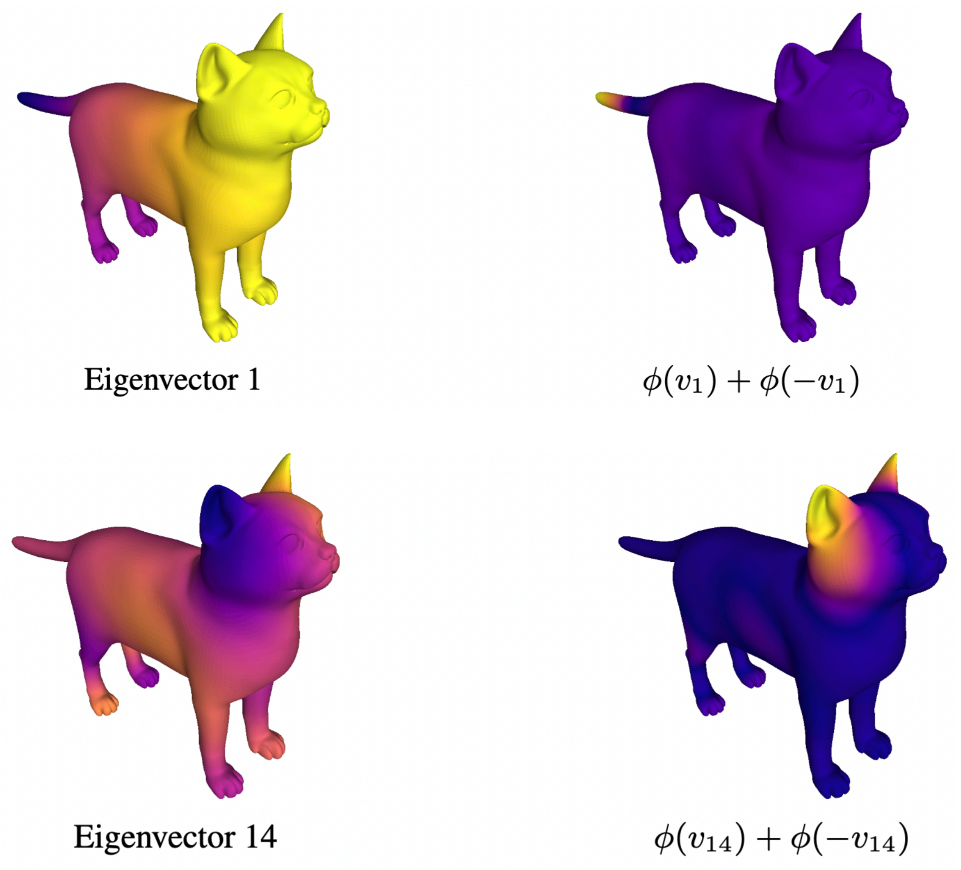
- **SignNet**: $f(v_1, \ldots, v_\ell) = \rho\big(\phi(v_1) + \phi(-v_1), \ldots, \phi(v_\ell) + \phi(-v_\ell)\big)$

  ϕ, ρ : DeepSets, Transformer, or GNN

- **BasisNet**: $f(V_1, \ldots, V_\ell) = \rho\Big([\phi_{d_i}(V_i V_i^\top)]_{i=1}^{\ell} \Big)$

  $\phi_d = \mathrm{IGN}_d$ order 2 (efficiency) or higher-order (universality)

  $\rho =$ MLP, DeepSets, Transformer



Theoretical and empirical benefits

- Can approximate many positional encodings
  based on heat kernels *(Feldman et al 2022),* random walks *(Dwivedi et al 2022),* diffusion, p-step random walks *(Mialon et al 2021),* generalized PageRank, landing probability distance encodings *(Li et al 2019)*

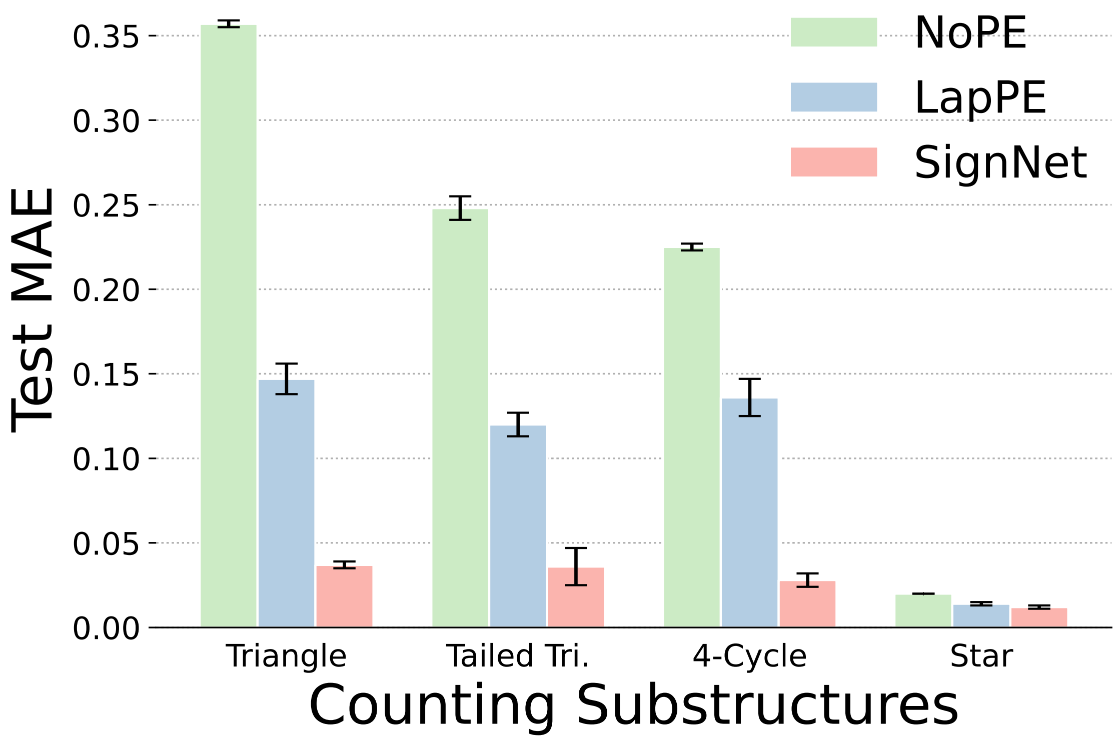
- Can approximate Spectral Graph Convolutions *(Bruna et al 2014, Defferard et al 2016, Li et al 2019)*:

