Impact of pooling methods on over-squashing and over-smoothing

Stevan Stanovic

Table of contents

Sommaire

Over-squashing Issue

Over-squashing is characterized by the fact that GNNs are almost unable to transfer information between distant nodes.

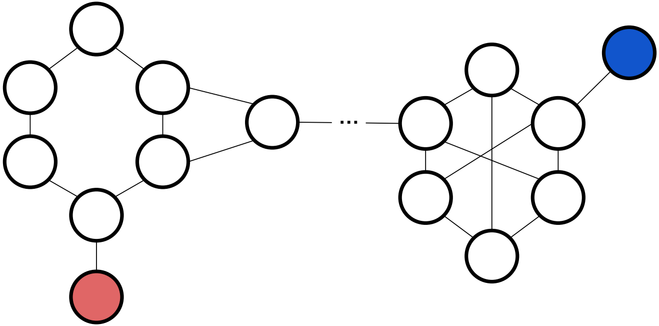


Figure: Red and blue nodes are the more distant nodes in the graph. Transferring information between them is more challenging than another pair of nodes.

Why is difficult to transfer information ?

It exists two main reasons:

- Propagation obstruction due to access/commute time [1]:
  - Access time: number of iterations needed to transfer information between two nodes
  - Commute time: number of iterations needed to transfer information and come-back between two nodes
  - Obstruction: information from nearest nodes is higher than distant nodes
- Bottleneck effect: as the number of layers increases, the number of nodes in each node's receptive field grows exponentially and messages that are propagated from distant nodes are distorted [2, 3]. Particularly if it exists few edges between dense regions in the graph.

Illustration of a graph with a bottleneck

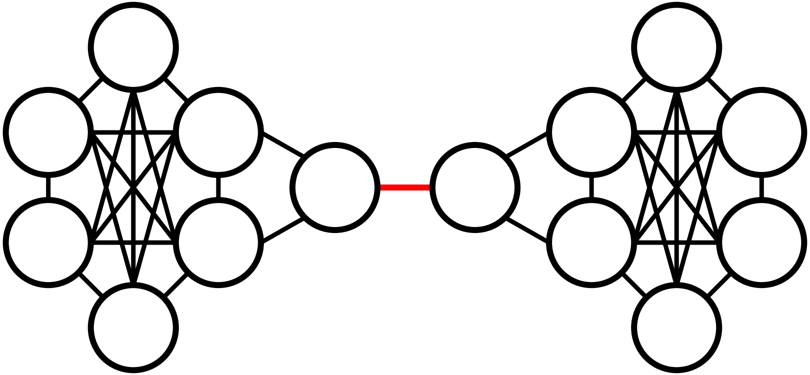


Figure: Information transiting through the red edge is concerned by over-squashing.

Commute time and existing solution to mitigate over-squashing

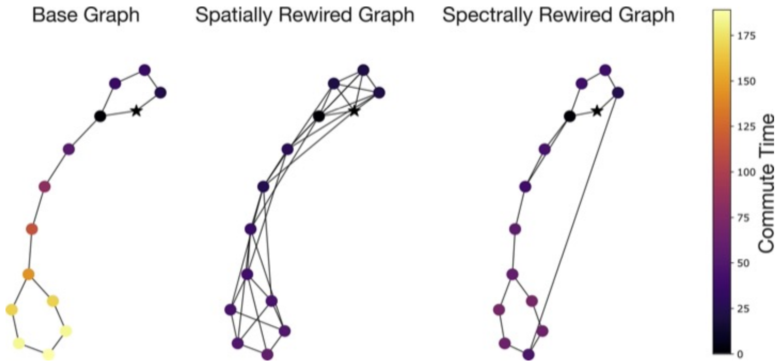


Figure: Example of rewiring methods with commute time measures on edges.

Over-smoothing Issue

Over-smoothing occurs when the number of layers, e.g. the depth, in a GNN increases. By iteratively combining neighbors node features together, all nodes representation in a graph are computed using the same information.

For a GCN, node features tend towards a common representation described by a combination of the square roots of the degrees of the graph [4, 5, 6].

