November 23rd
SJTU AI + Math Colloquia, Shanghai

Weisfeiler and Lehman Go Cellular: CW Networks

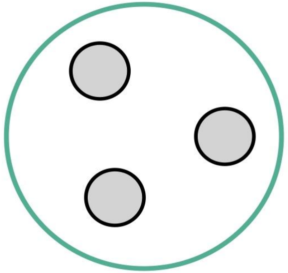
NeurIPS 2021

Cristian Bodnar[1,*] Fabrizio Frasca[2,3,*] Nina Otter[7]
Yu Guang Wang[4,5,6] Pietro Liò[1] Guido Montúfar[4,7] Michael Bronstein[2,3]

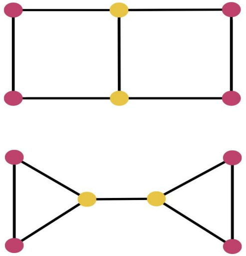
University of Cambridge[1] [2] Imperial College London[3] Max-Planck-Gesellschaft[4] 上海交通大學 Shanghai Jiao Tong University[5] UNSW Sydney[6] UCLA[7] erc

Motivation: Limitations of GNNs

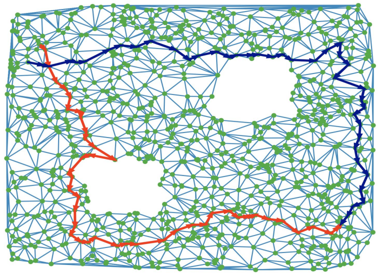
Groupwise interactions



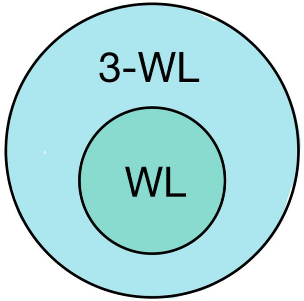
Higher-order structures



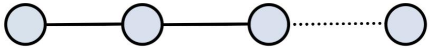
Higher-order signals



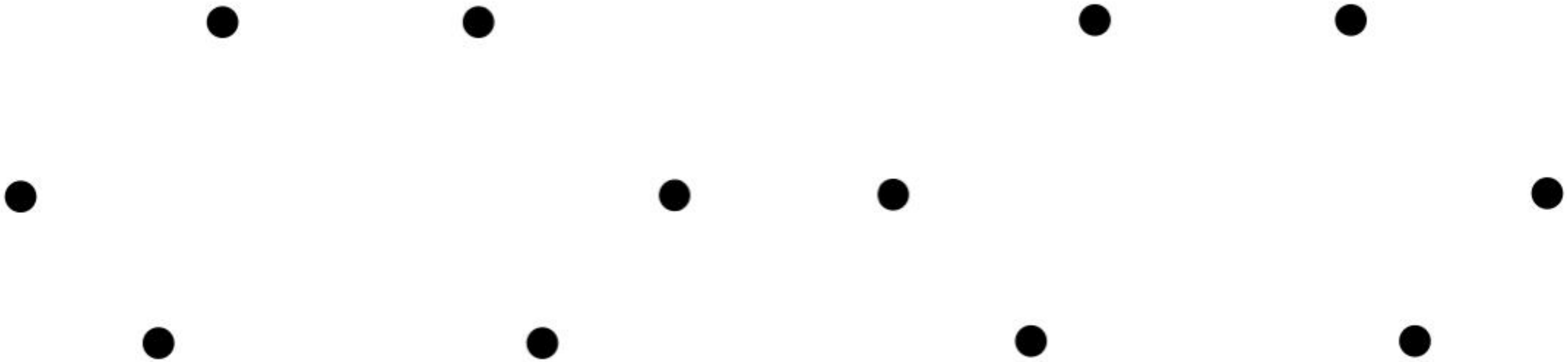
Expressive power



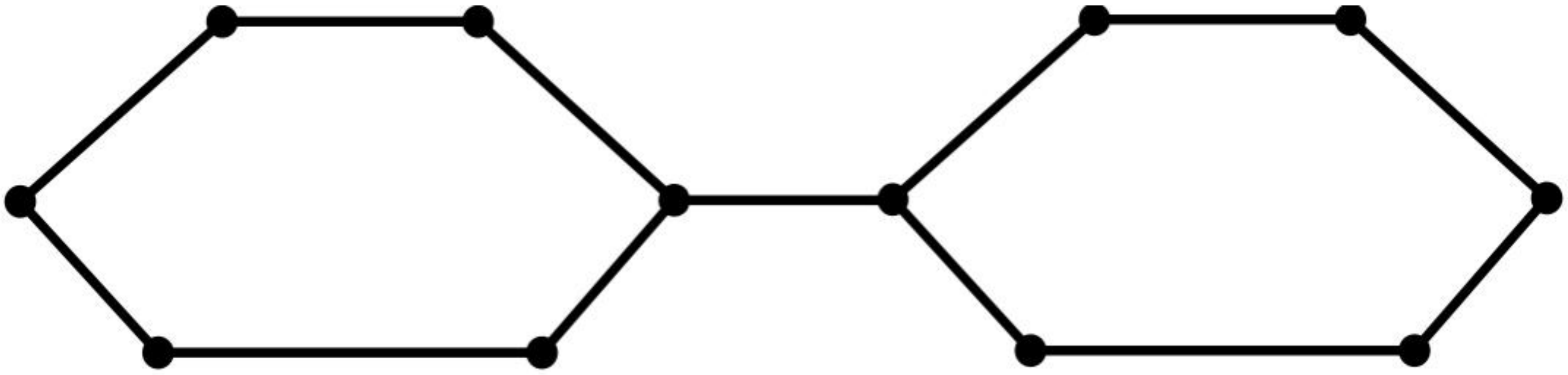
Long-range interactions



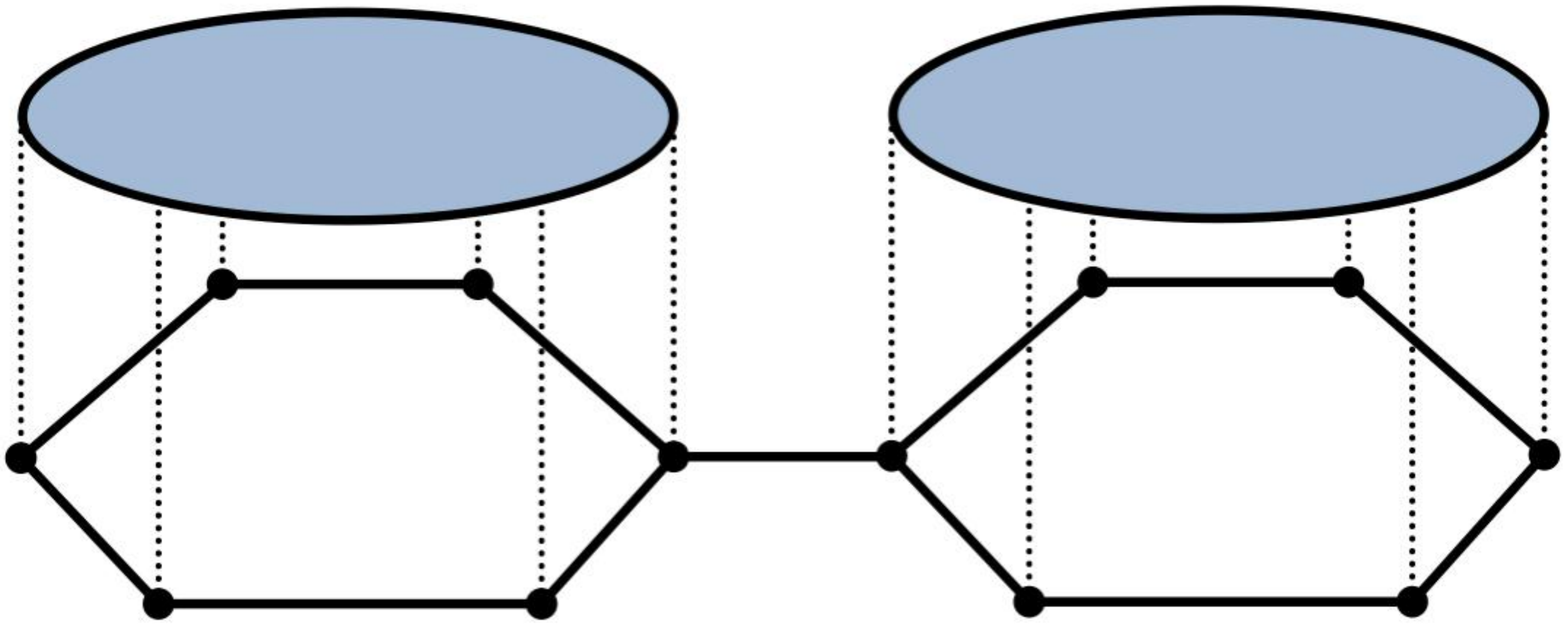
CW Complexes



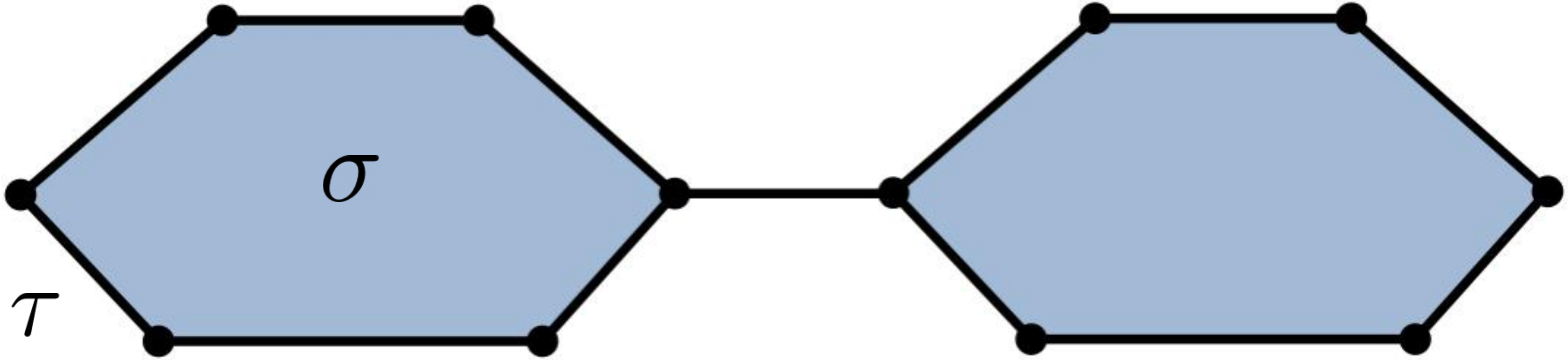
CW Complexes



CW Complexes

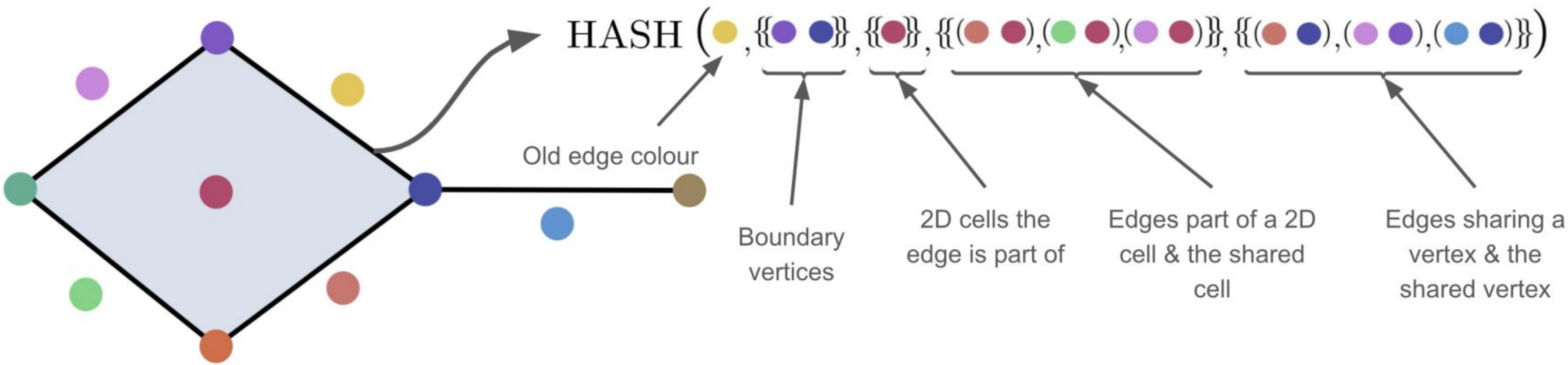


CW Complexes