$$h(V, \Lambda, X) = \sum_{i=1}^{n} p(\lambda_i) v_i v_i^\top X$$

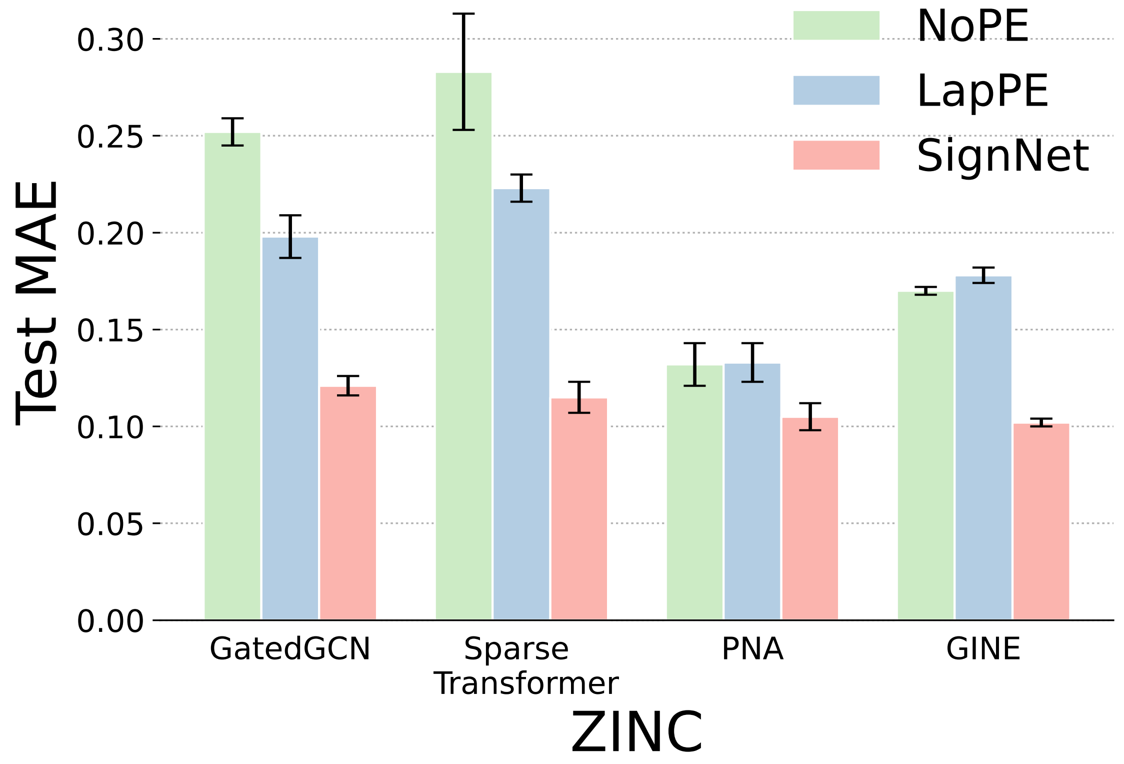
Lemma: There exist infinitely many pairs of graphs that BasisNet can distinguish, but spectral GNNs cannot.

Theoretical and empirical benefits

Lemma: can approximate spectral graph invariants *(Cvetković 1991)*, e.g. graph angles
➔ can express number of 3-, 4-, 5-cycles, connectivity, length-k closed walks.
Message passing GNNs cannot!



$\phi = \text{GIN}$ *(Xu et al 2019)* $\rho = \text{Transformer}$ *(Vaswani et al)*



ϕ = GIN *(Xu et al 2019)*, ρ = MLP

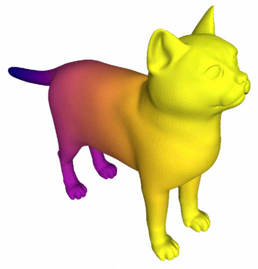
Texture reconstruction

- Neural fields on manifolds: eigenfunctions of Laplace-Beltrami operator as positional encodings

$$f(p) = \mathrm{NN}(v_1(p), \ldots, v_k(p))$$

Table 3: Test results for texture reconstruction experiment on cat and human models, following the experimental setting of (Koestler et al., 2022). We use 1023 eigenvectors of the cotangent Laplacian.

