Towards Graph Foundation Models for Dynamics in Complex Networked Systems: Lessons from Super-Spreader Identification in Multilayer Networks

Michał Czuba [*1] **Mateusz Stolarski** [*1] **Adam Piróg** [2] **Piotr Bielak** [1] **Piotr Bródka** [1]

Abstract

Network dynamics — including spreading, influence maximisation, and epidemic modelling — remain largely confined to the transductive paradigm, where models are trained on a single network and cannot be reused on unseen graphs without retraining. We argue that inductive cross-network generalisation is a necessary prerequisite for Graph Foundation Models (GFMs) in this domain and propose four design properties towards this goal. As a proof of concept, *ts-net* (TopSpreadersNetwork), trained solely on synthetic multilayer networks (MLNs), demonstrates zero-shot generalisation to real-world MLNs of varying size and layer count, outperforming classical heuristics and transductive baselines on three of four metrics. Based on *ts-net*'s performance, we further outline five open challenges towards building GFMs for network dynamics: scale, many-layer generalisation, self-supervised pretraining, cross-task transfer, and node-attribute integration.

1. Introduction

Graph Foundation Models (GFMs) have transformed the graph learning domain, with a growing body of work demonstrating zero-shot and few-shot transfer across node and graph classification tasks. The graph machine learning community is beginning to follow suit, with models such as OFA (Liu et al., 2024) and GraphGPT (Tang et al., 2024) providing early evidence of cross-graph generalisation. One corner of the graph machine learning landscape, however, remains largely rooted in the transductive paradigm: *network dynamics* — including spreading processes, influence maximisation, and epidemic modelling (Salehi et al., 2015). These tasks are still primarily solved by fitting one model per dataset, per graph, and per task, with only limited progress towards transferable or inductive solutions.

Multilayer networks (MLNs) (Kivelä et al., 2014) constitute a powerful framework for modelling complex systems with multiple types of interactions, such as social networks with different communication channels (Jankowski et al., 2017), transportation networks with various modes of transit (Cardillo et al., 2013), or brain connectivity networks with distinct functional layers (Wilson et al., 2021). Formally, an MLN is a quadruple $G = ([n], [\ell], V, E)$, where $[n]$ is a set of actors, $[\ell]$ a set of layers, $V \subseteq [n] \times [\ell]$ a set of nodes, and $E \subseteq \bigcup_{i \in [\ell]} V_i \times V_i$ a set of edges; each layer $G_i = (V_i, E_i)$ captures a distinct mode of interaction amongst the same actor set. Their rich structure challenges cross-dataset generalisation: layer counts and types vary across datasets, so that even within-domain transfer is non-trivial — a model trained on a social network with two layers cannot be straightforwardly applied as-is to a transportation network with thirty-seven layers.

In this position paper, we draw on our experience with super-spreader identification in MLNs to argue that **inductive generalisation across networks, while a necessary prerequisite for any GFM, remains largely unmet in the domain of network dynamics**. We provide both a conceptual argument for why this is so and empirical evidence that it is achievable today.

2. The transductive status quo

The dominant paradigm for network dynamics tasks is transductive: a model is fitted to a specific graph and cannot be reused on unseen networks without retraining. This limitation is structural rather than incidental, and it cuts across all classes of existing methods.

Classical heuristics like degree centrality (*deg-c*), neighbourhood size (*nghb-s*), and their discounting variants (Chen et al., 2009; Czuba & Bródka, 2025), carry no learnable parameters and accumulate no knowledge across graphs. They apply fixed formulae to each network's adjacency structure independently and cannot improve with more data.

[*]Equal contribution [1]Department of Artificial Intelligence, Wrocław University of Science and Technology, Wrocław, Poland [2]4Semantics, Warsaw, Poland. Correspondence to: Michał Czuba <michal.czuba@pwr.edu.pl>.

Accepted at *GFM @ ICML 2026* (Workshop on Graph Foundation Models: A New Era for Graph Machine Learning at the 43rd International Conference on Machine Learning), Seoul, South Korea, 2026. Copyright 2026 by the author(s).