Generalisation to symmetric normalized convolution

We consider the following convolution : $X_{n+1} = CX_n = C^{n+1}X_0$. Let note $-1 \leq \lambda_1 \leq ... \leq \lambda_m = 1$ be the m eigenvalues of C and $v_1, ..., v_m$ the corresponding eigenvectors. Assuming that $X_0 = \alpha_1 v_1 + ... + \alpha_m v_m$ and using the power iteration algorithm, we know that:

$$\exists p \ \mid \ \forall k \geq p \quad : \quad \|C^k X_0 - \alpha_m \lambda_m^k v_m\| < \gamma(\frac{\lambda_{m-1}}{\lambda_m})^k \text{ where } \gamma \text{ is a scalar.}$$

Note that v_m is equal to $\mathbf{D}^{\frac{1}{2}}\mathbb{1}$.

Dirichlet Energy

Over-smoothing can be efficiently characterized by the mean Dirichlet energy defined for graph $\mathcal{G}^{(l)}$ as $E(X^{(l)}) = Trace((X^{(l)})^T L X^{(l)})/|\mathcal{G}^{(l)}|$ where L is the Laplacian associated with the convolution operator. By utilizing the convolution described in [7], the mean Dirichlet energy can be written as:

$$E(\mathbf{X}^{(l)}) = \frac{1}{|\mathcal{G}^{(l)}|} \sum_{i \in \mathcal{V}^{(l)}} \sum_{j \in \mathcal{N}_i^{(l)}} \mathbf{A}_{ij}^{(l)} \| \frac{\mathbf{X}_i^{(l)}}{\sqrt{d_i + 1}} - \frac{\mathbf{X}_j^{(l)}}{\sqrt{d_j + 1}} \|_2^2$$

where $\mathcal{N}_i^{(l)}$ and d_i are respectively the neighborhood and the degree of node i.

Sommaire

Definitions

Given a vertex v surviving at layer l we denote by $RW^l(v)$ the Reduction Window of v at level l.

The receptive field of v at level l $RF^l(v)$ corresponds to the set of vertices defined at the base level graph which are merged onto v at level l. More formally, the receptive field at level l is defined recursively :

Definition

Let $\mathcal{G}^{(l)}, \dots, \mathcal{G}^{(1)} = (V^1, E^1)$ denote a sequence of reduced graphs. The receptive fields at level l are defined for any vertex $v \in V^l$ as:

$$RF^l(v) = \bigcup_{u \in RW^l(v)} RF^{l-1}(u) \text{ with } RF^1(u) = RW^1(u)$$

Application to Maximal Independent Sets

Reduction windows produced by these strategies satisfy the following equations at any layer l and for any vertex $w \in V^l$:

$$\begin{cases} RW^l(w) & = & \{w\} \text{ or} \\ RW^l(w) & = & \{w, v_1, \ldots, v_n\} \text{ with } \forall i \in \{1, \ldots, n\} \; d_{G_{l-1}}(w, v_i) = 1 \end{cases} \quad (1)$$

where $d_{G_{l-1}}(.,.)$ is the distance within the graph $\mathcal{G}^{(l-1)}$ defined at layer $l-1$.

Theoretical result on over-squashing

For a convolution, we need a linear relation to intersect features from two nodes:

$$m = \frac{d_{G_0}(u,v)}{2}$$

Using a decimation scheme satisfying equation 1 we have a log relation:

$$m \approx log(d_{G_0}(u,v))$$

Theoretical result on over-smoothing

We prove that:

1. If each vertex belongs to only one reduction window $\{RF^l(v)\}_{v \in V^l}$ forms partition of V^0.

   $\Rightarrow$ applying just pooling layers like MIVSPool prevent to have over-smoothing because the set of receptive fields forms a partition of the initial vertex set

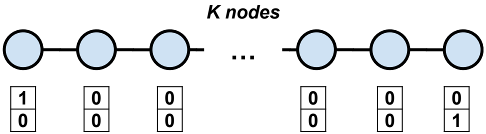
2. Let a GNN built by successive applications of GNN+MIVS-topk. At any level two non adjacent vertices are associated to disjoint receptive fields

   $\Rightarrow$ applying an alternation of GCN and Top-k methods conditioned by a MIVS (MIVS_{top-k}) also prevent over-smoothing.

