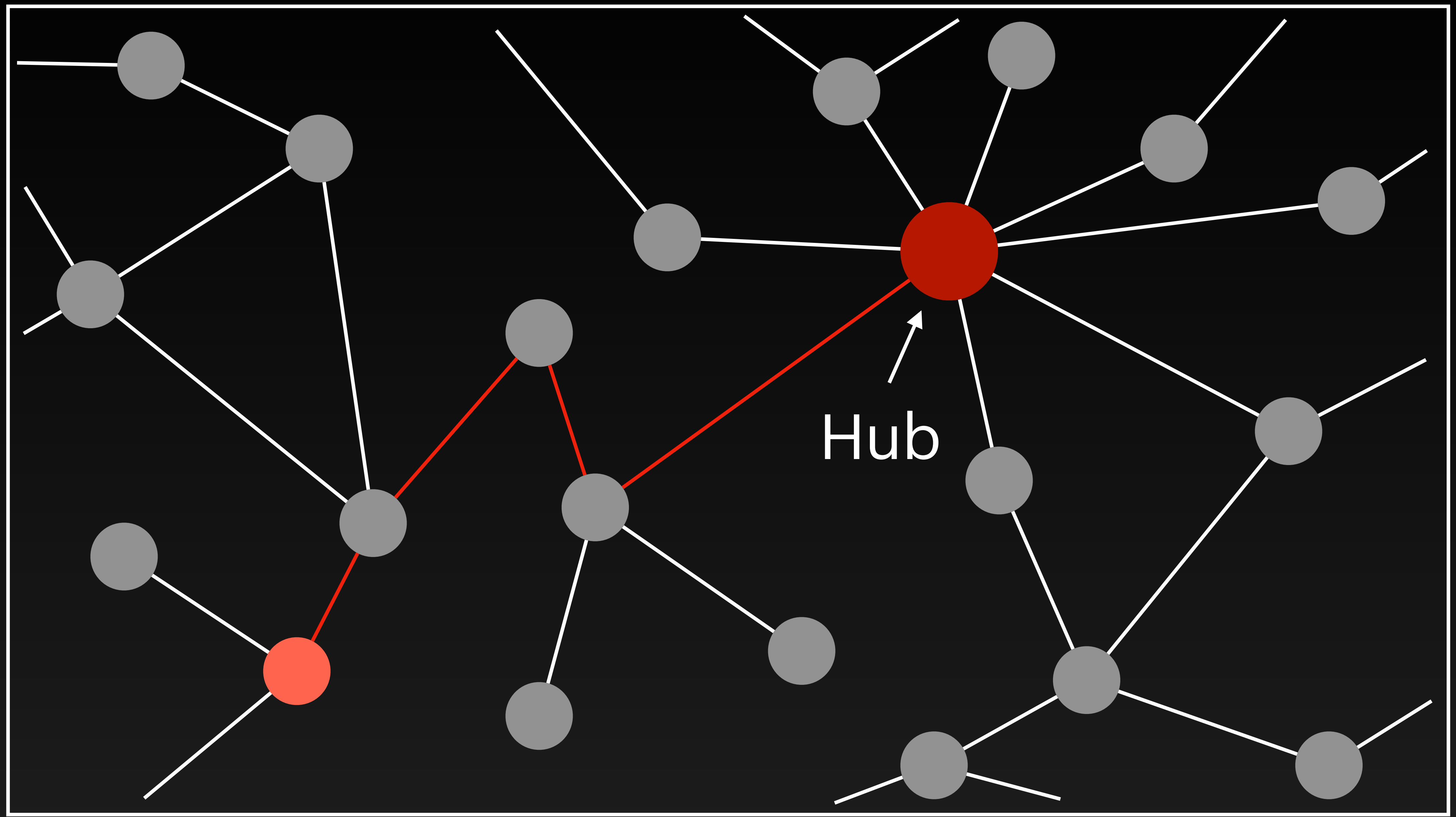
Are Graph Convolutional Networks Fully Exploiting Graph Structure?