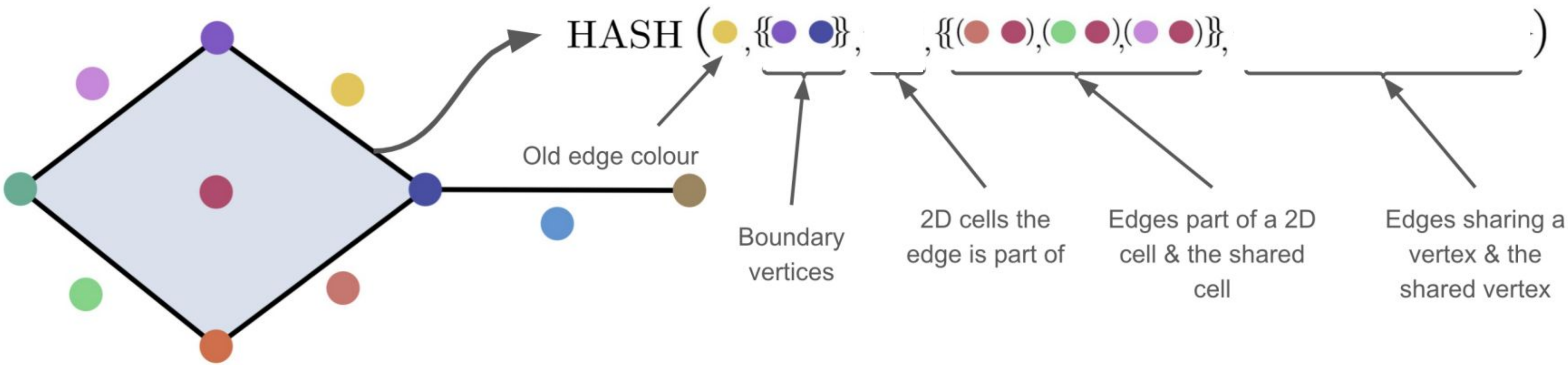


$\tau \leq \sigma \Leftrightarrow$ Cell τ is on the boundary of cell σ

Cellular Weisfeiler-Lehman

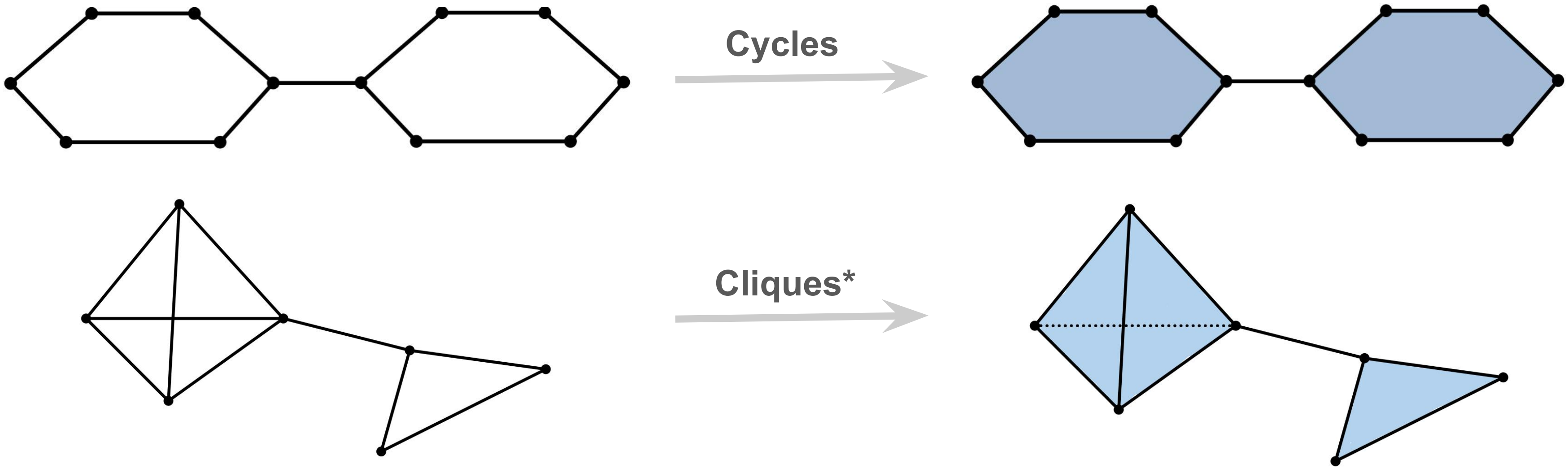


Cellular Weisfeiler-Lehman



Theorem 7. *CWL without coboundary and lower-adjacencies has the same expressive power in distinguishing non-isomorphic cell complexes as CWL with the complete set of adjacencies.*

Cellular Lifting Maps



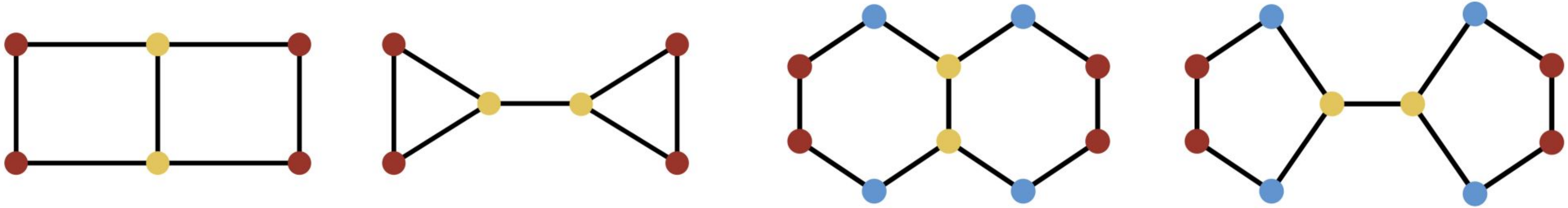
Theorem 13. *Let f be a skeleton-preserving lifting map. Then CWL(f) (i.e. CWL using