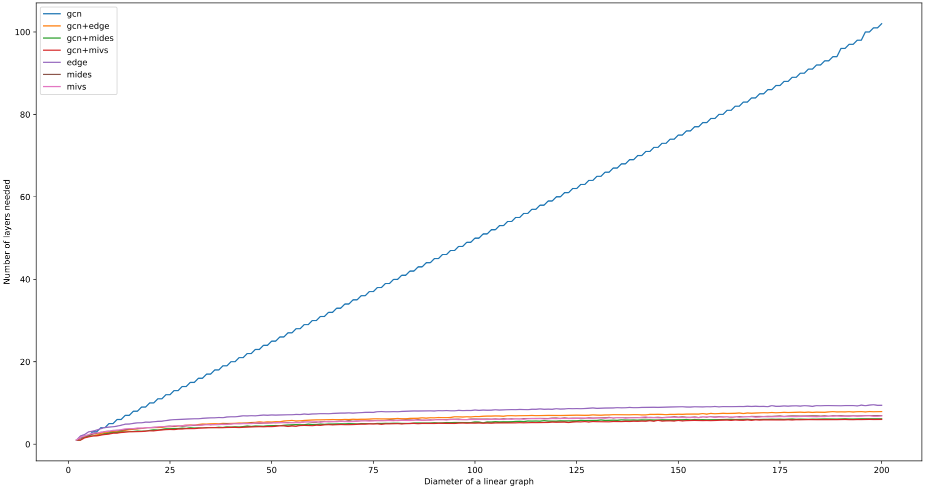
Sommaire

Experiment

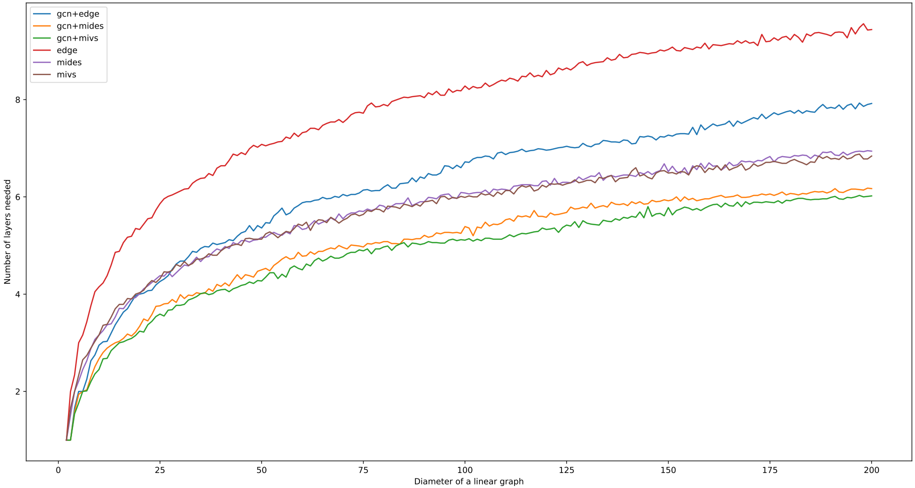
We study the number of layers needed to have the intersection of the two nodes for three different architectures: only convolutions layers (GCN), only pooling layers and an alternation of convolution and pooling.



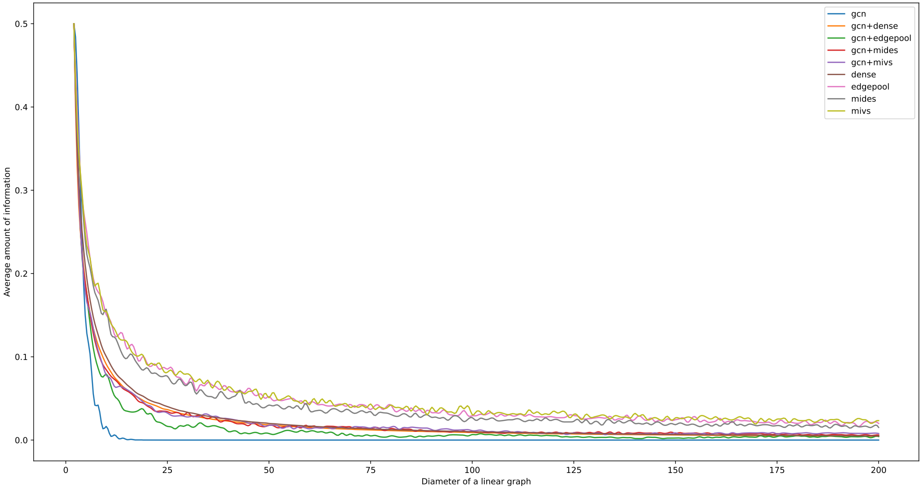
Number of layers necessary to cross receptive fields with GCN



