Facebook Page Classification with Deep Learning Models Proposal

Eileen Ip

Contents

1.0 Introduction

1.1 Context

The content of social media is growing exponentially, putting tremendous pressure to companies such as Meta. One of the main platforms provided by Meta, Facebook, has a huge network of interconnected pages, which represent companies, politicians, television shows, and government organisations. As of 2025, Facebook had over three billion active monthly users that continuously created new pages, posts, and connections (Dixon, 2025). This unprecedented scale complicates research interpretation through manual or simple rule-based classification systems, which are not sufficiently flexible and capable of comprehending connections between data.

1.2 Challenge

The problem is not only a technical issue to classify the Facebook pages, but it also has a direct effect on the user experience, relevance of the ads and brand reputation. Recommendations can be skewed or achieve significant reduction in engagement due to misclassifying political or organisational pages and spread misinformation (White, 2024). Therefore, Meta is particularly motivated to deploy advanced, automated, and interpretable systems to identify page categories correctly and leverage the graph structured data of social connections (Balendra, 2025).

1.3 Deep Learning Suitability

Deep learning is well suited for this problem, more specifically, Graph Neural Networks (GNNs). Unlike traditional machine learning models that analyse each of the pages independently without considering their relationships, GNNs directly learn the graph structured data to learn nodes' attributes and relational dependencies (Thomas Kipf, 2017). GNNs can help Facebook to discern categories and enhance personalised recommendations by learning pages within its neural network.

1.4 Proposal

The proposal outlines a deep learning project which compares modern GNN models, Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs) and Graph Sample and Aggregation Networks (SAGEs), on the Facebook Large Page-Page Network dataset. The project aims to deliver a semi supervised multi-class node classification with reasonable accuracy. This can effectively identify Facebook pages, which will align with the aims of Meta in content moderation, recommendation accuracy, and network interpretability.

2.0 Project Objectives

2.1 Domains Considered

Facebook classify and moderate millions of pages accurately in accordance with both their user characteristics and social connections (Meta, 2023). Traditional classifiers that are based on text or metadata, overlook relational dependencies between pages (Zhang, Shen, Dong, Wang, & Han, 2021), which leads to their generalisation is poor and their explainability is limited. This problem affects multiple domains such as in recommendation systems, content moderation and targeted advertising (Meta, 2023).

2.2 Project Goals

The project will implement deep learning models capable of semi-supervised, multi-class node classification on Facebook's Large Page-Page Network dataset. The primary goals are:

- Develop three scalable GNN architectures to classify nodes into four categories, which include politicians, governmental organisations, television shows, and companies.
- Compare model performance across architectures based on both quantitative metrics (accuracy and loss) and qualitative metrics (training and validation curves and 2D embedding separability plots).
- Demonstrate business applicability by translating model outputs into actionable insights for recommendation ranking, content moderation, and targeted advertising segmentation.
- Evaluate ethical and operational risks (data privacy, algorithmic bias, and interpretability).

2.3 Measurable Outcomes

- Achieve a minimum of 90% classification accuracy on the test data.
- Show strong generalisation on the training and validation curves (for accuracy and loss).
- Generate visually separable clusters in embedding space into four categories.

2.4 Alignment and Feasibility

Each of these objectives will further help the goals of Meta to reduce misinformation, enhance content discovery, and maintain user trust. The project is both technically feasible and can be accomplished using the current open-source libraries and can be done in a typical university semester.

3.0 Methodology

3.1 Data Source

The Facebook Large Page-Page Network dataset is a publicly accessible graph dataset that has (Rozemberczki, Allen, & Sarkar, 2019):

- Nodes: 22,470 Facebook pages.
- Edges: 171,002 undirected connections representing mutual page likes.
- Features: 47,000 high-dimensional word vector representations of page descriptions
- Labels: The four categories, which are politician, governmental organisation, television show, company.

This dataset provides a realistic simulation of Meta's social network structure, enabling extensive investigation without risks of privacy associated with live user data.

3.2 Data Preprocessing

The preprocessing process efficiently prepares data into a usable dataset for this classification project. It is reproducible and does the following steps:

1. Edges, labels and features are read from CSV and JSON files.
2. All node IDs are mapped into contiguous indices to construct tensors.
3. A sparse count matrix is created from feature indices, converted to TF-IDF weights to emphasise informative tokens (Cahyani & Patasik, 2021).
4. Using truncated SVD reduces the number of features to 256 dimensions, retaining variance while reducing computational cost (Brownlee, 2020).
5. L2 normalisation is used to achieve consistent embedding magnitudes.
6. The nodes are randomly divided into 80% training, 10% validation, 10% testing subsets to support semi-supervised learning.

