ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

On the Rank Properties of the Renormalization Trick in GCNs

Anna Bison and Alessandro Sperduti

University of Padova - Department of Mathematics “Tullio Levi-Civita”
Padova - Italy

Abstract. We analyze the renormalization trick in GCNs beyond its established spectral smoothing effect. We prove that self-loops can increase the rank of the propagation matrix by resolving local symmetries that otherwise induce linear dependencies, providing a rigorous explanation for the trick’s effectiveness: through Oono and Suzuki’s framework, the rank increment counteracts the loss of expressive power. From a spectral point of view, the addition of self-loops in GCNs ensures that some information located in the normalized adjacency’s kernel is preserved and propagated rather than discarded.

1 Introduction

Self-loops in Graph Neural Networks (GNNs) were originally introduced by Kipf and Welling [1] via the so-called *renormalization trick*. Derived as a first-order approximation of a ChebNet spectral filter, this modification was primarily validated through its empirical effectiveness. Following this work, several studies investigated the theoretical reasons behind the observed performance boost, mainly focusing on spectral properties and signal propagation dynamics. For instance, Wu et al. [2] (Simple GCN) demonstrated that self-loops cause the spectrum of the normalized Laplacian to shrink towards zero, thereby enforcing a stronger low-pass filtering effect that smooths the signal. More recently, Lampert and Scholtes [3] analyzed the phenomenon from a random walk perspective, highlighting the trade-off between local feature preservation and global diffusion rates. Similarly, Bose and Das [4] provided an extensive analysis of spectral shifts under varying self-loop weights, correlating these shifts with performance on heterophilic graphs. These works extensively cover the dynamical effects of self-loops (i.e., how they alter the speed of smoothing or the frequency response), while in this paper we focus the attention on their structural impact on the rank dynamics of the adjacency matrix of the graph, a metric recently highlighted as a robust proxy for over-smoothing [5]. We characterize the renormalization trick as a spectral shift that modifies the adjacency rank by changing the dimension of its kernel. Since this rank variation is invariant under degree normalization, despite the misalignment of eigenspaces, the rank dynamics of the adjacency matrix extends to its normalized forms, and thus to the operators typically used in GCNs.

*This work was supported by the Italian Ministry of University and Research under the PRIN program - Progetti di Rilevante Interesse Nazionale - PRIN 2022 (Secretary-General’s Decree n. 1401 of 18/09/2024) - CUP B53C24006570006, project title “DEEP-GRAPH: Design and Theory of Deep Graph Learning”.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

2 Notation

Vectors and matrices are denoted by arrows $\vec{v}$ and uppercase letters X, respectively. We write v_i for vector entries, X_{ij} for matrix entries, and X_i for the i-th row of a matrix. We denote the standard basis vectors as e_i and the symbol $\mathbf{1} = \sum_{i=1}^{N} e_i$ denotes the vector of all ones, with N the dimension of the vector space. We denote by δ_{ij} the Kronecker delta, defined as $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected, connected and unweighted graph *without self loops*, where $\mathcal{V} = \{1, \dots, |\mathcal{V}|\}$ is the set of nodes with associated features $X \in \mathbb{R}^{|\mathcal{V}| \times m}$, $\mathcal{E}$ is the set of edges represented into the adjacency matrix A ($A_{ij} = 1$ if connected, 0 otherwise), and D the diagonal degree matrix. Neighborhood of node i is denoted as $\mathcal{N}_i$. Based on these, we define the unnormalized Laplacian $\Delta = D - A$, the normalized Laplacian $\Delta_{\text{norm}} = I - A_{\text{norm}}$, where $A_{\text{norm}} = D^{-1/2} A D^{-1/2}$, and the normalized Laplacian augmented with self-loops $\tilde{\Delta}_{\text{norm}} = I - \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}$ with $\tilde{D} = D + I$, $\tilde{A} = A + I$, $\tilde{A}_{\text{norm}} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}$ in particular $\tilde{\Delta}_{\text{norm}} = \tilde{D}^{-1/2} \Delta \tilde{D}^{-1/2}$.