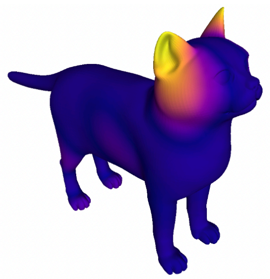
		Cat			Human		
Method	Params	PSNR ↑	DSSIM ↓	LPIPS ↓	PSNR ↑	DSSIM ↓	LPIPS ↓
Intrinsic NF	329k	34.25	.099	.189	32.29	**.119**	.330
Absolute value	329k	34.67	.106	.252	**32.42**	.132	.363
Sign flip	329k	23.15	1.28	2.35	21.52	1.05	2.71
SignNet	324k	**34.91**	**.090**	**.147**	**32.43**	.125	**.316**



Eigenvector 1

$\phi(v_1) + \phi(-v_1)$

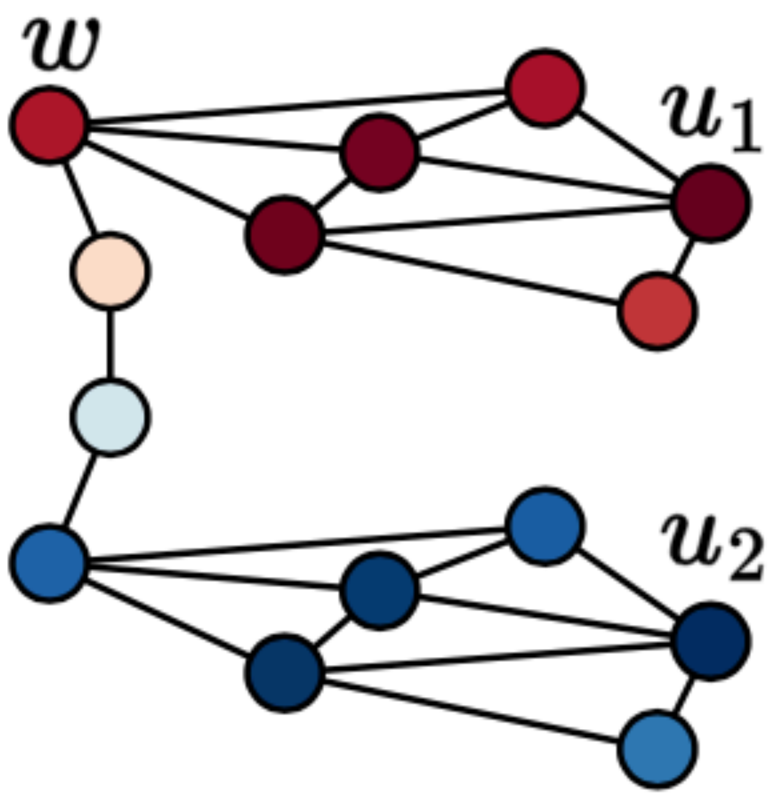
Eigenvector 14



$\phi(v_{14}) + \phi(-v_{14})$

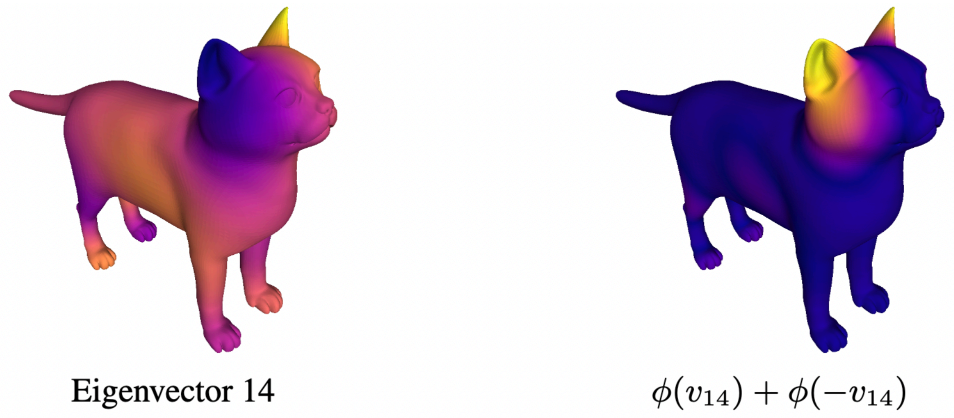
Beyond sign invariance

Beyond sign invariance



eigenvector

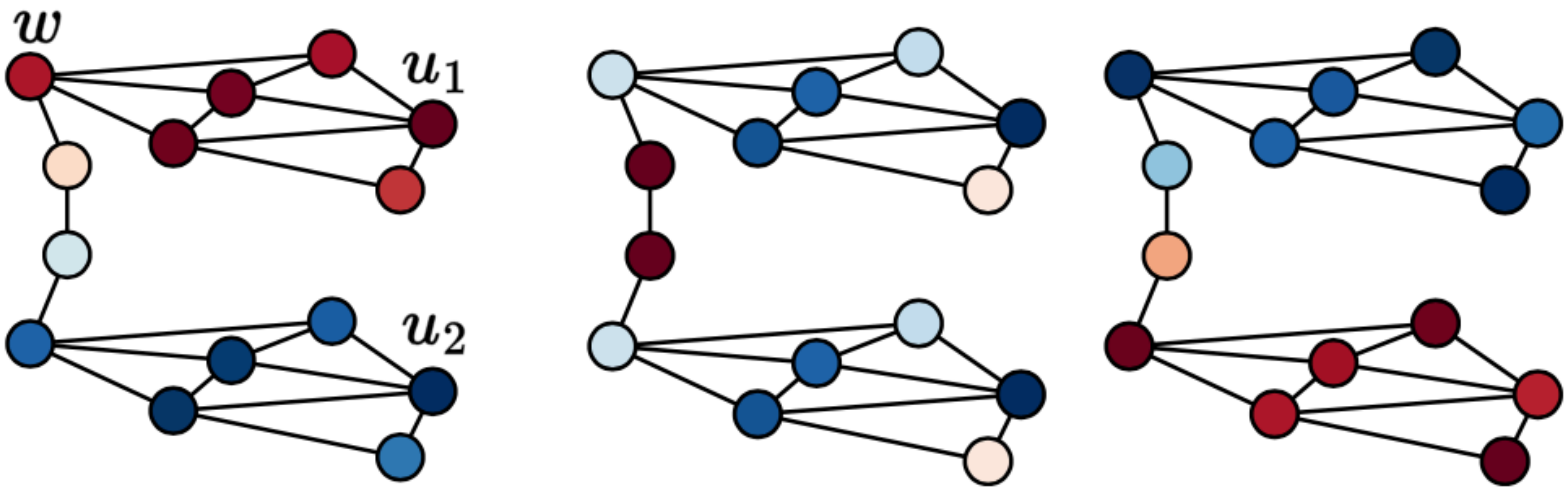
sign invariant



Example: link prediction

sign invariant representation => automorphic nodes have same representation

Beyond sign invariance



eigenvector

sign invariant

sign equivariant

Sign equivariance helps!

$$f(s_1 v_1, \ldots, s_k v_k)_{:,j} = s_j f(v_1, \ldots, v_k)_{:,j}$$