lifting f) is at least as powerful as WL in distinguishing non-isomorphic graphs.*

*Studied in **arxiv:2103.03212: Weisfeiler and Lehman Go Topological (ICML 2021)**

Cellular Lifting Maps

Definition 14. *Let* k-CL*,* k-IC*,* k-C *be the lifting maps attaching cells to all the cliques, induced cycles and simple cycles, respectively, of size at most* k*.*

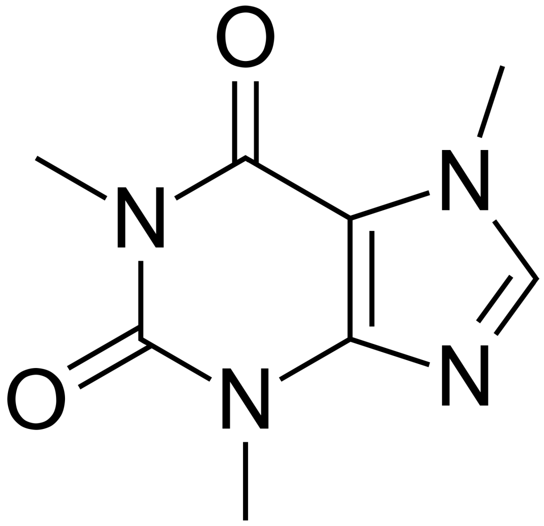
Corollary 15. *For all* $k \geq 3$*, CWL(*k-CL*), CWL(*k-IC*) and CWL*$(k\text{-C})$ *are strictly more powerful than WL.*