3.3 Model Architecture

The architecture of all models consists of two layers, with ReLU activation and 0.6 dropout to prevent overfitting. Their advantages and explanations are indicated below:

GCNs uses the principles of convolutional neural networks to directly learn the graph structured data (Thomas Kipf, 2017). They leverage the graph's spectral domain to capture the node features' transformation (Thomas Kipf, 2017). GCNs are appropriate for this classification since they aggregate the features from the node's neighbours in the graph, which effectively learning the local structures.

GATs use an attention mechanism for the graph convolution process, to allow nodes to compare their neighbours' impacts based on their significance (Velickovic, et al., 2018). The GATs employ self-attention for each edge to capture the coefficients, which helps adapt connected nodes' impact dynamically. They are suitable for this project since the model improves robustness against noise, improves interpretability and handles heterogenous connectivity through focusing on the related neighbours.

SAGE models are designed for capturing large graphs by sampling each node of a fixed size neighbourhood rather than using the whole graph (Hamilton, Ying, & Leskovec, 2018). It also uses different aggregation functions, such as mean, LSTM and pooling, to merge the information from the sampled neighbours with the target node representation. The SAGE models are applicable for this project because of its scalability and effectiveness in generalising to unseen nodes during the training process.

3.4 Training Process

The training parameters provides balance stability, efficiency, and generalisation on mid-sized graphs. The configuration includes:

- Optimiser: AdamW was selected and used with learning rate of 0.01 and weight decay of 0.0005 for fast convergence.
- Epochs: 300 was used as is a sufficient number of epochs to train the model.
- Early stopping Function: Checks if the validation loss does not improve over a specific number of epochs (the specific number is 80), the training loop terminates early to prevent overfitting.
- Loss Function: The cross-entropy function was used as the loss criterion due to its suitability with multi-class problems. This minimises the loss through aligning its actual class distribution with the estimated probability distribution.
- Scheduler Function: StepLR was utilised to enhanced convergence, improved stability, and more exploration in early stages, to allow it to sweep the parameters more broadly and prevent local minima (Yadav, 2024). With step size of 50 and gamma of 0.5.

The training process consists of:

- Performing a forward pass and the backpropagation.
- The predictions and loss with the cross-entropy criterion are calculated.
- Gradients are reset and computed.
- The optimiser and scheduler update the model parameters according to the gradients.
- The early-stopping function checks based on the validation accuracy.
- The training loss, validation loss, training accuracy and the validation accuracy are saved for tracking progress.

3.5 Evaluation Process

The model performance is assessed on the test dataset (unseen nodes) using both quantitative and qualitative metrics (seen in 4.0 Evaluation). The quantitative metrics will be evaluated with the tr trained model. Two nonlinear dimensionality reduction techniques will provide qualitative validation, which include t-SNE and UMAP. These techniques are displayed with plots to help visualise how well nodes belonging to the same category cluster together

3.6 Ethical and Privacy Considerations

In spite of the fact that this project is based on anonymised information, this project must consider:

- Anonymous user data must be used in order to abide by the Australian Privacy Principles (Australian Government, 2025).
- GNNs could increase structural discrimination (e.g. over-representation of a sensitive topic through affiliated pages), so frequent bias audits will be used.
- The model security should resist against adversarial attacks to avoid constant manual moderation, since these attacks can attempt to invalidate node embeddings.

3.7 Computational Resources

The models will be executed with sufficient computational resources, which needs a standard laptop using GPU (or CPU). The approximate runtime per model is 5 to 10 minutes. The memory consumption is limited, however, with feature reduction the models are able to operate efficiently. This project can be entirely completed in R Studio in R code.

3.8 Alternative Approaches

Other alternatives that may be considered are Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNS), however, GNNs are preferred with its interpretability and suitability with graph structured data (Thomas Kipf, 2017).

4.0 Evaluation

4.1 Quantitative and Qualitative Metrics

The main quantitative metrics are accuracy and loss, since these measures will show the model's precision and reliability. These results are collected during the training and evaluation process. The qualitative metrics include visualisations which are training/validation curves and clusters plots, to visually assess its ability to learn patterns and accurately categorise. The training and validation curves are produced using the quantitative metrics collected during the training process while the visual separability will be built using t-SNE and UMAP plots.