Sign equivariant network

- **Idea 1**: learn *linear* equivariant maps, interleave with equivariant nonlinearities
- **Problem:** linear equivariant maps do not allow interaction between eigenvectors:

$$W(V) = [W_1 v_1, W_2 v_2, \ldots, W_k v_k]$$

Sign equivariant network

- characterize sign equivariant polynomials

Theorem. A polynomial $p : \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ is sign equivariant iff it can be written as

$$p(V) = W^{(2)} \left((W^{(1)} V) \odot p_{\text{inv}}(V) \right)$$

linear sign equivariant

nonlinear sign invariant

Neural Network architecture: layer l has the form

$$f_l(V) = [W_1^{(l)} v_1, \ldots, W_k^{(l)} v_k] \odot \text{SignNet}_l(V)$$

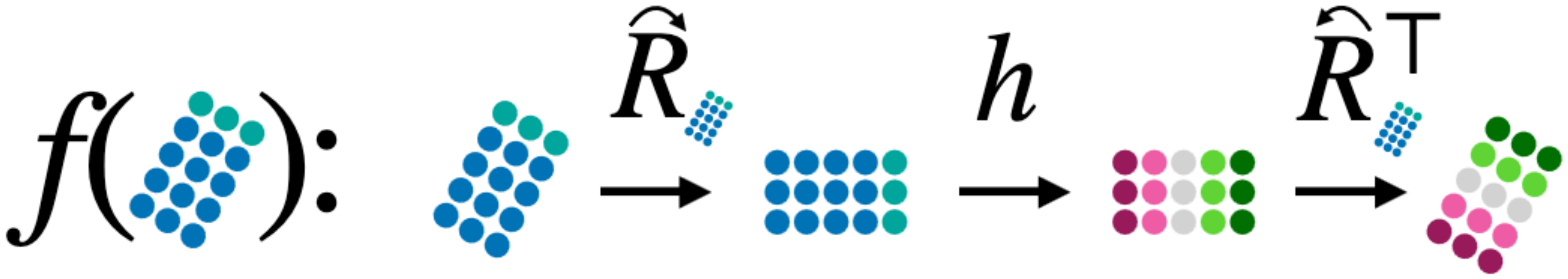
Example empirical results: link prediction

Table 2: Link prediction AUC and runtime per epoch for structural edge models.