Davide Buffelli, Fabio Vandin - *University of Padova*
{davide.buffelli, fabio.vandin}@unipd.it



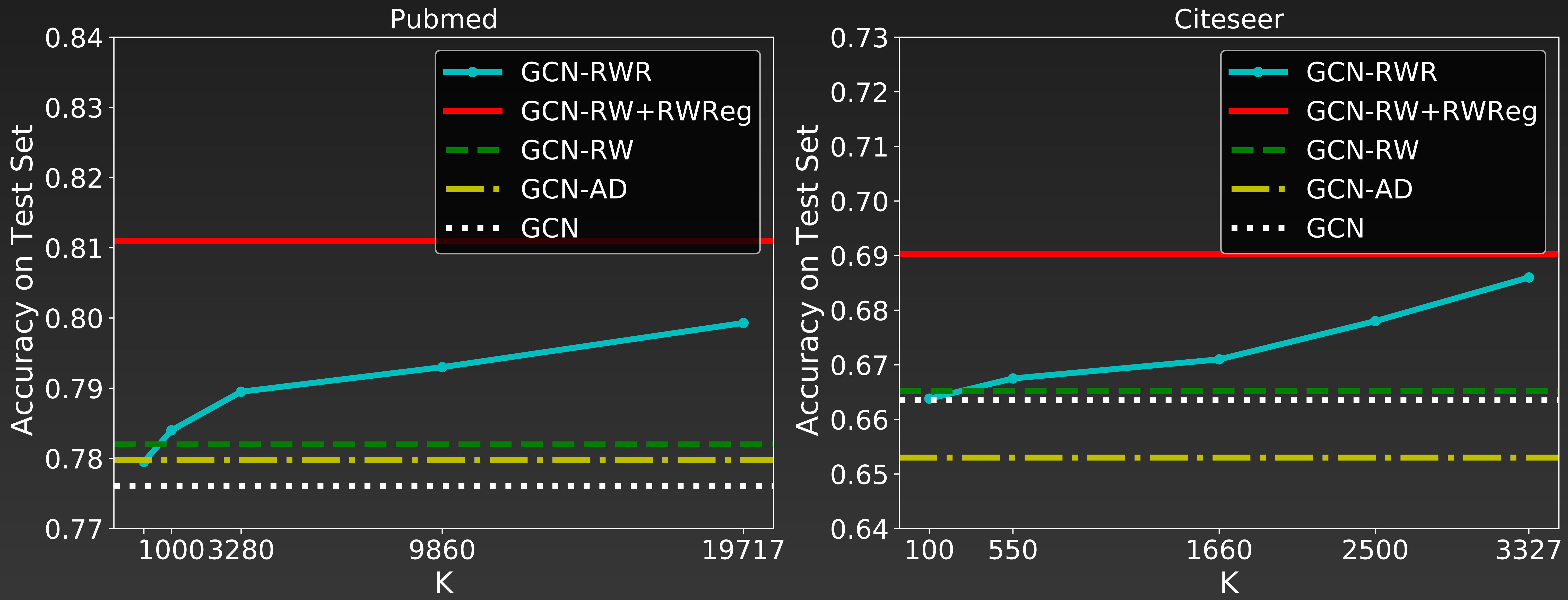
What is the impact of missing long range dependencies?

Three levels of structural knowledge injection:

- Adjacency Matrix [AD] $\hat{X} = [X, A]$
- RWR Statistics [RW] $\hat{X} = [X, S]$
- RWR Statistics + Regularization Term [RW+RWR]
  $\hat{X} = [X, S]$ + $\mathscr{L}_{\text{RWReg}} = \sum_{i,j} s_{i,j} ||H_{i,:} - H_{j,:}||^2$

Why Random Walks with Restart?

- Empirical Results
  RWR capture non-trivial dependencies between nodes
- Theoretical Results
  RWR-based colourings speed up the WL algorithm



RWR Regularization [RWR] ⬆️ >5% accuracy

Are Graph Convolutional Networks Fully Exploiting Graph Structure?