3 Rank Dynamics of the Renormalization Trick

In this section, we analyze the impact of the renormalization trick introduced in [1]. We consider the procedure applied to an undirected, connected and unweighted graph $\mathcal{G}$ which is initially *without self-loops*.
Since the transformation $A \mapsto \tilde{A} = A + I$ shifts the spectrum by +1, the kernel of A (associated with $\lambda = 0$) is mapped to $\lambda_{new} = 1$. Conversely, the eigenspace associated with $\lambda = -1$ in A is mapped to the kernel of $\tilde{A}$ ($\lambda_{new} = 0$), potentially reducing the rank. Therefore, the rank variation follows the balance:

$$\text{rk}(\tilde{A}) - \text{rk}(A) = \dim(\text{Ker}(A)) - \dim(E_{-1}), \tag{1}$$

where E_{-1} denotes the eigenspace of A associated with the eigenvalue -1. In the following, we analyze these two opposing contributions. Notice that the previous equation implies a sufficient condition for full expressivity: if -1 is not in the spectrum of A, then 0 is not in the spectrum of $\tilde{A}$, guaranteeing it is full-rank.

3.1 Rank Gain

A contribution to rank gain occurs when the addition of self-loops removes existing linear dependencies among the rows of A. In connected, unweighted graphs without self-loops, row-wise linear dependence may arise from *generalized neighborhood equivalence*: sets of pairwise non-adjacent nodes with pairwise disjoint neighborhoods, whose union of neighborhoods coincides with the union of neighborhoods of another such set. The addition of the identity matrix breaks this symmetry by introducing a unique feature, the self-loop, that distinguishes each node from its structurally equivalent peers. An example considering singletons follows:

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Example 1. *The addition of self-loops corresponds to the transformation $A \mapsto A+I$, which, from a row-wise perspective, maps each row A_i to $\tilde{A}_i = A_i + e_i^T$. Let $i, j \in \mathcal{V}$ with $i \neq j$. Under the assumptions on $\mathcal{G}$, linear dependence between two rows implies $A_i = A_j$. Consequently, since the graph has no self-loops, nodes i and j cannot be connected ($A_{ij} = A_{ji} = 0$). After the addition of self-loops, the new rows $\tilde{A}_i$ and $\tilde{A}_j$ become linearly independent, as the presence of the identity term e_i in $\tilde{A}_i$ (with value 1 at index i) cannot be matched by $\tilde{A}_j$ (which has value 0 at index i, since $A_{ji} = 0$ and e_j is zero at i).*

If before the transformation two unconnected nodes shared the same neighborhood, the addition of self-loops ensures their respective augmented adjacency rows become linearly independent. This precise set of nodes is formally defined as (see Fig. 1a):

$$V_1 = \{i \in \mathcal{V} \mid \exists\, j \neq i \text{ s.t. } A_i = A_j\}.$$

3.2 Rank Loss

The renormalization trick acts destructively on the rank when it introduces new linear dependencies. While the spectral cause is definitively identified as the presence of the eigenvalue $\lambda = -1$, the topological origin is generally linked to local structural symmetries, such as those found in tightly connected clusters, that induce linear dependencies among node neighborhoods.

Example 2. *Addition of self loops creates linear dependencies in a set of connected nodes sharing the same external connectivity, i.e.:*

$$V_2 = \{i \in \mathcal{V} \mid \exists\, j \neq i \text{ s.t. } A_{ij} = 1 \land \forall\, k \neq i, j, A_{ik} = A_{jk}\}.$$

In fact, in the original matrix A, the rows associated with these nodes are linearly independent because the zero diagonal entries distinguish a node from its neighbors: specifically, for any pair i, j in the cluster, the structural symmetry is broken by $A_{ii} = 0 \neq A_{ji} = 1$. The addition of self-loops removes this distinction. In the augmented matrix $\tilde{A}$, the diagonal entry becomes equal to the off-diagonal connection ($\tilde{A}_{ii} = \tilde{A}_{ji} = 1$). Consequently, the rows become identical, causing the linear dependency that reduces the rank (see Fig. 1b).