Number of layers necessary to cross receptive fields without GCN



Average amount of information transmitted during the crossing

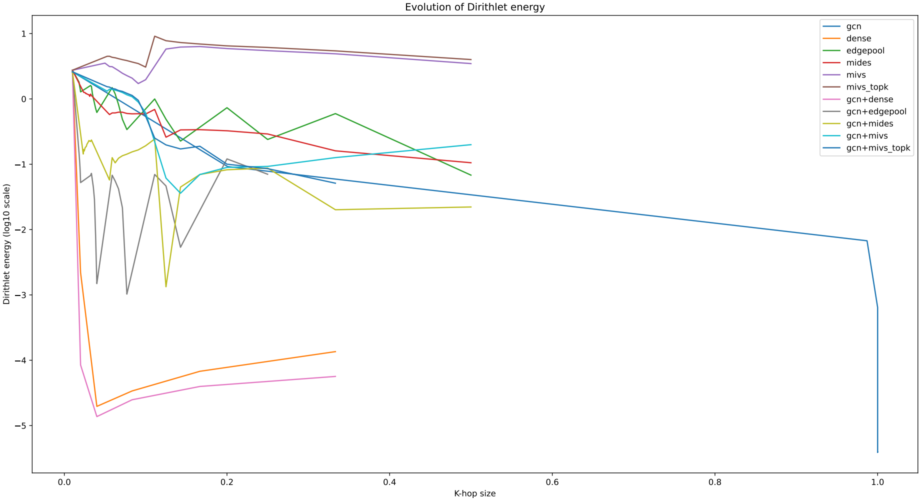


Sommaire

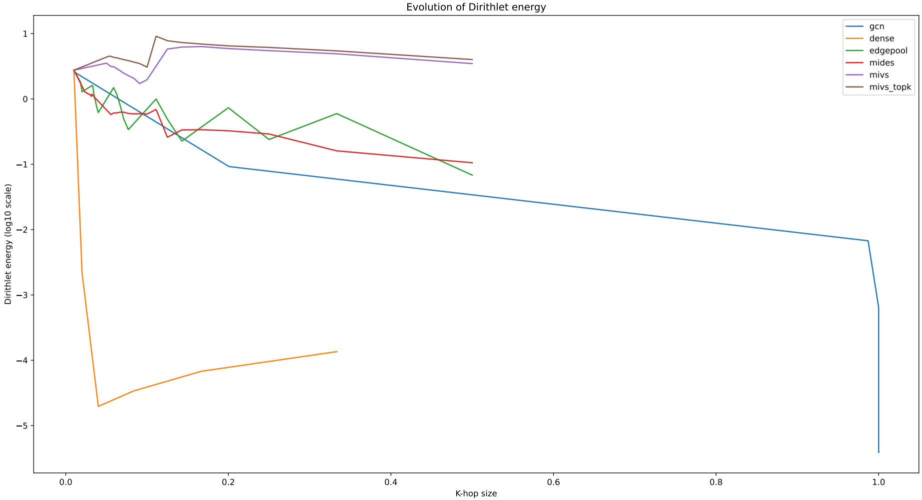
Experiment

We study the variation of the Dirichlet energy according to the number of receptive field. We compare the Dirichlet energy for different methods: only convolutions layers (GCN), only pooling layers and an alternation of convolution and pooling.

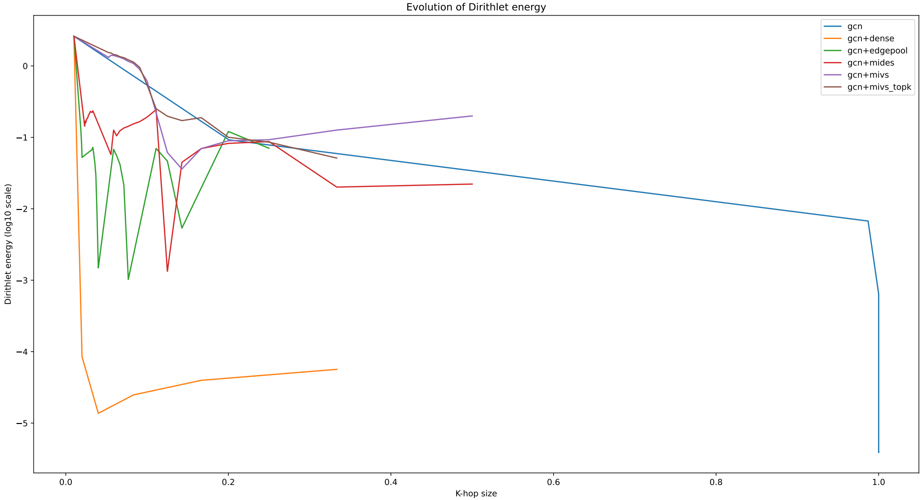
Evolution of Dirichlet energy for all methods