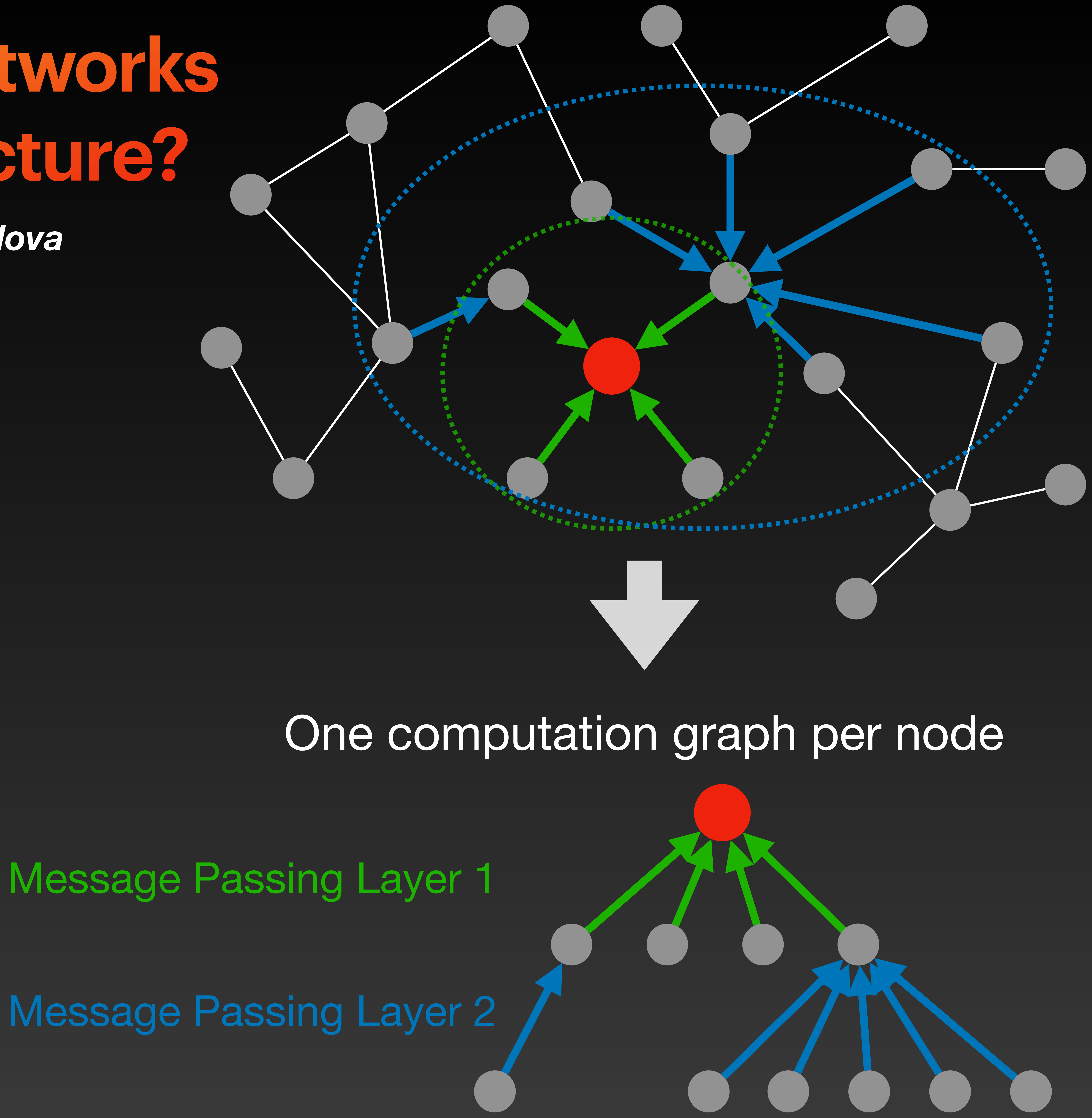
<u>Davide Buffelli</u>, Fabio Vandin - *University of Padova*
{davide.buffelli, fabio.vandin}@unipd.it

Graph Convolutional Networks

Message Passing Iteration l:

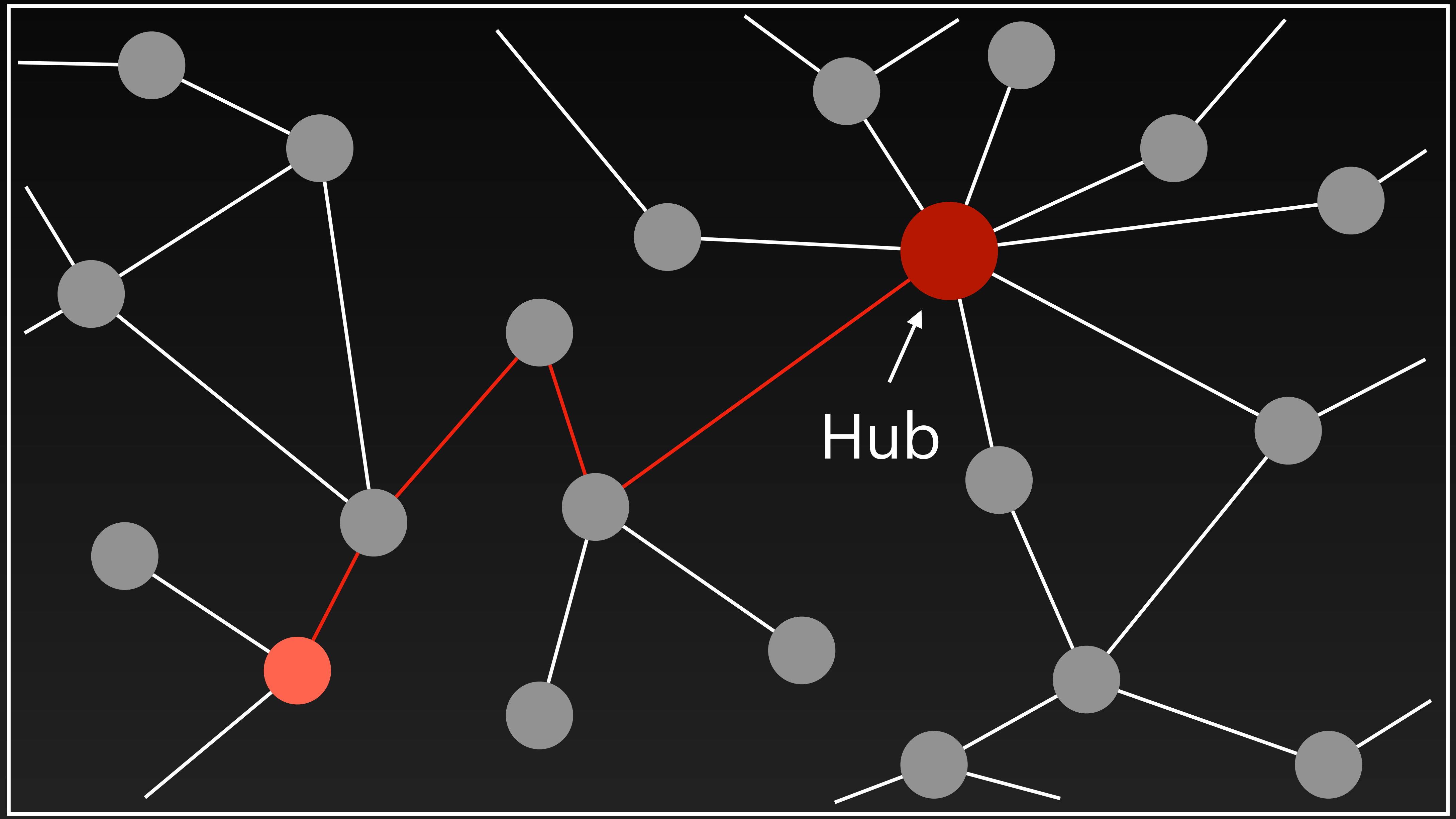
- Create Messages $$m_{ij}^{l} = MSG(h_i^{l-1}, h_j^{l-1}, e_{ij})$$
- Aggregate messages from neighbours $$M_i^l = AGG(\{m_{ij}^l | v_j \in N(v_i)\})$$
- Update node representation $$h_i^l = UPDATE(\{M_i^l, h_i^{l-1}\})$$

After k message passing iterations a node will access information from its k-hop neighbourhood



Are Graph Convolutional Networks Fully Exploiting Graph Structure?

Davide Buffelli, Fabio Vandin - *University of Padova*
{davide.buffelli, fabio.vandin}@unipd.it



Motivation:
Information in a graph is not limited to local neighbourhoods, however oversmoothing is preventing GCNs from long-range communications.

How much information is lost?
Can we try to quantify it?
How can we introduce long-range dependencies?