4.2 Interpretability

Accuracy will evaluate correct label prediction and loss is for assessing mistakes. These curves are used to monitor overfitting and convergence stability. The well-formed clusters in the clustering plots denote that the model's ability to learn significant groups in regard to every category of the pages. From a business perspective, these insights will show the clear grouped categories from clusters and detect anomalous communities that indicate of misinformation. This knowledge would help improved recommendation ranking, more effective moderating and enhanced marketing segmentation.

4.3 Baseline Comparison

The initial simple logistic regression as a baseline without encompassing any relational features, is typically unsuitable and inaccurate for graph structured data. Therefore, using different GNN models provide a better advantage, demonstrating the practical benefit of factoring in the network structured data. Such accuracy improvements would show a substantial decrease in false-positive instances and an increase in the precision.

5.0 Timeline

Phase	Weeks	Key Tasks	Deliverables	Roadblocks	Contingency Plans
Planning	1 to 2	Review research on GNNs.	Planned methodology	Over-scope	Limit review to relevant GNN methods.
Data Preparation	3 to 4	Load Facebook dataset and apply preprocessing.	Cleaned processed dataset	Sparse features can cause instability	Try different SVD dimensions (128/256/512).
Task Dependency: Training the model depends on effective feature reduction preprocessing.					
Model Development and Training	5 to 7	Implement GCN, GAT, SAGE architectures and run training process while tuning its hyperparameters.	Trained models	Poor generalisations	Apply early stopping and dropout regularisation.
Task Dependency: The evaluation relies on the accurately trained models.					
Evaluation and Improvement	8	Use the test set and generate t-SNE/UMAP plots, training/validation curves and performance metrics. Edit models, if necessary, when the models achieve low accuracy or poor clustering.	Visual figures and metrics	GPU runtime issues	Change runtime type to CPU, the batch size will be reduced and models pruned.
Task Dependency: Assessing ethics needs all metrics above to determine any potential biases.					
Ethical Assessment & Discussion	9 to 10	Analyse bias, fairness, and governance issues.	Written responses	Ambiguous ethical evidence	Add references about Meta's responsibility with AI.
Reporting and Final Submission	11 to 12	Write and proofread the report.	Final written Report	Missing the deadline	Create schedules to provide sufficient time for editing.

6.0 Conclusion

6.1 Potential Business Impact

The potential business value of implementing GNN-based classification to Meta includes:

- Better content recommendations by understanding network-based relationships, Facebook can offer more personalised and relevant content, enhancing engagement for its users.
- Enhanced content moderation with node embeddings that can identify misinformation in outliers, which minimises manual moderation labour.
- Optimised targeted advertising since page categorisation is improved with accuracy, which allow advertisers to reach audiences more precisely.

6.2 Contribution to Deep Learning

The project contributes to deep learning by:

- Demonstrating how graph-based architectures can be suitable for complex data where basic machine learning models cannot capture.
- Providing a distinct comparison of significant GNN variations.
- Showcasing interpretability to explain hidden patterns through clear quantitative measurements and embedded visualisations as qualitative metrics.

Therefore, this project should be pursed because it contributes to a deeper understanding of advanced deep learning.

6.3 Ethical Reflections

Deep learning systems at a large-scale hold significant social impact. The responsible deployment relies on consistent fairness audits, transparent reporting, and collaboration between engineers, policymakers, and ethicists. With interdisciplinary governance, GNNs can turn nonlinear data into explainable information, these models would help support Meta's mission of making the world a closer place to each other in a responsible manner (Meta, 2023).

6.4 Limitations

Although the models achieved strong accuracy, the following several limitations remain:

- The dataset contains only four categories, which could simplify the classification problem and may not generalise to networks with higher diversity.
- The random split of nodes for training, validation and testing sets does not consider the presence of temporal or structural dependencies that might be exist in real social networks.
- The features are derived only from page descriptions, which can introduce noise or bias if the text is incomplete or inconsistent.
- The models were trained on CPU or GPU, which limit the experimentation of larger architectures when scaling the models in the future.

These limitations should be considered for future improvements to the project to enhance the understanding of the data and improve the models' capabilities.

6.5 Future Works

Future works are suggested at aim at extending this project beyond the current deep learning models, the following improvements can be made:

- The use of heterogeneous graph neural networks that have the capability of integrating multiple types of nodes and relationships, such as user interactions, posts, or shared media content. This could provide a more insightful understanding into user preferences.
- Investigate on other deep learning methods, such as different node representation learning. This could reduce the reliance to use labelled data and allow adaptation to other social platform page data.
- Incorporating time-based data could help capture how relationships evolve over time, providing insights into changing engagement patterns or page popularity.
- Experimenting with more advanced architectures including graph transformers or GNNs with positional encodings may improve the ability to learn more complex dependencies between nodes.