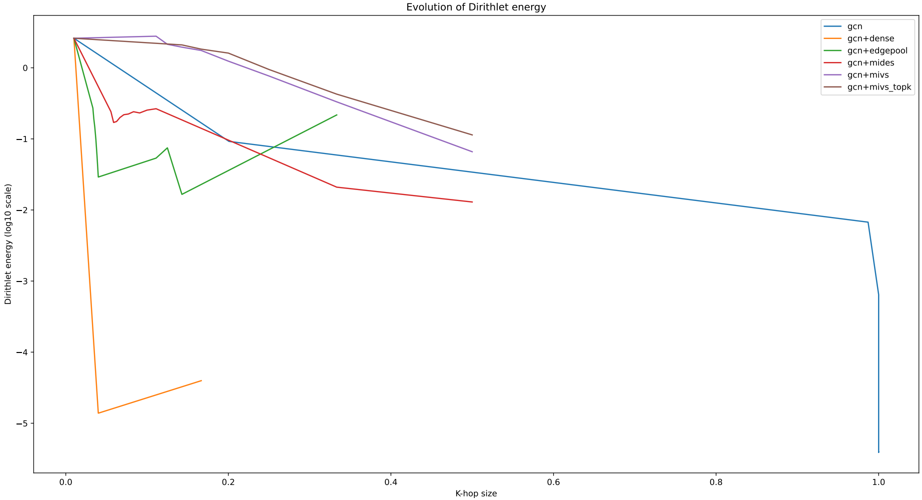
Evolution of Dirichlet energy for GCN and pooling



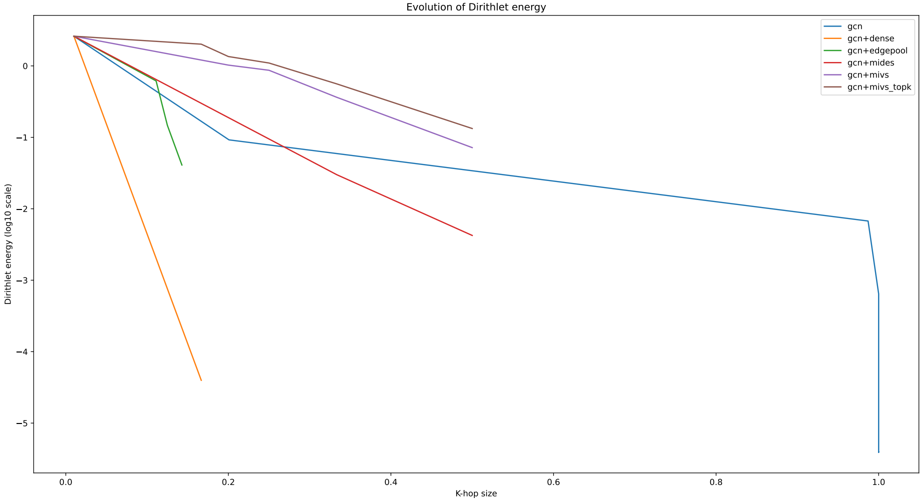
Evolution of Dirichlet energy for GCN and alternation methods



For GCN and alternation of GCN and 2 pooling



For GCN and alternation of GCN and 4 pooling



Sommaire

Conclusion

1. Accumulating normalized convolution leads to converge to a linear combination of the square root of degrees
2. Applying pooling help to intersect faster features
3. Applying our pooling strategy methods prevents to have over-smoothing