Quantifying the impact of missing long range dependencies

Three levels of structural knowledge injection:

- Adjacency Matrix $\hat{X} = [X, A]$
- RWR Statistics $\hat{X} = [X, S]$
- RWR Statistics + Regularization Term
  $\hat{X} = [X, S] \; + \; \mathscr{L}_{\text{RWReg}} = \sum_{i,j} s_{i,j} ||H_{i,:} - H_{j,:}||^2$

Empirical Evaluation

5 architectures:
- GCN
- GAT
- GraphSage
- DiffPool
- k-GNN

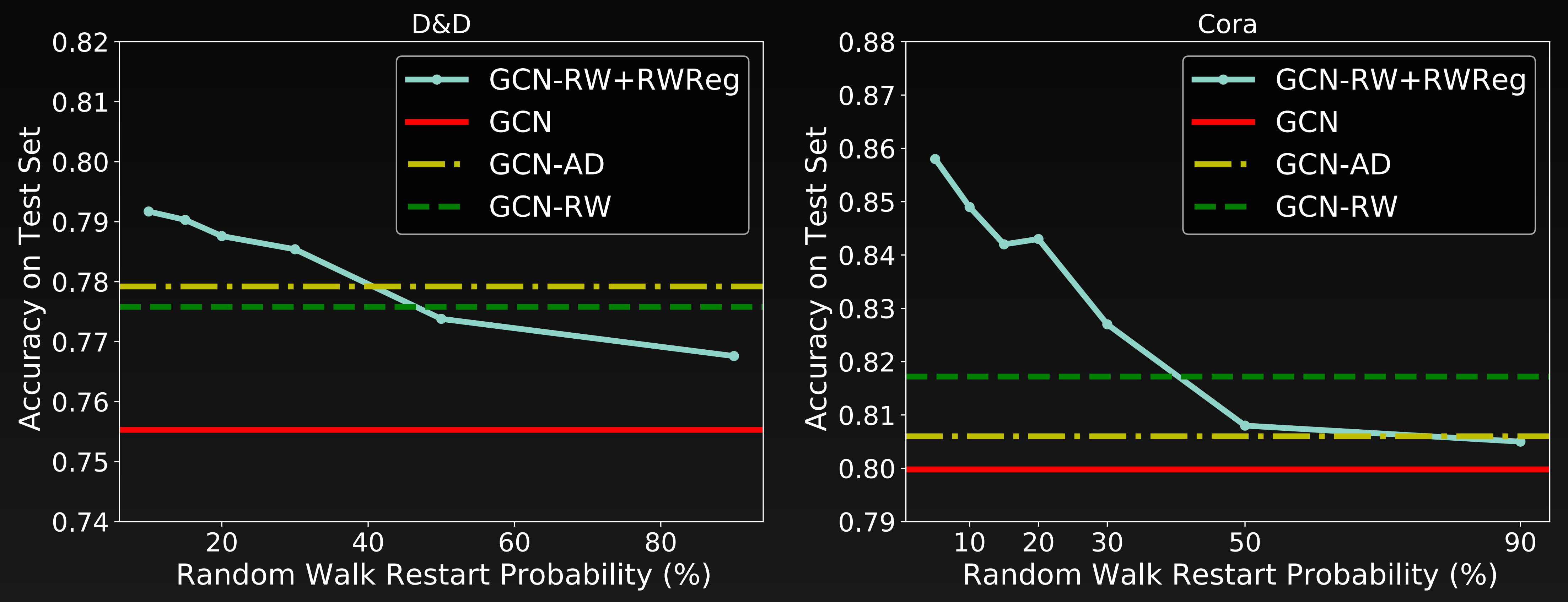
3 tasks:
- Node Classification
- Graph Classification
- Triangle Count

Results:
- ⬆️ Accuracy
- ⬆️ Generalization on Inductive Tasks
- All Architectures are affected

Are Graph Convolutional Networks Fully Exploiting Graph Structure?

Davide Buffelli, Fabio Vandin - *University of Padova*
{davide.buffelli, fabio.vandin}@unipd.it

Changing Restart Probability



Why Random Walks with Restart?