	Erdős-Rényi		Barabási-Albert	
Model	Test AUC	Runtime (s)	Test AUC	Runtime (s)
GCN (constant input)	.497±.06	.058±.00	.705±.01	.048±.00
SignNet	.498±.00	.120±.00	.707±.00	.095±.00
$V_{i,:}^{\top} V_{j,:}$	.570±.01	.010±.01	.597±.01	.008±.00
MLP($V_{i,:} \odot V_{j,:}$)	.614±.02	.050±.00	.651±.03	.040±.00
Sign Equivariant	**.751**±.00	.063±.00	**.773**±.01	.054±.00

Another application: rotational equivariance

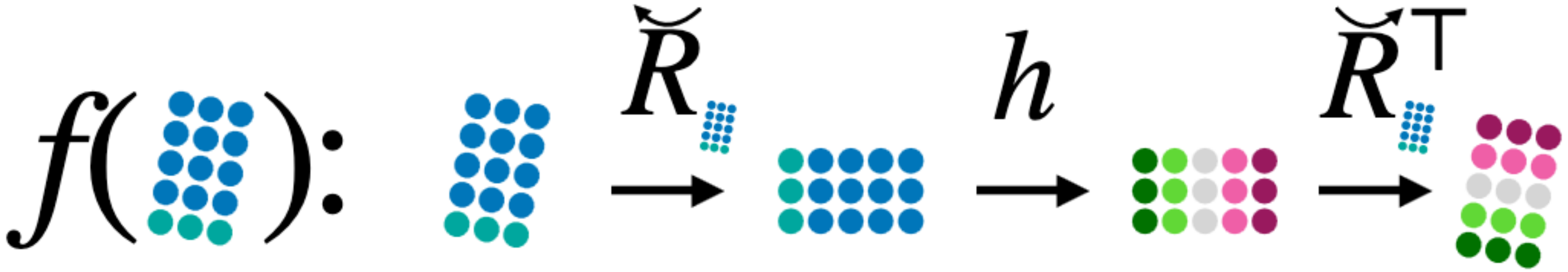
- orthogonally equivariant models, e.g. for physical systems
- "align" data with axes for canonical representation:



- But: ambiguity of eigenvector sign flips!
  Sum over all sign flips *(Puny et al., 2022, Atzmon et al., 2022, Xiao et al., 2020)*

$$f(X) = \sum_{S} h(XRS)R^\top$$

- Instead: make *h* sign equivariant
  => $f(X) = h(XR)R^\top$ orthogonally equivariant (and universal)



Outline

- A concrete example: (Graph) Neural Networks on eigenvectors
  D. Lim, J. Robinson, L. Zhao, T. Smidt, S. Sra, H. Maron, S. Jegelka. Sign and Basis Invariant Networks for Spectral Graph Representation Learning, ICLR 2023.
  D. Lim, J. Robinson, S. Jegelka, H. Maron. Expressive Sign Equivariant Networks for Spectral Geometric Learning. Neural Information Processing Systems (NeurIPS), 2023.

- Sample complexity gains from learning with invariances
  B. Tahmasebi, S. Jegelka. The Exact Sample Complexity Gain from Invariances for Kernel Regression on Manifolds, Neural Information Processing Systems (NeurIPS), 2023.

Provable gains in sample complexity?

- In practice, learning with invariant models “works better”.
- In theory, do we get gains in sample complexity?

Setting: Kernel Ridge Regression

- data $\{(x_i, y_i)\}_{i=1}^n$, on a compact boundaryless manifold $x_i \in \mathcal{M}$
- $y_i = f^*(x_i) + \epsilon_i$ with iid $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
- f^* is invariant to actions of group G
- Estimator:

$$\hat{f} = \arg\min_{f \in \mathcal{F}_{\text{inv}}} \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 + \eta \|f\|^2$$

- without invariances, excess population risk (Sobolev space, d dimensions):

$$\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right] \leq \mathcal{O}\left(\frac{C_d}{n}\right)^{s/(s+d/2)} \cdot \|f^*\|_{H^s(\mathcal{M})}^{d/(s+d/2)}$$

Previous work

in the best case, effectively, sample size n is multiplied by group size $|G|$

- finite groups, isometric actions, on spheres; result in the limit *(Bietti-Ventura-Bruna 2021)*
- Random features, different regime *(Mei-Misiakiewicz-Montanari 2021)*

- our generalization:
  - arbitrary groups — *additional* reduction for infinite groups
  - arbitrary manifold, not necessarily isometric (or linear) action

other works:
Sokolic et al 2017, Elesedy & Zaidi 2021, Elesedy 2021, Zhu et al 2021, Sannai et al 2021, Mroueh et al 2015

New results

- **Classical** bound without invariances (Sobolev space):

$$\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right] \leq O\left(C_d \sigma^2 \frac{\mathrm{vol}(\mathcal{M})}{n}\right)^{s/(s+d/2)} \cdot \|f^*\|_{H^s(\mathcal{M})}^{d/(s+d/2)}$$

- New: If functions also invariant to actions of group G:

effective dimension of quotient space, in best case $\dim(\mathcal{M}) - \dim(G)$

$$\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right] \leq O\left(C_d \sigma^2 \frac{\mathrm{vol}(\mathcal{M}/G)}{n}\right)^{s/(s+d_{\mathrm{inv}}/2)} \cdot \|f^*\|_{H^s_{\mathrm{inv}}(\mathcal{M})}^{d_{\mathrm{inv}}/(s+d_{\mathrm{inv}}/2)}$$

$\mathrm{vol}(\mathcal{M})/|G|$ for finite groups

quotient space $\mathscr{M}/G$: set of all orbits $\{g\,.\,x : g \in G\}, \forall x \in \mathscr{M}$