4 Extension to Normalized Dynamics

From a Graph Signal Processing perspective, we must verify that the rank variations discussed for the augmented adjacency $\tilde{A}$ translate to the normalized matrix $\tilde{A}_{\text{norm}} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}$. Since the augmented degree matrix $\tilde{D} = D + I$ contains strictly positive diagonal entries, the normalization matrix $\tilde{D}^{-1/2}$ is always invertible. A fundamental property of linear algebra states that rank is invariant under multiplication by non-singular matrices. Therefore:

$$\text{rank}(\tilde{A}_{\text{norm}}) = \text{rank}(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}) = \text{rank}(\tilde{A}).$$

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Case 1: disconnected symmetric nodes (V_1)

$$A = \begin{pmatrix} \mathbf{0} & \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{0} & 1 \\ 1 & 1 & 0 \end{pmatrix} \xrightarrow{\text{Rank=2}} \text{Singular}$$

$$\tilde{A} = \begin{pmatrix} \mathbf{1} & \mathbf{0} & 1 \\ \mathbf{0} & \mathbf{1} & 1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{Rank=3}} \text{Full Rank}$$

(a) Nodes i, j share neighbor k but are not connected. A has identical rows. Adding self-loops restores independence ($e_i \neq e_j$).

Case 2: connected symmetric nodes (V_2)

$$A = \begin{pmatrix} \mathbf{0} & \mathbf{1} & 1 \\ \mathbf{1} & \mathbf{0} & 1 \\ 1 & 1 & 0 \end{pmatrix} \xrightarrow{\text{Rank=3}} \text{Full Rank}$$

$$\tilde{A} = \begin{pmatrix} \mathbf{1} & \mathbf{1} & 1 \\ \mathbf{1} & \mathbf{1} & 1 \\ 1 & 1 & 1 \end{pmatrix} \xrightarrow{\text{Rank=1}} \text{Singular}$$

(b) Connected nodes i, j share neighbor k. A rows are distinct. Adding self-loops makes rows identical ($\tilde{r}_i = \tilde{r}_j = [1, 1, 1]$).

Fig. 1: Scenarios where adding self-loops affects the rank of A in different ways.

Furthermore, under the assumption of connected graphs, the original degree matrix D is also non-singular (as $d_i \geq 1$ for all nodes). Thus, the invariance $\text{rank}(A_{\text{norm}}) = \text{rank}(A)$ holds strictly. This ensures that the kernel dimensions are preserved through normalization in both configurations. Consequently, the rank difference established in Eq. (1) holds equivalently for the normalized operators:

$$\text{rk}(\tilde{A}_{\text{norm}}) - \text{rk}(A_{\text{norm}}) = \text{rk}(\tilde{A}) - \text{rk}(A).$$

Thus, the rank dynamics derived in the previous section apply directly to the GCN propagation layer: any increase in rank corresponds to a reduction in the dimension of the signal subspace that is annihilated during convolution. Interpreting these results through the theoretical framework established by Oono and Suzuki [6] (specifically Theorem 2), the rank dynamics have direct implications for signal preservation. Components of the input signal $X^{(0)}$ residing in the kernel of A_{norm} are associated with an eigenvalue $\lambda = 0$ and are thus annihilated after a single propagation step. Following the spectral shift induced by self-loops, these same components may shift to a subspace associated with non-zero eigenvalues ($\lambda \neq 0$) in $\tilde{A}_{\text{norm}}$. Consequently, rather than suffering immediate deletion, these signal components enter the regime of exponential decay (or conservation), effectively surviving the first layer. Thus, adding self-loops modulates the set of signals that are instantaneously filtered out by the architecture. A reduction in the kernel dimension implies that information previously discarded at the onset is now retained for subsequent processing. This mechanism suggests that the accuracy improvements observed in many datasets may stem from the architecture's recovered ability to propagate expressive signal components that, without self-loops, would have been prematurely suppressed.