Illustration of the Definition of the Receptive Field

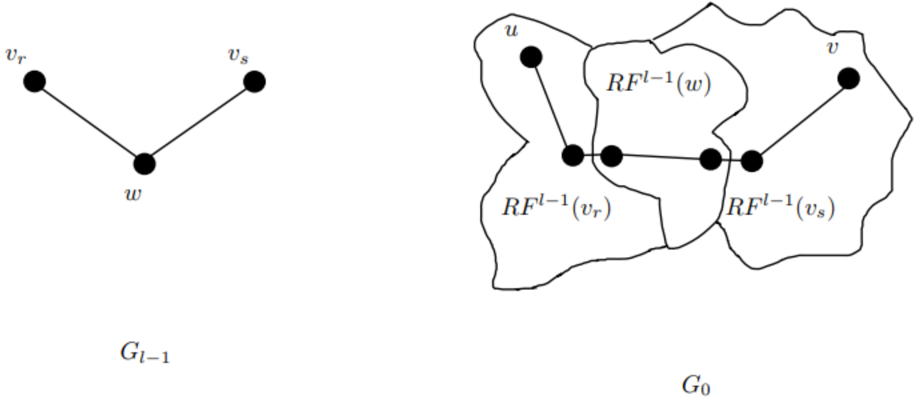


Figure: Connection between receptive fields in the base level graph. Note the two edges between $RF^{l-1}(v_r)$ and $RF^{l-1}(w)$ and between $RF^{l-1}(w)$ and $RF^{l-1}(v_s)$.

Theoretical results on over-squashing

Using Equation **??**, we can deduce a lower bound for the number of iterations needed to cluster two nodes in the same Receptive Field:

$$
\begin{aligned}
& d_{G_0}(u,v) \leq 2 * 3^m - 1 \\
\Rightarrow \quad & m \geq \log_3\left(\frac{d_{G_0}(i,j)+1}{2}\right)
\end{aligned}
\tag{2}
$$

Theoretical results on over-squashing

Supposing the existence of a constant $\gamma > 1$ such that for any level p and any couple of surviving vertices (x, y) we have:

$$d_{G_{p-1}}(x, y) \geq \gamma d_{G_p}(x, y)$$

We thus have:

$$d_{G_{m-1}}(i, j) = 1 \leq \left(\frac{1}{\gamma}\right)^{m-1} d_{G_0}(i, j) \Rightarrow \gamma^{m-1} \leq d_{G_{0]}}(i, j)$$

We thus can deduce an upper bound for the number of iterations needed to cluster two nodes in the same Receptive Field:

$$m \leq \frac{\log(d_{G_0}(i, j))}{\log(\gamma)} + 1 \qquad (2)$$

Theoretical results on over-squashing

Combining Equations 2 and **??**, we have:

$$\log_3\left(\frac{d_{G_0}(i,j)+1}{2}\right) \leq m \leq \frac{\log(d_{G_0}(i,j))}{\log(\gamma)}+1 \tag{2}$$

References I

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio, and Michael M. Bronstein.
On over-squashing in message passing neural networks: The impact of width, depth, and topology.
In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 7865–7885. PMLR, 2023.

Uri Alon and Eran Yahav.
On the bottleneck of graph neural networks and its practical implications.
In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

References II

Ja[illegible] Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M. Bronstein.
Understanding over-squashing and bottlenecks on graphs via curvature.
In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net, 2022.

Kenta Oono and Taiji Suzuki.
Graph neural networks exponentially lose expressive power for node classification.
In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020.

Qimai Li, Zhichao Han, and Xiao-Ming Wu.
Deeper insights into graph convolutional networks for semi-supervised learning.
In Sheila A. McIlraith and Kilian Q. Weinberger, editors, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances*

References III

in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pages 3538–3545. AAAI Press, 2018.

Chen Cai and Yusu Wang.
A note on over-smoothing for graph neural networks.
CoRR, abs/2006.13318, 2020.

Thomas N. Kipf and Max Welling.
Semi-supervised classification with graph convolutional networks.
In *International Conference on Learning Representations (ICLR)*, 2017.

Yann A LeCun, Léon Bottou, Genevieve B Orr, and Klaus-Robert Müller.
Efficient backprop.
In *Neural networks: Tricks of the trade*, pages 9–48. Springer, 2012.

References IV

Marco Gori, Gabriele Monfardini, and Franco Scarselli.
A new model for learning in graph domains.
In *Proceedings. 2005 IEEE international joint conference on neural networks*, volume 2, pages 729–734, 2005.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini.
The graph neural network model.
IEEE transactions on neural networks, 20(1):61–80, 2008.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst.
Convolutional neural networks on graphs with fast localized spectral filtering.
Advances in neural information processing systems, 29, 2016.

Will Hamilton, Zhitao Ying, and Jure Leskovec.
Inductive representation learning on large graphs.
Advances in neural information processing systems, 30:1024–1034, 2017.