New results

effective dimension of quotient space, in best case $\dim(\mathcal{M}) - \dim(G)$

$$\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right] \leq O\left(C_d \sigma^2 \frac{\mathrm{vol}(\mathcal{M}/G)}{n}\right)^{s/(s+d_{\mathrm{inv}}/2)} \cdot \|f^*\|_{H^s_{\mathrm{inv}}(\mathcal{M})}^{d_{\mathrm{inv}}/(s+d_{\mathrm{inv}}/2)}$$

$\mathrm{vol}(\mathcal{M})/|G|$ **for finite groups**

- 2 ways of reducing sample complexity:
  - volume term generalizes *Bietti et al 2021:* #samples multiplied by $\frac{\mathrm{vol}(\mathcal{M})}{\mathrm{vol}(\mathcal{M}/G)}$
    $n \longrightarrow n|G|$ for finite groups
  - exponent: new, relevant for infinite groups

Minimax optimality

Theorem 1: Upper bound

$$\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right] \leq 32 \left(\frac{1}{\kappa} C_d \sigma^2 \frac{\mathrm{vol}(\mathcal{M}/G)}{n}\right)^{s/(s+d_{\mathrm{inv}}/2)} \cdot \|f^*\|_{H^s_{\mathrm{inv}}(\mathcal{M})}^{d_{\mathrm{inv}}/(s+d_{\mathrm{inv}}/2)}$$

Theorem 2: Lower bound

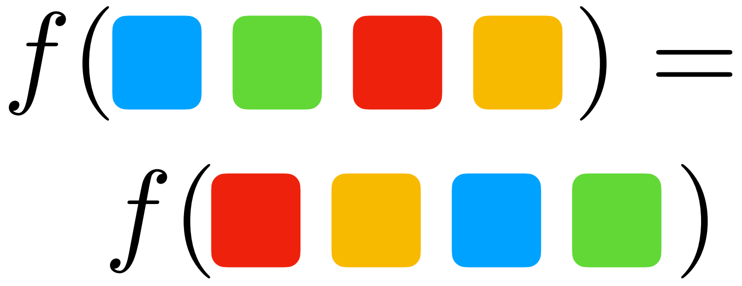
$$\inf_{\hat{f}} \sup_{f^* \in H^s_{\mathrm{inv}}(\mathcal{M}), \|f^*\|=1} \mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right] \geq C_\kappa \left(C_d \sigma^2 \frac{\mathrm{vol}(\mathcal{M}/G)}{n}\right)^{s/(s+d_{\mathrm{inv}}/2)}$$

$s = \frac{d}{2}(\kappa + 1)$

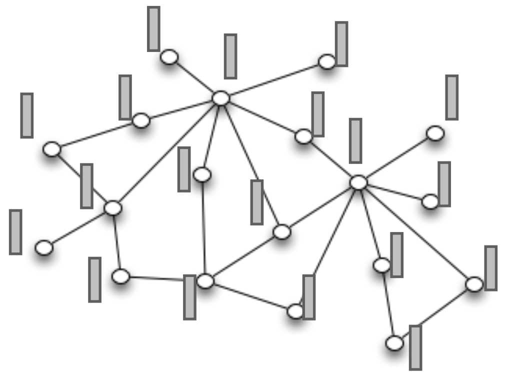
$d_{\mathrm{inv}} = \dim(\mathcal{M}/G)$

Examples

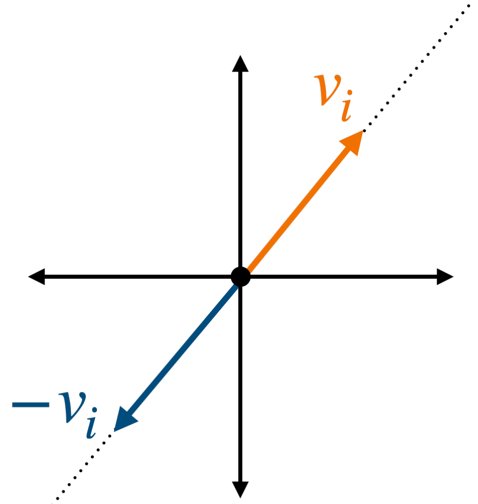
- Sets, m elements, with permutation invariance:
  effective #samples $n \times m!$



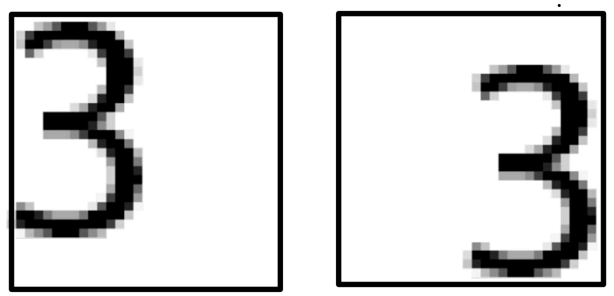
- Graphs on m nodes, with permutation invariance:
  effective #samples $n \times m!$