Learning-based methods are transductive for a different but equally fundamental reason: they learn representations tied to a specific graph, whether that representation captures its spreading dynamics, its embedding space, or its centrality structure. Compounding this, most operate on single-layer graphs and cannot directly handle the multilayer setting. *deep-im* (Ling et al., 2023) takes the conceptually correct step of internalising spreading dynamics as a learnable, per-graph surrogate (the analogue, at graph scale, of what a GFM spreading model should do at corpus scale) yet re-learns that surrogate from scratch for each new network. PIANO (Li et al., 2023) recasts influence maximisation as a reinforcement learning problem but is primarily evaluated within structurally similar networks, without explicitly addressing cross-graph transfer. Inductive methods exist in the single-layer setting (Panagopoulos et al., 2024), but variable layer counts and layer-type heterogeneity make the multilayer case qualitatively harder. Unsupervised multilayer embedding methods such as Multi-node2vec (Wilson et al., 2021) address the MLN structure and can be combined with a prediction head, e.g. k-means (*mn2v-km*), but remain per-graph, providing no cross-network transfer.

GBIM (Yuan et al., 2024) is a notable exception in addressing a multiplex graph structure, modelling simultaneous multi-information propagation across concurrent layers; nonetheless, it fits a surrogate model per problem instance and transfers nothing across graphs. Models more broadly applicable to ranking in heterogeneous or directed graph settings, such as GNNRank (He et al., 2022), are typically applied in a transductive setting and additionally require adaptation to handle MLNs.

The cost of this paradigm is concrete: limited knowledge transfer across networks, reduced robustness under structural distribution shift, and (for surrogate-based approaches) limited scalability to very large graphs. This is precisely the per-dataset workflow that GFMs have begun to displace in other domains, and we see no fundamental obstacle to the same shift occurring in network dynamics.

3. Desiderata for GFMs in MLNs' dynamics

We propose four design properties satisfied by *ts-net* and put forward as a candidate minimal set for GFMs in this domain; they are not sufficient on their own, and Section 5 addresses what remains beyond them.

***D1*: Relationship-agnostic encoding** Layer and edge types should not require a fixed schema known at training time. *ts-net* achieves this via a shared encoder with no layer-type lookup; schema-dependent encoders fix the vocabulary at training time, precluding transfer to networks with unseen layer types.

***D2*: Size-agnostic inference** The model should generalise to arbitrary actor and layer counts without retraining or architectural modification. Neighbourhood sampling provides actor-side flexibility; adaptive cross-layer aggregation provides layer-side flexibility, enabling zero-shot evaluation on graphs structurally different from those seen during training.

***D3*: Corpus-based training** A GFM must be trained on a *distribution* of graphs rather than a single graph. Even a corpus of synthetic networks constitutes a form of pretraining set, exposing the model to structural diversity (varying density, layer count, and degree distribution) that single-graph training cannot provide.

***D4*: Topology-focused input** When the target dynamic is governed primarily by network structure, as in the Independent Cascade Model (ICM) (Goldenberg et al., 2001), node features may introduce graph-specific signals that hinder cross-network transfer. Zero-feature input (i.e., zero or degree-derived features rather than node-specific metadata) can therefore serve as a principled design choice in such settings; we provide empirical evidence in Section 4.

4. A proof of concept — *ts-net*

We show that *ts-net* (*TopSpreadersNetwork* (Czuba et al., 2025)) can be treated as a proof of concept for the four desiderata introduced in Section 3: trained entirely on synthetic graphs, it demonstrates zero-shot generalisation to unseen real-world MLNs of widely varying size.

ts-net ranks actors by spreading potential under ICM, augmented with a protocol function δ that generalises it to the multirelational setting. Two boundary cases of δ, representing opposite ends of the diffusion-difficulty spectrum, are considered: AND requires an actor to receive activation signals in *all* its layers before becoming active; OR requires activation in *any* single layer. The architecture comprises a shared GAT+GIN encoder (Veličković et al., 2017; Xu et al., 2018) applied identically to every layer; actor-side neighbourhood sampling and a soft-attention cross-layer aggregation (*WiseAverage*) jointly support *D2* by enabling inference on networks of arbitrary actor and layer count. Per-node outputs are equivariant to within-layer node ordering (via GNN message-passing), and WiseAverage is invariant to layer ordering. An MLP prediction head for the downstream task completes the pipeline with $\sim$458,000 parameters in total, with all-zero input feature vectors. Training uses the TopSpreaders dataset: more than 200 synthetic MLNs generated by Erdős–Rényi and preferential-attachment models (Bollobas et al., 2003), with ICM simulated under 40 configurations $\times$ 40 repetitions per actor, so that each actor serves as a sole diffusion seed across different spreading regimes; no real-world network is seen during training.

Table 1. Average T and S across artificial and real-world networks (Czuba et al., 2025). Bold: best per column.