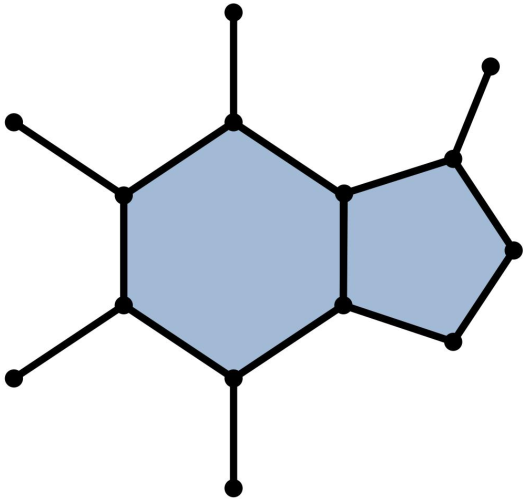


Theorem 16. *There exists a pair of graphs indistinguishable by 3-WL but distinguishable by CWL(*k*-*CL*) with* $k \geq 4$*, CWL(*k-IC*) with* $k \geq 4$ *and CWL*$(k\text{-C})$ *with* $k \geq 8$*.*

Molecules as Cell Complexes

Graph representations of molecules date back to the nineteenth century [1]. However, it is not necessarily the best representation. We propose modelling **molecules as cell complexes**.





[1]: Danail Bonchev, D.H. Rouvray (eds.) (1991) "Chemical Graph Theory: Introduction and Fundamentals"

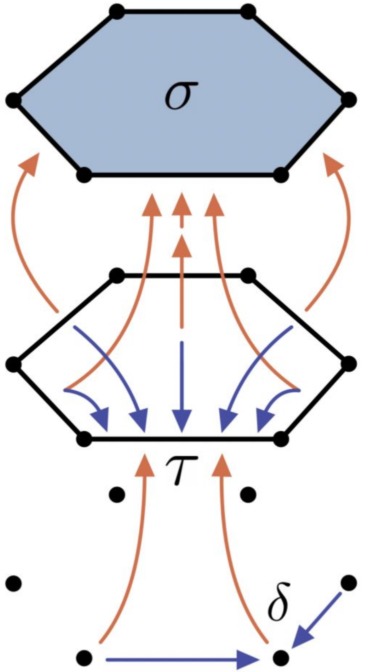
CWNs and Molecular Message Passing

$$m_{\mathcal{B}}^{t+1}(\sigma) = \mathrm{AGG}_{\tau \in \mathcal{B}(\sigma)}\Big(M_{\mathcal{B}}\big(h_\sigma^t, h_\tau^t\big)\Big)$$

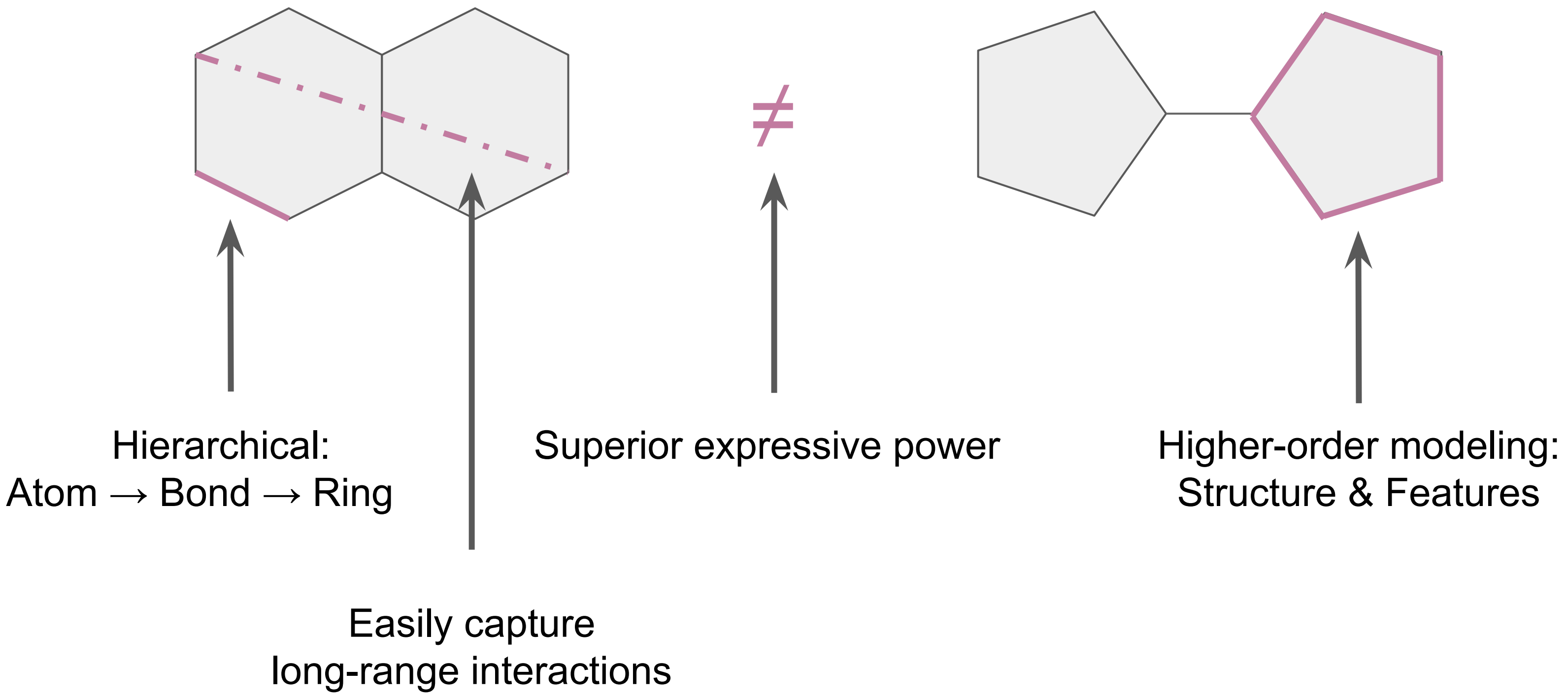
$$m_{\uparrow}^{t+1}(\sigma) = \mathrm{AGG}_{\tau \in \mathcal{N}_\uparrow(\sigma), \delta \in \mathcal{C}(\sigma,\tau)}\Big(M_{\uparrow}\big(h_\sigma^t, h_\tau^t, h_\delta^t\big)\Big)$$