These improvements would allow the models to be more flexiable, scalable, and applicable to a wider range of network analysis tasks in both research and the industry.

6.6 Summary

In this proposal, scalable, interpretable, and ethically aligned GNN architectures are suggested to enhance the systems of Meta to classify and moderate Facebook pages. Our proposal leveraged on the Facebook Large Page-Page Network dataset indicate that deep learning models are superior to traditional classifiers, which produces significant improvements in accuracy and interpretability.

The priorities in Meta for implementations of responsible AI and content governance are consistent with the project's objectives, which include the training, evaluation, and comparison of GCN, GAT, and SAGE architectures. This comprehensive approach integrates advanced data preprocessing, model optimisation, while ensuring that the algorithms are both technically robust and ethically responsible.

The quantitative metrics are accuracy and loss that quantifies the models' effectiveness to classify. While the qualitative include training/validation curves that show the models' ability to capture patterns and visualisations using t-SNE and UMAP provides clusters that represent the grouped categories. The completion period is 12 weeks with the proposed timeline being a feasible and risks assessed (mitigated by the contingency plans) roadmap.

Overall, this proposal should be conducted as this project addresses Meta's classification problem as mentioned before, by implementing these scalable, robust, and ethically aligned deep learning models.

7.0 References

Australian Government. (2025). *Australian Privacy Principles*. Retrieved from Australian Government Office of the Australian Information Commissioner: https://www.oaic.gov.au/privacy/australian-privacy-principles

Balendra, S. (2025). *Meta's AI moderation and free speech: Ongoing challenges in the Global South*. Retrieved from Cambridge University Press: https://www.cambridge.org/core/journals/cambridge-forum-on-ai-law-and-governance/article/metas-ai-moderation-and-free-speech-ongoing-challenges-in-the-global-south/2DB952F896DB5744A43CD3E6C1A6DCB4

Brownlee, J. (2020, August 18). *Singular Value Decomposition for Dimensionality Reduction in Python*. Retrieved from Machine Learning Mastery: https://machinelearningmastery.com/singular-value-decomposition-for-dimensionality-reduction-in-python/

Cahyani, D., & Patasik, I. (2021). *Performance comparison of TF-IDF and Word2Vec models for emotion text classification*. Retrieved from Bulletin of Electrical Engineering and Informatics: https://beei.org/index.php/EEI/article/view/3157

Dixon, S. (2025, October 16). *Most popular social networks worldwide as of February 2025, by number of monthly active users*. Retrieved from Statista: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Hamilton, W., Ying, R., & Leskovec, J. (2018, September 10). *Inductive Representation Learning on Large Graphs*. Retrieved from Cornell University: https://arxiv.org/abs/1706.02216

Meta. (2023, June 29). *The AI behind unconnected content recommendations on Facebook and Instagram*. Retrieved from Meta: https://ai.meta.com/blog/ai-unconnected-content-recommendations-facebook-instagram/

Rozemberczki, B., Allen, C., & Sarkar, R. (2019). *Multi-scale Attributed Node Embedding*. Retrieved from SNAP: https://snap.stanford.edu/data/facebook-large-page-page-network.html

Thomas Kipf, M. W. (2017, February 22). *Semi-Supervised Classification with Graph Convolutional Networks*. Retrieved from Cornell University: https://arxiv.org/abs/1609.02907

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018, February 4). *Graph Attention Networks*. Retrieved from Cornell University: https://arxiv.org/abs/1710.10903

White, K. (2024, June 17). *Digital News Report: Australia 2024: AI, social media, misinformation and distrust – what the data tells us about the news landscape in 2024*. Retrieved from University of Canberra: https://www.canberra.edu.au/about-uc/media/newsroom/2024/june/digital-news-report-australia-2024-ai,-social-media,-misinformation-and-distrust-what-the-data-tells-us-about-the-news-landscape-in-2024

Yadav, A. (2024, October 28). *Guide to Pytorch Learning Rate Scheduling*. Retrieved from Medium: https://medium.com/data-scientists-diary/guide-to-pytorch-learning-rate-scheduling-b5d2a42f56d4

Zhang, Y., Shen, Z., Dong, Y., Wang, K., & Han, J. (2021). *MATCH: Metadata-Aware Text Classification in*. Retrieved from Microsoft Research: https://arxiv.org/pdf/2102.07349