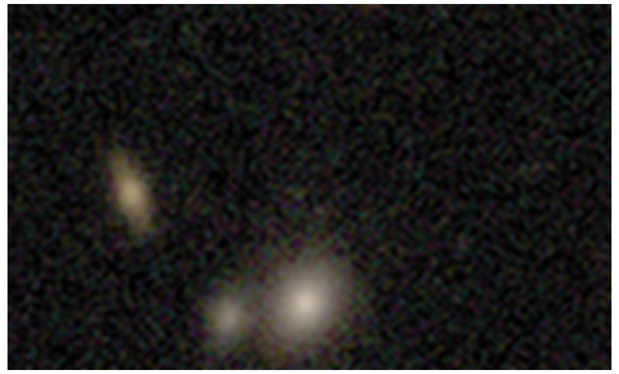
- m Eigenvectors, with invariance to sign flips (SignNet)
  effective #samples $n \times 2^m$

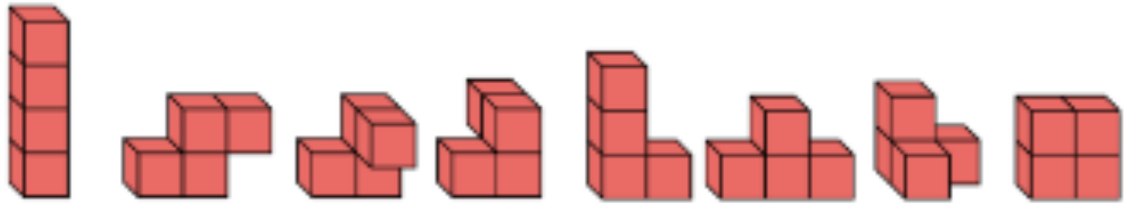


- $m \times m$ images, with 2d translation invariance
  effective #samples $n \times m^2$



Examples





- 3d point clouds, with invariance to translations, permutations, rotations

1. effective #samples $n \times m!$

2. effective dimension (exponent):

$$d = \dim(\mathcal{M}) \quad \longrightarrow \quad d_{\text{inv}} = \dim(\mathcal{M}) - \dim(G) = 3m - 6$$

Proof idea

- $\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right]$ governed by approximation error/bias and degrees of freedom/variance
- controlled by "effective" dimension of function space

- smooth invariant function $f : \mathcal{M} \to \mathbb{R}$ corresponds to smooth function $\tilde{f} : \mathcal{M}/G \to \mathbb{R}$

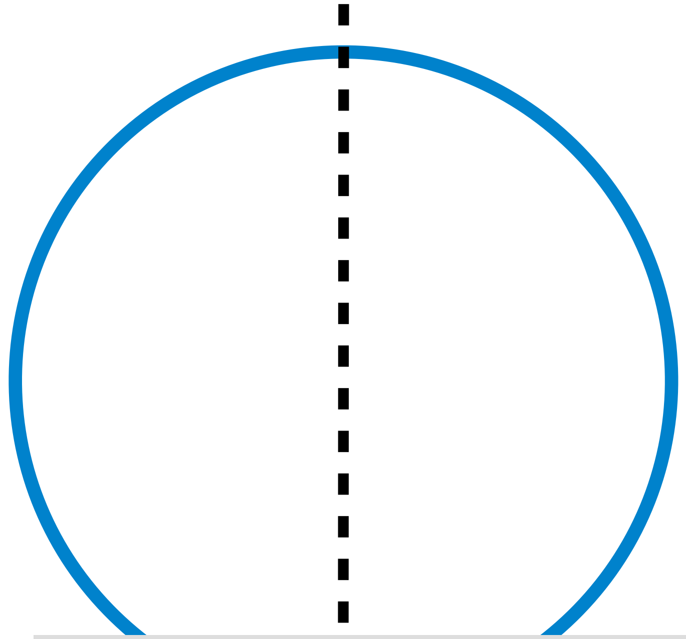
- project eigenspaces of Laplace-Beltrami operator

Proof idea

- smooth invariant function $f : \mathcal{M} \to \mathbb{R}$ corresponds to smooth function $\tilde{f} : \mathcal{M}/G \to \mathbb{R}$
- project eigenspaces of Laplace-Beltrami operator

Intuition:

- Sphere $\mathbb{S}^1$ with Fourier basis, $\sin(k\theta), \cos(k\theta), \phi_0 \equiv 1$
- G: reflection about y axis
- invariant bases: $\sin((2k+1)\theta), \cos(2k\theta), \phi_0 \equiv 1$



$\frac{N(\lambda; G)}{N(\lambda)}$ = fraction of eigenfunctions invariant to G

Proof idea

- $\mathbb{E}\left[\mathcal{R}(\hat{f}) - \mathcal{R}(f^*)\right]$ governed by approximation error/bias and degrees of freedom/variance
- controlled by “effective” dimension of function space
- e.g. variance:

dim of eigenspace of Laplace-Beltrami operator associated with eigenvalue λ

$$\mathrm{tr}\left[(\Sigma + \eta I)^{-1}\right] = \sum_{\lambda \in \mathrm{Spec}(\mathcal{M})} \sum_{\ell=1}^{\dim(V_{\lambda,G})} \frac{\mu_\lambda}{\mu_\lambda + \eta}$$

$N(\lambda)$ = #{eigenvalues < λ of Laplace-Beltrami operator}

Challenge: we need this for the quotient space $\mathscr{M}/G$...