	Artificial		Real	
Method	T	S	T	S
random	0.538	0.759	0.520	0.749
deg-c	0.841	0.872	0.829	0.896
deg-cd	0.830	0.872	0.829	**0.901**
mn2v-km	0.611	0.790	0.586	0.786
deep-im	0.548	0.761	0.643	0.788
ts-net	**0.850**	**0.881**	**0.921**	0.897

Table 2. Cross-regime transfer: *ts-net* under different train/test δ configurations (Czuba et al., 2025).

Train δ	Test δ	Art. T	Real T
AND	AND	0.808	0.899
OR	OR	0.955	0.924
AND	OR	0.982	0.940

The spreading potential score of an actor quantifies its capacity to propagate under ICM when seeded alone; T measures that for the top-ranked actor relative to the optimal seed, and S the analogous score for the top-s super-spreaders (s is the 80th-percentile cutoff); both $\in [0, 1]$, higher is better (Czuba et al., 2025). *ts-net* leads in 3 out of 4 reported metrics (Table 1), trained only on synthetics yet evaluated zero-shot on real-world MLNs, which provides empirical support for *D3* and *D2*. Notably, *deep-im* could not be evaluated on *timik* (Jankowski et al., 2017) (61,702 actors) due to memory exhaustion, a limitation inductive approaches can mitigate. Figure 1 traces the cumulative relative score $y_{rel}(k)$ — the fraction of max. spreading potential recovered by the top-k actors — across budgets, illustrating *ts-net*'s consistent lead and *deep-im*'s collapse on real-world networks.

ts-net, trained exclusively on AND, achieves $T = 0.982$ (artificial) and 0.940 (real) when evaluated under OR — *exceeding* the in-distribution OR→OR performance of 0.955 / 0.924 (Table 2). This cross-regime transfer suggests the model captures structural patterns rather than overfitting to regime-specific diffusion behaviour, corroborating *D4*.

Replacing zero-feature input with centrality features reduces T from $0.808 \rightarrow 0.764$ (artificial) and $0.899 \rightarrow 0.730$ (real), providing empirical support for *D4*. In this setting, where dynamics are driven purely by network structure, node features introduce graph-specific noise, and topology alone is sufficient for cross-network transfer.

5. Open challenges

While the results of *ts-net* provide encouraging evidence for the feasibility of inductive generalisation across MLNs in the domain of spreading dynamics, several obstacles need to be addressed before GFMs become practical in this field.

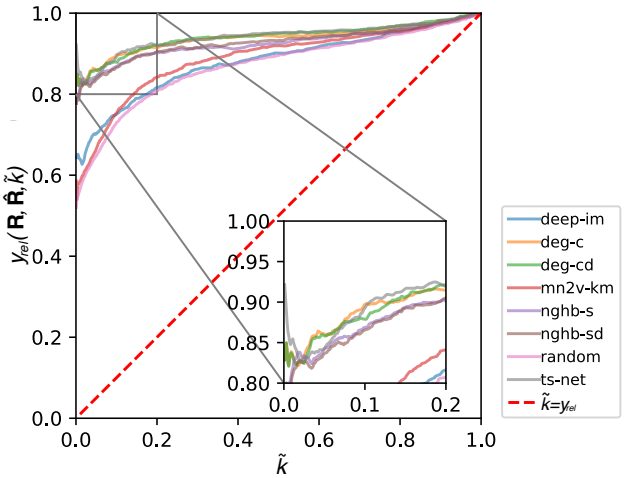

Figure 1. Averaged $y_{rel}(k)$ on real-world networks. *ts-net* leads throughout; *deep-im* degrades to near-random performance on most networks.

Scale The training corpus used here (on the order of 10^2 synthetic networks) is only a first step. Foundation models in other domains rely on orders of magnitude more data. Achieving comparable scale will require both richer synthetic graph generators and the systematic curation of large, diverse repositories of real-world multilayer networks.

Many-layer generalisation *ts-net*'s performance analysis revealed that performance degrades on networks with substantially more relationships than those seen during training (training graphs had at most 5 layers; arxiv and eu-trans have 13 and 37) (Czuba et al., 2025). This suggests that actor-side scalability via neighbourhood sampling must be complemented by layer-side strategies, such as chunk-based or hierarchical processing of layers.

Self-supervised pretraining The discussed approach has been evaluated in a fully supervised manner and depends on costly simulations. Self-supervised objectives, for example contrastive learning or diffusion-prediction surrogates defined on unlabelled graph corpora, could provide a scalable alternative and reduce reliance on expensive labels.