4.1 A Note on Spectral Analysis

In general, the adjacency matrix transformed with self-loops $\tilde{A}_{\text{norm}}$ has a different spectrum w.r.t. A_{norm}, as it results from the analysis in [2]. However, a

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

particular that must be kept into account and that seems not specified in [2], is that the orthonormal base of $\tilde{A}_{\text{norm}}$ is in general different from the orthonormal base of A_{norm}, and hence the decomposition of graph signals induced by the two operators is different: eigenspaces that were associated to the high eigenvalues of the first operator are not in general the same eigenspaces that are associated to the high eigenvalues of the second one. To analyze the effect of the impact of renormalization trick from this point of view, it's useful to consider the canonical operator w.r.t. usually the graph signals are decomposed: Δ_{norm}. When self loops are added, the Laplacian obtained after the transformation is $\tilde{\Delta}_{\text{norm}} = I - \tilde{A}_{\text{norm}} = (D+I)^{-1/2}\Delta(D+I)^{-1/2} = \tilde{D}^{-1/2}\Delta\tilde{D}^{-1/2}$. As a consequence[1], the kernel of $\tilde{\Delta}_{\text{norm}}$ (i.e. the eigenspace of eigenvalue 1 of $\tilde{A}_{\text{norm}}$) is now generated by $\tilde{D}^{1/2}\mathbf{1}$, which is different from the eigenspace of Δ_{norm}, that instead is generated by $D^{1/2}\mathbf{1}$. This due to the fact that for non-regular graphs, and hence in general, $\tilde{\Delta}_{\text{norm}}$ does not commute with Δ_{norm}:

Theorem 1. *In general,* $[\Delta_{norm}, \tilde{\Delta}_{norm}] = \Delta_{norm}\tilde{\Delta}_{norm} - \tilde{\Delta}_{norm}\Delta_{norm} \neq 0_{|\mathcal{V}|\times|\mathcal{V}|}$ *and then they are not simultaneously diagonalizable.*

Proof. The proof proceeds analogously to the one presented in [7] for Δ and Δ_{norm}. To show that $[\Delta_{\text{norm}}, \tilde{\Delta}_{\text{norm}}] \neq 0_{|\mathcal{V}|\times|\mathcal{V}|}$, we will show that in general $[\Delta_{\text{norm}}, \tilde{\Delta}_{\text{norm}}]D^{1/2}\mathbf{1} \neq \vec{0}$:

$$[\Delta_{\text{norm}}, \tilde{\Delta}_{\text{norm}}]D^{1/2}\mathbf{1} = \Delta_{\text{norm}}\tilde{\Delta}_{\text{norm}}D^{1/2}\mathbf{1} - \tilde{\Delta}_{\text{norm}}\Delta_{\text{norm}}D^{1/2}\mathbf{1} =$$

$$\Delta_{\text{norm}}\tilde{\Delta}_{\text{norm}}D^{1/2}\mathbf{1} = \Delta_{\text{norm}} \sum_{i=1}^{|\mathcal{V}|}\left(\sqrt{d_i} - \frac{\sqrt{d_i}}{d_i+1} - \sum_{j\in\mathcal{N}_i}\frac{\sqrt{d_j}}{\sqrt{(d_i+1)(d_j+1)}}\right)e_i \neq \vec{0}$$

since $\Delta_{\text{norm}}\vec{v} = \vec{0} \iff \exists\, c \in \mathbb{R} \mid v_i = \sqrt{d_i}c \;\forall\, i \in \mathcal{V}$, and in general, due to local degree differences $\nexists\, c \in \mathbb{R} \mid \forall\, i \in \mathcal{V}\; (\sqrt{d_i} - \frac{\sqrt{d_i}}{d_i+1} - \sum_{j\in\mathcal{N}_i}\frac{\sqrt{d_j}}{\sqrt{(d_i+1)(d_j+1)}}) = c\sqrt{d_i}$, from which the thesis follows. □