Theoretical Results

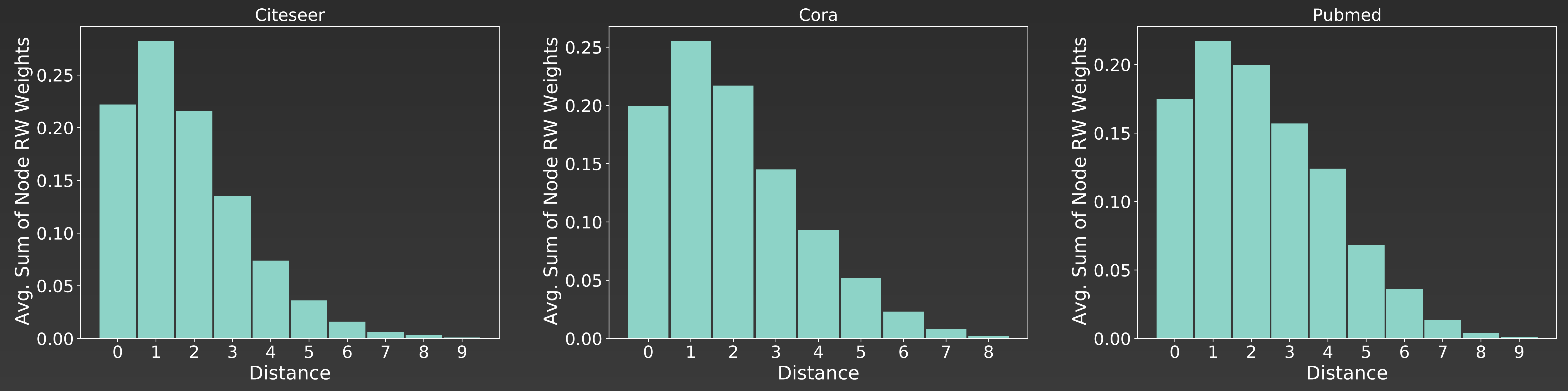
RWR-based colourings speed up the WL algorithm

Proposition Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two non-isomorphic graphs for which the 1-WL algorithm terminates with the correct answer after k iterations and starting from the labelling of all 1's. Then the k-long random walk representations of G_1 and G_2 are different.

Empirical Results

RWR capture non-trivial dependencies between nodes

Dataset	Average Kendall Tau-b
Cora	0.729 ± 0.082
Pubmed	0.631 ± 0.057
Citeseer	0.722 ± 0.171



Are Graph Convolutional Networks Fully Exploiting Graph Structure?

Davide Buffelli, **Fabio Vandin** - ***University of Padova***
{davide.buffelli, fabio.vandin}@unipd.it

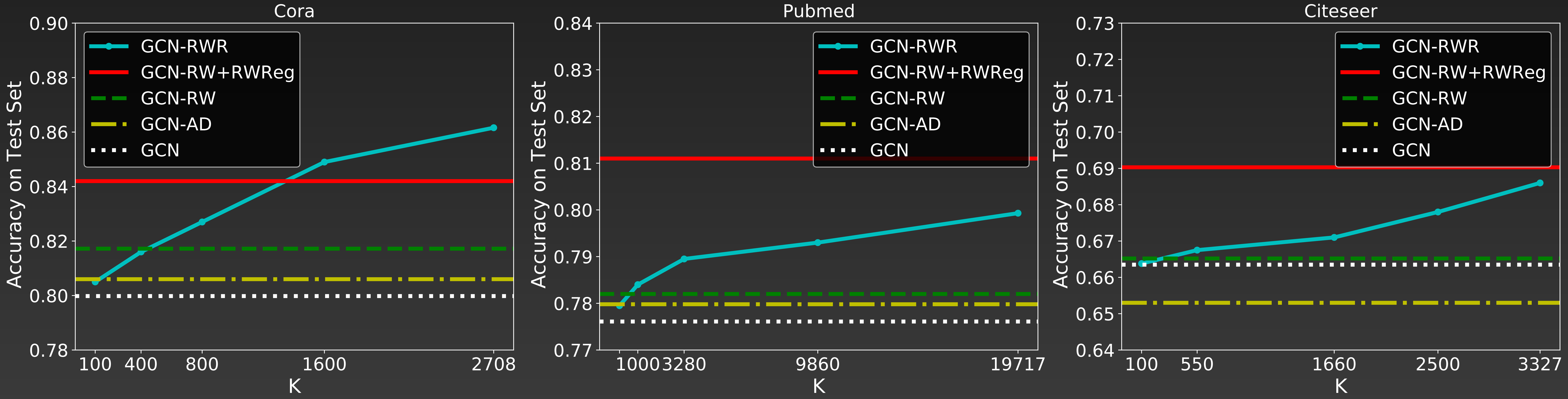
RWReg: a Tractable Method

[RWR] $\mathscr{L}_{\textbf{RWReg}} = \sum_{i,j} s_{i,j} ||H_{i,:} - H_{j,:}||^2$

- Permutation Invariant
- No additional operations at inference time
- No additional features
- ⬆️ > 5% accuracy improvement
- No additional parameters
- Control scalability with K

Results on Node Classification (Accuracy)

	GCN	GCN RW+RWReg	GCN RWR
Cora	0.7998 ±0.029	0.8420 ±0.026	**0.8616 ±0.025**
Pubmed	0.7761 ±0.022	**0.8110 ±0.037**	0.7993 ±0.034
Citeseer	0.6635 ±0.095	**0.6903 ±0.102**	0.6860 ±0.096



Are Graph Convolutional Networks Fully Exploiting Graph Structure?