$$h_\sigma^{t+1} = U\Big(h_\sigma^t, m_{\mathcal{B}}^t(\sigma), m_{\uparrow}^{t+1}(\sigma)\Big)$$

$$h_X = \mathrm{READOUT}(\{\!\!\{h_\sigma^L\}\!\!\}_{dim(\sigma)=0}, \{\!\!\{h_\sigma^L\}\!\!\}_{dim(\sigma)=1}, \{\!\!\{h_\sigma^L\}\!\!\}_{dim(\sigma)=2})$$



Practical Benefits of CWNs



Computational Complexity

(Message passing)

In practice: B_p fixed (small) constant

$$\mathcal{O}\Big(\sum_{p=1}^{d} \underbrace{\overbrace{B_p}^{\text{Max boundary size}} S_p}_{\text{Boundary msgs}} + \underbrace{2 * \binom{B_p}{2} \overbrace{S_p}^{\text{Number of p-cells}}}_{\text{Upper msgs}}\Big) \dashrightarrow \Theta\Big(\sum_{p=1}^{d} S_p\Big)$$

Empirically, GIN is only 19% – 35% faster at inference time (ZINC)

(Lifting)

linear trend →

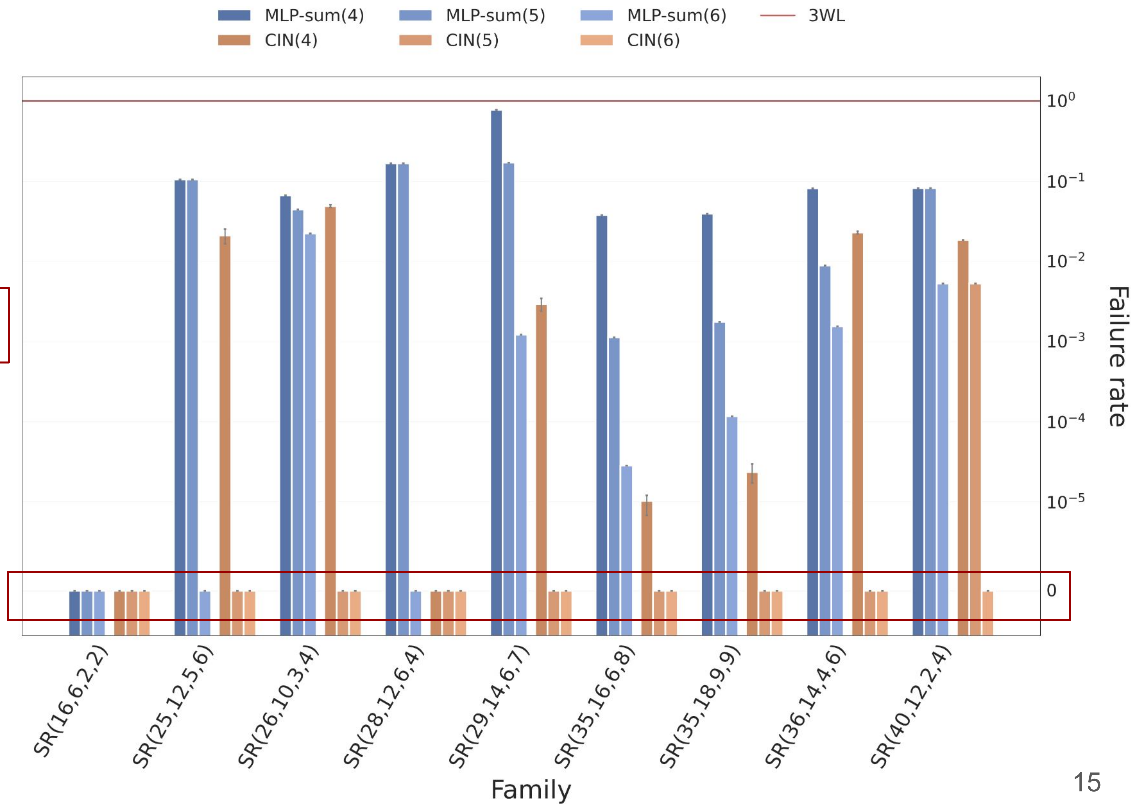
Dataset ↓ / Processes →	Seq.	2	4	8	16	32
ZINC (12k)	320.27 ± 0.54	169.95 ± 0.32	84.90 ± 0.21	43.38 ± 0.07	23.17 ± 0.68	18.59 ± 0.68
Mol-HIV (41k)	1178.98 ± 3.90	635.58 ± 0.83	319.01 ± 0.40	164.26 ± 0.52	86.92 ± 0.77	60.62 ± 2.05
ZINC-FULL (250k)	6805.35 ± 16.50	3549.16 ± 7.73	1782.41 ± 3.84	918.38 ± 3.46	492.77 ± 6.13	383.92 ± 3.30