References V

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka.
How powerful are graph neural networks?
arXiv preprint arXiv:1810.00826, 2018.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio.
Graph attention networks.
arXiv preprint arXiv:1710.10903, 2017.

Zheng Ma, Junyu Xuan, Yu Guang Wang, Ming Li, and Pietro Liò.
Path integral based convolution and pooling for graph neural networks.
Advances in Neural Information Processing Systems, 33:16421–16433, 2020.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl.
Neural message passing for quantum chemistry.
In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.

References VI

Junhyun Lee, Inyeop Lee, and Jaewoo Kang.
Self-attention graph pooling.
In *International conference on machine learning*, pages 3734–3743. PMLR, 2019.

Hongyang Gao and Shuiwang Ji.
Graph u-nets.
In *international conference on machine learning*, pages 2083–2092. PMLR, 2019.

Cătălina Cangea, Petar Veličković, Nikola Jovanović, Thomas Kipf, and Pietro Liò.
Towards sparse hierarchical graph classifiers.
arXiv preprint arXiv:1811.01287, 2018.

Boris Knyazev, Graham W Taylor, and Mohamed Amer.
Understanding attention and generalization in graph neural networks.
Advances in neural information processing systems, 32, 2019.

References VII

Zhitao Ying, Jiaxuan You, Christopher Morris, Xiang Ren, Will Hamilton, and Jure Leskovec.
Hierarchical graph representation learning with differentiable pooling.
Advances in neural information processing systems, 31:4805–4815, 2018.

Amirhossein Nouranizadeh, Mohammadjavad Matinkia, Mohammad Rahmati, and Reza Safabakhsh.
Maximum entropy weighted independent set pooling for graph neural networks.
ArXiv, abs/2107.01410, 2021.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun.
Spectral networks and locally connected networks on graphs.
arXiv preprint arXiv:1312.6203, 2013.

Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen.
An end-to-end deep learning architecture for graph classification.
Proceedings of the AAAI Conference on Artificial Intelligence, 32(1):4438–4445, 2018.

References VIII

Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard Zemel.
Gated graph sequence neural networks.
arXiv preprint arXiv:1511.05493, 2015.

Oriol Vinyals, Samy Bengio, and Manjunath Kudlur.
Order matters: Sequence to sequence for sets.
arXiv preprint arXiv:1511.06391, 2015.

Luc Brun and Walter Kropatsch.
Hierarchical graph encodings.
Image processing and analysis with graphs: theory and practice, 2012.

Peter Meer.
Stochastic image pyramids.
Computer Vision, Graphics, and Image Processing, 45(3):269–294, 1989.

References IX

Paul D Dobson and Andrew J Doig.
Distinguishing enzyme structures from non-enzymes without alignments.
Journal of molecular biology, 330(4):771–783, 2003.

Karsten M Borgwardt, Cheng Soon Ong, Stefan Schönauer, SVN Vishwanathan, Alex J Smola, and Hans-Peter Kriegel.
Protein function prediction via graph kernels.
Bioinformatics, 21(suppl_1):47–56, 2005.

Nikil Wale, Ian A Watson, and George Karypis.
Comparison of descriptor spaces for chemical compound retrieval and classification.
Knowledge and Information Systems, 14(3):347–375, 2008.

Jean-Michel Jolion and Annick Montanvert.
The adaptive pyramid: a framework for 2d image analysis.
CVGIP: Image Understanding, 55(3):339–348, 1992.

Diego Mesquita, Amauri Souza, and Samuel Kaski.
Rethinking pooling in graph neural networks.
Advances in Neural Information Processing Systems, 33:2220–2231, 2020.

Frederik Diehl, Thomas Brunner, Michael Truong Le, and Alois Knoll.
Towards graph pooling by edge contraction.
In *ICML 2019 workshop on learning and reasoning with graph-structured data*, 2019.

Stevan Stanovic, Benoit Gaüzère, and Luc Brun.
Maximal independent vertex set applied to graph pooling.
In *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, S+ SSPR 2022, Montreal, QC, Canada, August 26–27, 2022, Proceedings*, pages 11–21. Springer, 2023.

Nitika Verma, Edmond Boyer, and Jakob Verbeek.
Feastnet: Feature-steered graph convolutions for 3d shape analysis.
In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2598–2606, 2018.

References XI

Yll Haxhimusa.
The structurally Optimal Dual Graph Pyramid and its application in image partitioning, volume 308.
IOS Press, 2007.

