SCIENCE CHINA
Information Sciences

• LETTER •

January 2026, Vol. 69, Iss. 1, 119203:1–119203:2
https://doi.org/10.1007/s11432-024-4546-0

Neural algorithmic approach to network dismantling

Peng ZHANG[2], Jun FU[1,2*], Xiaojie SUN[1], Tonglei CHENG[3] & Guanrong CHEN[4]

[1]*State Key Laboratory of Synthetical Automation for Process Industries, Northeastern University, Shenyang 100819, China*
[2]*School of Future Technology, Northeastern University, Shenyang 100819, China*
[3]*College of Information Science and Engineering, Northeastern University, Shenyang 100819, China*
[4]*Department of Electrical Engineering, City University of Hong Kong, Hong Kong 999077, China*

Received 27 October 2024/Revised 7 March 2025/Accepted 22 April 2025/Published online 4 January 2026

Citation Zhang P, Fu J, Sun X J, et al. Neural algorithmic approach to network dismantling. Sci China Inf Sci, 2026, 69(1): 119203, https://doi.org/10.1007/s11432-024-4546-0

The network dismantling (ND) problem, which involves identifying the minimum set of nodes that will cause a system to collapse if removed, remains a critical challenge in network science. Inspired by their advancements in optimization, deep learning techniques have been applied to study the ND problem. Deep learning methods, including FINDER [1] (a reinforcement learning-based approach) and GDM [2] (a supervised learning-based approach), have demonstrated potential but have also encountered limitations: FINDER incurs high computational costs by iteratively recalculating node importance during removal, while GDM's reliance on manual feature engineering introduces human bias. These limitations are primarily attributed to the uncharted relationship between their model design and reasoning ability, which hinders the advancement of deep learning in ND.

To gain further insight into this relationship, we analyze it using the principle of algorithmic alignment [3] and propose a more general supervised learning framework, stable-aligned graph dismantling with machine learning (Stable-AGDM), for the ND problem. In Figure 1, the model under consideration comprises three distinct hierarchical learning modules, each corresponding to a specific learning level for the ND problem. These modules are graph structure learning, topological correlation learning, and feature selection. Notably, the processing graph structure and topological correlation information necessitate graph neural network architectures with different abilities. To mitigate the influence of imbalanced training data (Appendix B.2.1), Stable-AGDM integrates the stable learning framework [4] in its feature selection module. This integration balances information distribution across samples to enhance model robustness. A concise overview is provided below, with more exhaustive details relegated to Appendix B.

Neural feature engineering. The fundamental distinction lies in the replacement of GDM's manual feature engineering (e.g., degree, clustering coefficient) with a neural feature engineering module in Stable-AGDM. Manual feature engineering, in its rigid implementation, extracts predefined structural features and limits the model's adaptability. This, in turn, results in suboptimal dismantling performance on specific complex networks (Figure S6). In contrast, neural feature engineering employs an adaptive learning process to identify network-specific structural patterns, thereby reducing human bias and enhancing performance and interpretability.

The design of neural feature engineering is motivated by the contrasting efficacy of two approaches: supervised learning models like GDM (reliant on manual feature engineering) and reinforcement learning frameworks like FINDER (feature-free). Despite methodological divergence, both approaches achieve comparable performance for the ND problem, prompting further investigation into the influence of manual feature engineering on the performance of these models. In Figure S6(a), GDM with k-core features exhibits superior performance in comparison to variants devoid of this feature. This finding underscores the relevance of global location information in addressing the ND problem. In Figure S6(b), a robust correlation is observed among degree, clustering coefficient, and k-core, while the χ^2 of degree demonstrates distinct heterogeneity.

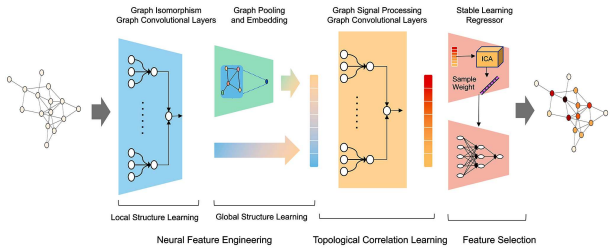

Figure 1 (Color online) Overview of the Stable-AGDM framework.

When analyzed with Figure S6(a), it was determined that the GDM model with the χ^2 of degree outperforms the one without the χ^2 of degree. Therefore, incorporating features with substantial heterogeneity can enhance model performance by distinguishing nodes across topological positions. These observations demonstrate that the efficacy of manual feature engineering is contingent upon its capacity to capture graph structural features that are imperative for ND. In a similar manner, GNNs can extract such structural features, thereby providing a data-driven alternative to manual feature engineering.