Synthetic Experiments: Expressive Power

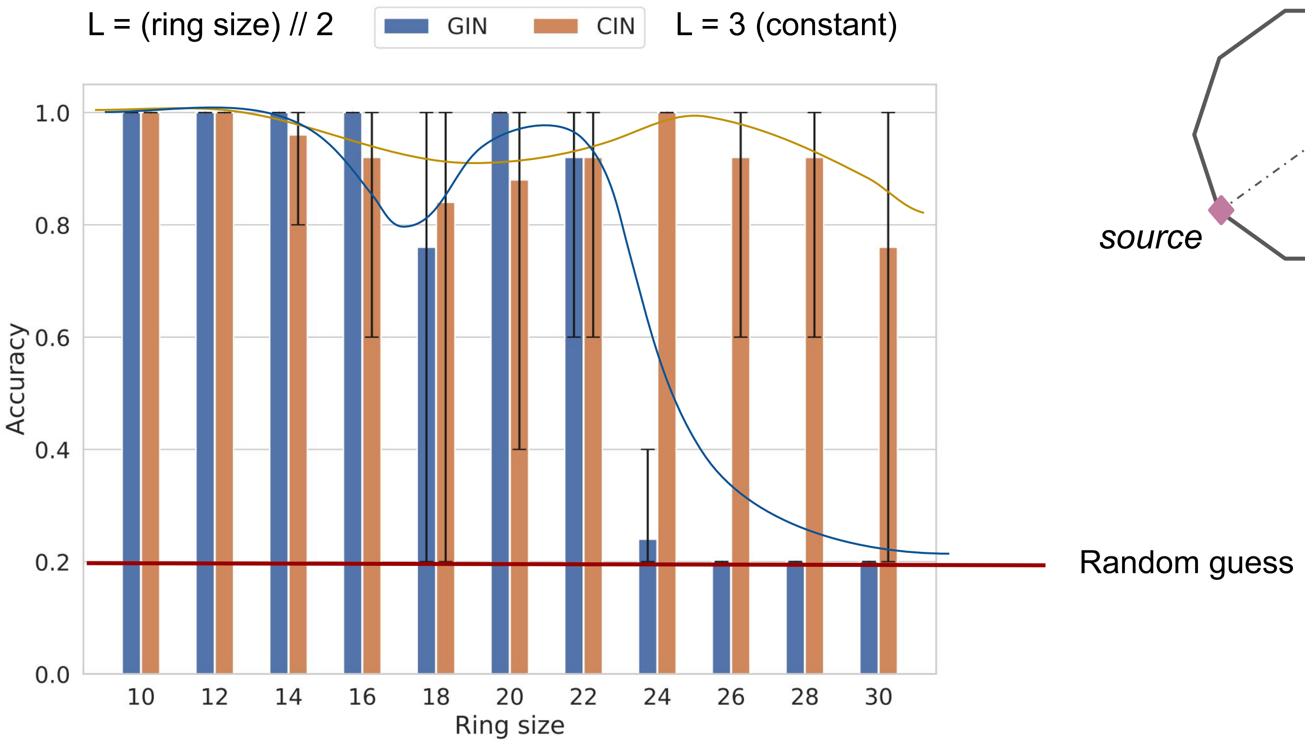
Circular Skip Links

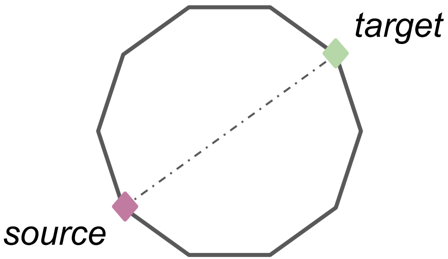
Method	Mean	Min	Max
MP-GNNs	10.000±0.000	10.000	10.000
RingGNN	10.000±0.000	10.000	10.000
3WLGNN	97.800±10.916	30.000	100.000
CIN (Ours)	100.000±0.000	100.000	100.000