Cross-task transfer *ts-net* is specialised to super-spreader ranking under ICM. A true GFM should support transfer across related tasks, including source detection, epidemic forecasting, alternative diffusion models (e.g., LTM), and seed set evaluation.

Rich node attributes Real-world networks often include textual, tabular, or temporal node attributes. Integrating such signals without sacrificing cross-network generalisation re-

mains an open question. While topology-only input may be preferable in settings where the dynamics are purely structural, combining structural and attribute-based information in a transferable manner remains an important direction.

6. Conclusion

In this article, we argue that inductive generalisation across networks, while a necessary prerequisite for any GFM, remains largely unmet in the domain of network dynamics, and show, through *ts-net*, that such behaviour is achievable in practice, even in the more challenging setting of MLNs. We hope this work motivates the development of scalable, pre-trained, and transferable models for dynamics in complex networked systems, bringing this domain closer to the foundation model paradigm.

References

Bollobas, B., Borgs, C., Chayes, J., and Riordan, O. Directed scale-free graphs. In *Proceedings of the 14th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 132–139, 1 2003.

Cardillo, A., Gómez-Gardeñes, J., Zanin, M., Romance, M., Papo, D., Pozo, F. d., and Boccaletti, S. Emergence of network features from multiplexity. *Scientific Reports*, 3 (1344):1–6, 2 2013. ISSN 2045-2322.

Chen, W., Wang, Y., and Yang, S. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD internat. conf. on Knowledge discovery and data mining*, pp. 199–208, 2009.

Czuba, M. and Bródka, P. Rank-refining seed selection methods for budget constrained influence maximisation in multilayer networks under linear threshold model. *Social Network Analysis and Mining*, 15(1):46–25, April 2025. ISSN 1869-5469.

Czuba, M., Stolarski, M., Piróg, A., Bielak, P., and Bródka, P. Identifying super spreaders in multilayer networks [preprint with appendices]. *arXiv*, May 2025. doi: 10.48550/arXiv.2505.20980.

Goldenberg, J., Libai, B., and Muller, E. Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, 12(3):211–223, August 2001. ISSN 1573-059X.

He, Y., Gan, Q., Wipf, D., Reinert, G. D., Yan, J., and Cucuringu, M. Gnnrank: Learning global rankings from pairwise comparisons via directed gnns. In *international conference on machine learning*, 2022.

Jankowski, J., Michalski, R., and Bródka, P. A multilayer network dataset of interaction and influence spreading in a virtual world. *Scientific Data*, 4(1):170144, 2017. ISSN 2052-4463.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y., and Porter, M. A. Multilayer networks. *Journal of Complex Networks*, 2(3):203–271, 07 2014. ISSN 2051-1310.

Li, H., Xu, M., Bhowmick, S. S., Rayhan, J. S., Sun, C., and Cui, J. Piano: Influence maximization meets deep reinforcement learning. *IEEE Transactions on Computational Social Systems*, 10(3):1288–1300, 2023. doi: 10.1109/TCSS.2022.3164667.

Ling, C., Jiang, J., Wang, J., Thai, M. T., Xue, L., Song, J., Qiu, M., and Zhao, L. Deep graph representation learning and optimization for influence maximization. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Liu, H., Feng, J., Kong, L., Liu, N., Tao, D., Chen, Y., and Zhang, M. One for all: Towards training one graph model for all classification tasks. In *The Twelfth International Conference on Learning Representations*, 2024.

Panagopoulos, G., Tziortziotis, N., Vazirgiannis, M., Pang, J., and Malliaros, F. D. Learning graph representations for influence maximization. *Social Network Analysis and Mining*, 14(1):203, October 2024. ISSN 1869-5469. doi: 10.1007/s13278-024-01311-z.

Salehi, M., Sharma, R., Marzolla, M., Magnani, M., Siyari, P., and Montesi, D. Spreading processes in multilayer networks. *IEEE Transactions on Network Science and Engineering*, 2(2):65–83, 2015.

Tang, J., Yang, Y., Wei, W., Shi, L., Su, L., Cheng, S., Yin, D., and Huang, C. GraphGPT: Graph instruction tuning for large language models. In *Proc. of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 491–500, 2024.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Wilson, J. D., Baybay, M., Sankar, R., Stillman, P., and Popa, A. M. Analysis of population functional connectivity data via multilayer network embeddings. *Network Science*, 9 (1):99–122, 2021.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

Yuan, Z., Shao, M., and Chen, Z. Graph Bayesian Optimization for Multiplex Influence Maximization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, March 2024.