For this reason, when considering that the largest eigenvalue of $\tilde{\Delta}_{\text{norm}}$ is $\leq$ of the largest eigenvalue of Δ_{norm}, as shown in [2], when $\tilde{\Delta}_{\text{norm}}$ and Δ_{norm} are not simultaneously diagonal, it should be taken in account that components of the signals that are associated to the same eigenvalue in the spectrum of Δ_{norm}, are in general not associated to the same eigenvalue of $\tilde{\Delta}_{\text{norm}}$, and hence the "low" frequencies components that are damped after few iterations of Δ_{norm}, will be possibly a superposition of different eigenvectors of $\tilde{\Delta}_{\text{norm}}$, and hence split and then damped in a completely different way than before. Despite this spectral misalignment of eigenspaces between A, A_{norm}, and $\tilde{A}_{\text{norm}}$, the fundamental insight of our analysis holds: the *rank dynamics* discussed in the previous sections remains invariant, since the normalization matrices $D^{-1/2}$ and $\tilde{D}^{-1/2}$ are full-rank diagonal matrices.

[1] It is worth noting that, unlike the adjacency matrix, the rank of the Laplacian is invariant to the addition of self-loops. Indeed, the standard Laplacian satisfies $\tilde{\Delta} = \tilde{D} - \tilde{A} = (D+I) - (A+I) = D - A = \Delta$. Since the operator itself remains unchanged, its rank (and that of its normalized counterparts) is strictly preserved.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Example 3 (Complete Graphs). *Matrix A for complete graphs has maximum rank, and then the addition of the self-loops makes all rows the same, collapsing rank from maximum to 1. This means that in a space of graph signals of dimension n, a subspace of dimension $n-1$ will be completely damped after one single application of the operator $\tilde{A}_{norm}$. Consequently, even one single application of $\tilde{A}_{norm}$ to the embedding matrix of the graph will dump all the components of its signals, leaving only the components that live in the one-dimensional subspace orthonormal to the hyperspace generated by $\sum_{i=1}^{n} x_i = 0$, i.e., the subspace generated by constant signals $\mathbf{1}$.*

5 Conclusions

While [6] details asymptotic decay assuming the augmented operator, our analysis isolates the critical role of self-loops in initial signal preservation. We demonstrate that, by inducing operator incompatibility and altering rank, self-loops transition components from the kernel of A_{norm} to the active spectrum. Thus, self-loops actively rescue information from immediate annihilation, offering a theoretical framework to interpret how specific topological properties modulate their empirical effectiveness.

Usage of AI Tools The author used Gemini for providing LaTeX support, assisting in the drafting and refinement of specific text sections, and for acting as an interactive conversational agent to stress-test the mathematical intuition and verify the consistency of the formal statements through counter-examples.

References

[1] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[2] Felix Wu, Amauri H. Souza Jr., Tianyi Zhang, Christopher Fifty, Tao Yu, and Kilian Q. Weinberger. Simplifying graph convolutional networks. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 6861–6871. PMLR, 2019.

[3] Moritz Lampert and Ingo Scholtes. The self-loop paradox: Investigating the impact of self-loops on graph neural networks. https://arxiv.org/abs/2312.01721, 2023.

[4] Kushal Bose and Swagatam Das. Learning from heterophilic graphs: A spectral theory perspective on the impact of self-loops and parallel edges. https://arxiv.org/abs/2509.13139, 2025.

[5] Kaicheng Zhang, Piero Deidda, Desmond J. Higham, and Francesco Tudisco. Rethinking oversmoothing in graph neural networks: A rank-based perspective. https://doi.org/10.48550/arXiv.2502.04591, 2025.

[6] 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.

[7] Anna Bison and Alessandro Sperduti. Analysis of dirichlet energies as over-smoothing measures. https://arxiv.org/abs/2512.09890, 2025.