Davide Buffelli, **Fabio Vandin** - ***University of Padova***
{davide.buffelli, fabio.vandin}@unipd.it

Empirical Results

The impact of structural knowledge:

- Improvements on all architectures
- Improvements on all tasks
- RW & RW+RWReg give important contributions
- Graph Classification is more affected
- Increase in generalization in Triangle Counting

Node Classification (Accuracy)

Model	Dataset	Structural Information			
		none	AD	RW	RW+RWReg
GCN	Cora	0.7998 ± 0.029	0.8060 ± 0.035	0.8172 ± 0.025	$\mathbf{0.8420 \pm 0.026}$
	Pubmed	0.7761 ± 0.022	0.7798 ± 0.070	0.7820 ± 0.042	$\mathbf{0.8110 \pm 0.037}$
	Citeseer	0.6635 ± 0.095	0.6530 ± 0.104	0.6652 ± 0.098	$\mathbf{0.6903 \pm 0.102}$
GraphSage	Cora	0.8061 ± 0.017	0.8031 ± 0.014	0.8067 ± 0.014	$\mathbf{0.8270 \pm 0.015}$
	Pubmed	0.8071 ± 0.016	0.8037 ± 0.013	0.8076 ± 0.015	$\mathbf{0.8202 \pm 0.010}$
	Citeseer	0.6811 ± 0.021	0.6889 ± 0.020	0.6930 ± 0.019	$\mathbf{0.7280 \pm 0.020}$
GAT	Cora	0.8150 ± 0.021	0.8236 ± 0.019	0.8332 ± 0.020	$\mathbf{0.8480 \pm 0.019}$
	Pubmed	0.8046 ± 0.011	0.7966 ± 0.014	0.8114 ± 0.009	$\mathbf{0.8287 \pm 0.010}$
	Citeseer	0.6646 ± 0.008	0.6720 ± 0.017	0.6866 ± 0.009	$\mathbf{0.7011 \pm 0.011}$

Triangle Count (MSE)

Model	TRIANGLES Test Set		
	Global	Small	Large
GCN	2.2900	1.3111	3.6082
GCN-AD	4.7469	1.1628	5.9717
GCN-RW	2.0449	**1.1008**	2.9889
GCN-RW+RWReg	**2.0298**	1.1664	**2.8932**

Graph Classification (Accuracy)

Model	Dataset	Structural Information			
		none	AD	RW	RW+RWReg
GCN	ENZYMES	0.5702 ± 0.052	0.5916 ± 0.076	0.5845 ± 0.055	$\mathbf{0.6166 \pm 0.065}$
	D&D	0.7553 ± 0.028	0.7792 ± 0.022	0.7758 ± 0.023	$\mathbf{0.7903 \pm 0.023}$
	PROTEINS	0.7400 ± 0.035	0.7758 ± 0.042	0.7848 ± 0.034	$\mathbf{0.7957 \pm 0.032}$
DiffPool	ENZYMES	0.6610 ± 0.031	0.7113 ± 0.027	0.6876 ± 0.025	$\mathbf{0.7214 \pm 0.039}$
	D&D	0.7931 ± 0.022	0.8376 ± 0.020	0.8248 ± 0.028	$\mathbf{0.8402 \pm 0.024}$
	PROTEINS	0.8137 ± 0.017	0.8210 ± 0.039	0.7834 ± 0.043	$\mathbf{0.8349 \pm 0.038}$
k-GNN	ENZYMES	0.5152 ± 0.111	0.5728 ± 0.063	$\mathbf{0.5730 \pm 0.077}$	0.5715 ± 0.080
	D&D	0.7562 ± 0.021	0.7785 ± 0.020	$\mathbf{0.7948 \pm 0.022}$	0.7864 ± 0.021
	PROTEINS	0.7636 ± 0.043	0.7518 ± 0.034	0.7814 ± 0.028	$\mathbf{0.7857 \pm 0.026}$