To circumvent the potential for human experience bias in model learning, a neural feature engineering module has been devised,

* Corresponding author (email: junfu@mail.neu.edu.cn)

comprising two components: local and global structure learning. The GNNs employed in the context of local structure learning must possess the capacity to generate robust topological representations, thereby facilitating the acquisition of node structure information. A residual gated graph neural network (ResGatedG) is employed to learn local structure vectors $\boldsymbol{h}_{local} \in \mathbb{R}^{n\times p}$. However, it should be emphasized that any GNN that enhances topological representation ability can be used for this module. In the domain of global structure learning, graph pooling is employed to acquire a graph feature vector $\boldsymbol{h}_{graph} \in \mathbb{R}^{1\times p}$, which encompasses global structure information. This vector is then integrated with the local structure vectors to yield global structure vectors $\boldsymbol{h}_{global} \in \mathbb{R}^{n\times p}$ in the structure presentation space:

$$\boldsymbol{h}_{global_i} = \boldsymbol{W}_5 \,\mathrm{ReLU}\left(\boldsymbol{h}_{local;i}^{\top} \cdot \boldsymbol{h}_{graph} \cdot \boldsymbol{W}_6\right), \tag{1}$$

where $\boldsymbol{h}_{global_i} \in \mathbb{R}^{1\times p}$ denotes the updated global structure vector of the target node i, $\boldsymbol{h}_{local_i} \in \mathbb{R}^{1\times p}$ denotes the local structure vector of the target node i, and $\boldsymbol{W}_5 \in \mathbb{R}^{1\times p}$ and $\boldsymbol{W}_6 \in \mathbb{R}^{p\times p}$ are learnable parameters.

Topological correlation learning. This module is designed to generate node representation vectors by capturing correlations among global structure vectors $\boldsymbol{h}_{global} \in \mathbb{R}^{n\times p}$. The critical node selection in ND is significantly influenced by higher-order interactions, necessitating the use of graph signal processing-based GNNs to model topological correlations. A graph attention network (GAT) introduces adaptive edge weights, analogous to a self-adaptive graph shift operator in signal processing, to prioritize correlations between structural features. Accordingly, GAT is employed to generate the node representation vectors $\boldsymbol{H} \in \mathbb{R}^{n\times m_H}$ in the representation space. However, it should be emphasized that any graph convolutional layer that enhances feature correlation learning ability can be used for this module.

Feature selection. This module integrates the stable learning framework to identify stable features within the representation space. The design motivation of the module is derived from Figure S6(c), which illustrates the degradation in performance that occurs when neural architectures replace manual feature engineering (i.e., the AGDM model). This occurrence can be attributed to the sparsity of nodes in optimal dismantling sets, which results in severe training data imbalance. Consequently, models tend to overfit the majority class (zero-labeled nodes) while underfitting the minority class (nonzero-labeled nodes), particularly under conditions of limited samples and high model complexity. Consequently, existing supervised methods continue to depend on manual feature engineering despite its inherent limitations, as neural feature engineering increases model complexity. Moreover, the absence of proper data balancing frequently results in training failure. A more nuanced understanding of this phenomenon can be attained by employing a feature selection framework. Specifically, while neural feature engineering flexibly captures rich graph information, it inevitably introduces redundant or irrelevant features, making models struggle to identify task-critical features with limited training data.

To identify features that have a practical impact on ND, a feature selection module is designed to eliminate irrelevant and spurious correlations. The stable learning framework can be statistically understood as a feature selection mechanism based on regression coefficients [4], which can distinguish stable features from nonstable features in the representation space. Consequently, a feature selection module is adopted, based on the stable learning framework. This module selects stable feature variables through a weighted loss function, with the objective of increasing the robustness of the model.

Result. Figures S1 and S2 demonstrate that the performance of Stable-AGDM is consistently competitive across a range of network scales and types, thereby refuting the assertion of dataset-specific superiority. In Figure S4, Stable-AGDM accomplishes ND by removing an average of 12.74% of nodes. This efficiency is evident in the operational mechanism of Stable-AGDM, which functions as a one-pass method without the necessity of manual feature engineering. In contrast, conventional baselines necessitate iterative structural recomputation or manual feature engineering, thereby compromising efficiency. Extended results can be found in Appendix C.

Conclusion. The present study puts forth a more general deep learning framework, Stable-AGDM, for addressing the ND problem. The distinguishing feature of the proposed Stable-AGDM framework is its composition of three modules, each representing a distinct learning level for the ND problem: graph structure learning, topological correlation learning, and feature selection. As demonstrated in experimental trials, Stable-AGDM has been shown to outperform existing methods and exhibits generalization capacity across a range of ND scenarios. In this regard, it is proposed that a more profound investigation be undertaken into the structural design of module alignment within the ND problem. This investigation will entail the identification of empirical knowledge that is genuinely useful for addressing the problem, with a particular focus on the impact of diverse modules on model performance. This endeavor will facilitate the acquisition of more profound insights. Moreover, observations of low-rank phenomena in the neural feature engineering module further motivate us to investigate their underlying causes through larger-scale experiments and explainable machine learning techniques.

Acknowledgements This work was supported by National Nature Science Foundation of China (Grant No. 61825301).

Supporting information Appendixes A–C. The code and data for this study are available on Gitee at https://gitee.com/yzhangpeng/stable-agdm. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Fan C, Zeng L, Sun Y, et al. Finding key players in complex networks through deep reinforcement learning. Nat Mach Intell, 2020, 2: 317–324
2 Grassia M, De Domenico M, Mangioni G. Machine learning dismantling and early-warning signals of disintegration in complex systems. Nat Commun, 2021, 12: 5190
3 Cappart Q, Chételat D, Khalil E, et al. Combinatorial optimization and reasoning with graph Neural Networks. J Mach Learn Res, 2023, 24: 1–61
4 Xu R, Zhang X, Shen Z, et al. A theoretical analysis on independence-driven importance Weighting for covariate-shift generalization. In: Proceedings of the International Conference on Machine Learning (ICML), Baltimore, 2022. 24803–24829