William L Hamilton.
Graph representation learning.
Synthesis Lectures on Artifical Intelligence and Machine Learning, 14(3):1–159, 2020.

Francesco Landolfi.
Revisiting edge pooling in graph neural networks.
In *ESANN*, 2022.

Pinar Yanardag and SVN Vishwanathan.
Deep graph kernels.
In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1365–1374, 2015.

References XII

Filippo Maria Bianchi, Daniele Grattarola, Lorenzo Livi, and Cesare Alippi.
Hierarchical representation learning in graph neural networks with node decimation pooling.
IEEE Trans. Neural Networks Learn. Syst., 33(5):2195–2207, 2022.

Zhen Zhang, Jiajun Bu, Martin Ester, Jianfeng Zhang, Zhao Li, Chengwei Yao, Huifen Dai, Zhi Yu, and Can Wang.
Hierarchical multi-view graph pooling with structure learning.
IEEE Transactions on Knowledge and Data Engineering, 35(1):545–559, 2023.

Federico Errica, Marco Podda, Davide Bacciu, and Alessio Micheli.
A fair comparison of graph neural networks for graph classification.
arXiv preprint arXiv:1912.09893, 2019.

References XIII

Muhammet Balcilar, Renton Guillaume, Pierre Héroux, Benoit Gaüzère, Sébastien Adam, and Paul Honeine.
Analyzing the expressive power of graph neural networks in a spectral perspective.
In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021.

Aamir Anis, Akshay Gadde, and Antonio Ortega.
Towards a sampling theorem for signals on arbitrary graphs.
In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3864–3868. IEEE, 2014.

Yuichi Tanaka, Yonina C Eldar, Antonio Ortega, and Gene Cheung.
Sampling signals on graphs: From theory to applications.
IEEE Signal Processing Magazine, 37(6):14–30, 2020.

References XIV

Stevan Stanovic, Benoit Gaüzère, and Luc Brun.
Maximal independent sets for pooling in graph neural networks.
In Donatello Conte Pasquale Foggia, editor, *In 8th IAPR - TC-15 Workshop on Graph-based Representations in Pattern Recognition (GBR'23)*, Lecture Notes in Computer Science. Springer, September 2023.
to be published.

Filippo Maria Bianchi and Veronica Lachi.
The expressive power of pooling in graph neural networks.
CoRR, abs/2304.01575, 2023.

Deli Chen, Yankai Lin, Wei Li, Peng Li, Jie Zhou, and Xu Sun.
Measuring and relieving the over-smoothing problem for graph neural networks from the topological view.
In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The*

References XV

Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020, pages 3438–3445. AAAI Press, 2020.

Daniele Grattarola, Daniele Zambon, Filippo Maria Bianchi, and Cesare Alippi.
Understanding pooling in graph neural networks.
IEEE Transactions on Neural Networks and Learning Systems, pages 1–11, 2022.

Singh Akansha.
Over-squashing in graph neural networks: A comprehensive survey.
arXiv preprint arXiv:2308.15568, 2023.

Cedric Sanders, Andreas Roth, and Thomas Liebig.
Curvature-based pooling within graph neural networks.
arXiv preprint arXiv:2308.16516, 2023.

George Dasoulas, Johannes F. Lutzeyer, and Michalis Vazirgiannis.
Learning parametrised graph shift operators.
In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021.

Florian Dörfler and Francesco Bullo.
Kron reduction of graphs with applications to electrical networks.
IEEE Trans. Circuits Syst. I Regul. Pap., 60-I(1):150–163, 2013.

George Karypis and Vipin Kumar.
Metis – unstructured graph partitioning and sparse matrix ordering system, version 2.0.
01 1995.

References XVII

Davide Bacciu, Alessio Conte, and Francesco Landolfi.
Generalizing downsampling from regular data to graphs.
In Brian Williams, Yiling Chen, and Jennifer Neville, editors, *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, pages 6718–6727. AAAI Press, 2023.

T. Konstantin Rusch, Michael M. Bronstein, and Siddhartha Mishra.
A survey on oversmoothing in graph neural networks.
CoRR, abs/2303.10993, 2023.

Aamir Anis, Akshay Gadde, and Antonio Ortega.
Towards a sampling theorem for signals on arbitrary graphs.
In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014*, pages 3864–3868. IEEE, 2014.

Luc Brun and Walter Kropatsch.
Image processing and analysing with graphs: Theory and practice, chapter hierarchical graph encoding, 2012.