Proof idea

- project eigenspaces of Laplace-Beltrami operator to invariant functions

Dimension Counting Theorem for invariant functions (quotient space):

Smooth connected closed Riemannian manifold $\mathcal{M}$, compact Lie group G (isometries).

Then #eigenfunctions of the Laplace-Beltrami operator with eigenvalue at most λ is

$$N(\lambda; G) = \#\{\lambda_i : \lambda_i \leq \lambda\} \approx \frac{\omega_{d_{\text{inv}}}}{(2\pi)^{d_{\text{inv}}}} \text{vol}(\mathcal{M}/G)\lambda^{d_{\text{inv}}/2}$$

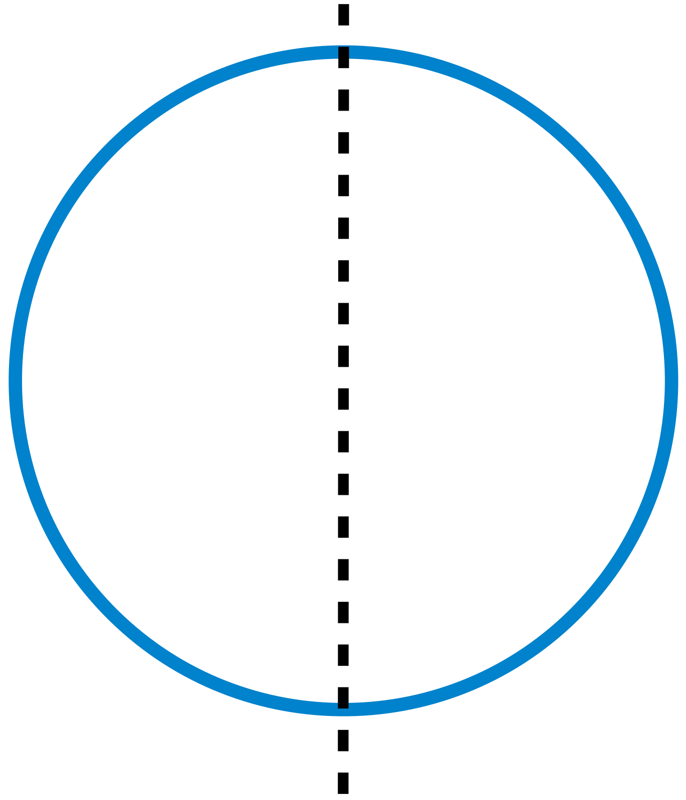
$d_{\text{inv}} = \dim(\mathcal{M}/G)$

- extends *Weyl's law* to quotient space
- BUT: it's not all that simple…

Complications

- Weyl's law does not always directly apply to quotient space $\mathcal{M}/G$
  - Quotient space $\mathcal{M}/G$ is not always a manifold
  - Principal part $\mathcal{M}_0/G$ can have a boundary

$$\mathcal{M} = \mathbb{S}^1$$
$$G = \text{reflections}$$

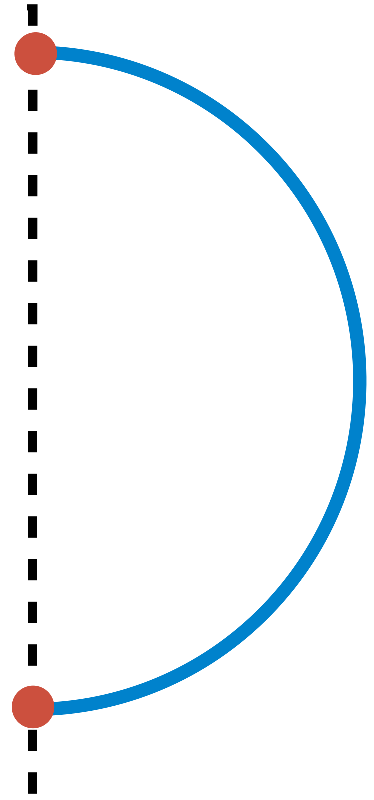


Complications

- Weyl's law does not always directly apply to quotient space $\mathcal{M}/G$
    - Quotient space $\mathcal{M}/G$ is not always a manifold
    - Principal part $\mathcal{M}_0/G$ can have a boundary

    => need boundary condition

$$\mathcal{M} = \mathbb{S}^1$$
$$G = \text{reflections}$$



Summary

- Encoding invariances into the machine learning model can have empirical and theoretical benefits
- Widely relevant
- Challenge: finding the "right" models, *proving* benefits (deep learning models?)

References:

- *D. Lim, J. Robinson, L. Zhao, T. Smidt, S. Sra, H. Maron, S. Jegelka. Sign and Basis Invariant Networks for Spectral Graph Representation Learning, ICLR 2023.*
- *D. Lim, J. Robinson, S. Jegelka, H. Maron. Expressive Sign Equivariant Networks for Spectral Geometric Learning. Neural Information Processing Systems (NeurIPS), 2023.*
- *B. Tahmasebi, S. Jegelka. The Exact Sample Complexity Gain from Invariances for Kernel Regression on Manifolds, Neural Information Processing Systems (NeurIPS), 2023.*