Strongly Regular



Synthetic Experiments: RingTransfer





Real-world Experiments: TUDatasets

Dataset	MUTAG	PTC	PROTEINS	NCI1	NCI109	IMDB-B	IMDB-M	RDT-B
RWK [28]	79.2±2.1	55.9±0.3	59.6±0.1	>3 days	N/A	N/A	N/A	N/A
GK ($k = 3$) [62]	81.4±1.7	55.7±0.5	71.4±0.31	62.5±0.3	62.4±0.3	N/A	N/A	N/A
PK [56]	76.0±2.7	59.5±2.4	73.7±0.7	82.5±0.5	N/A	N/A	N/A	N/A
WL kernel [63]	90.4±5.7	59.9±4.3	75.0±3.1	**86.0**±1.8	N/A	73.8±3.9	50.9±3.8	81.0±3.1
DCNN [3]	N/A	N/A	61.3±1.6	56.6±1.0	N/A	49.1±1.4	33.5±1.4	N/A
DGCNN [74]	85.8±1.8	58.6±2.5	75.5±0.9	74.4±0.5	N/A	70.0±0.9	47.8±0.9	N/A
IGN [50]	83.9±13.0	58.5±6.9	**76.6**±5.5	74.3±2.7	**72.8**±1.5	72.0±5.5	48.7±3.4	N/A
GIN [72]	89.4±5.6	64.6±7.0	76.2±2.8	82.7±1.7	N/A	75.1±5.1	52.3±2.8	**92.4**±2.5
PPGNs [51]	**90.6**±8.7	66.2±6.6	**77.2**±4.7	83.2±1.1	**82.2**±1.4	73.0±5.8	50.5±3.6	N/A
Natural GN [20]	89.4±1.6	**66.8**±1.7	71.7±1.0	82.4±1.3	N/A	73.5±2.0	51.3±1.5	N/A
GSN [9]	**92.2** ± 7.5	**68.2** ± 7.2	**76.6** ± 5.0	**83.5** ± 2.0	N/A	**77.8** ± 3.3	**54.3** ± 3.3	N/A
SIN [7]	N/A	N/A	76.4 ± 3.3	82.7 ± 2.1	N/A	**75.6** ± 3.2	**52.4** ± 2.9	**92.2** ± 1.0
CIN (Ours)	**92.7** ± 6.1	**68.2** ± 5.6	**77.0** ± 4.3	**83.6** ± 1.4	**84.0** ± 1.6	**75.6** ± 3.7	**52.7** ± 3.1	**92.4** ± 2.1

Real-world Experiments: ZINC + MolHIV

Table 3: ZINC (MAE), ZINC-FULL (MAE) and Mol-HIV (ROC-AUC).

Method	ZINC ↓		ZINC-FULL ↓	MOLHIV ↑
	No Edge Feat.	With Edge Feat.	All methods	All methods
GCN [45]	0.469±0.002	N/A	N/A	76.06±0.97
GAT [67]	0.463±0.002	N/A	N/A	N/A
GatedGCN [10]	0.422±0.006	0.363±0.009	N/A	N/A
GIN [72]	0.408±0.008	0.252±0.014	0.088±0.002	77.07±1.49
PNA [19]	0.320±0.032	0.188±0.004	N/A	79.05±1.32
DGN [5]	0.219±0.010	0.168±0.003	N/A	79.70±0.97
HIMP [26]	N/A	0.151±0.006	0.036±0.002	78.80±0.82
GSN [9]	0.139±0.007	0.108±0.018	N/A	77.99±1.00
CIN-small (Ours)	0.139±0.008	0.094±0.004	0.044±0.003	80.55±1.04
CIN (Ours)	**0.115±0.003**	**0.079±0.006**	**0.022±0.002**	**80.94±0.57**

Real-world Experiments: ZINC (ablation study)

Method	MAE
GatedGCN [10]	0.363±0.009
GIN [72]	0.252±0.014
PNA [19]	0.188±0.004
DGN [5]	0.168±0.003
HIMP [26]	0.151±0.006
GSN [9]	0.108±0.018
GIN-E Custom	0.196±0.007
CIN No-Rings small	0.174±0.006
CIN No-Rings	0.159±0.007
CIN-small	0.094±0.004
CIN	**0.079±0.006**

- "Standard"
- Updates edges
- Captures rings

Importance of updating edges

Importance of capturing rings

Thanks!

ArXiv

Github

Cristian Bodnar

Twitter: @CristianBodnar

Email: cb2015@cam.ac.uk

Fabrizio Frasca

Twitter: @ffabffrasca

Email: ffrasca@twitter.com