

Full Length Article



Weisfeiler–Lehman goes dynamic: An analysis of the expressive power of Graph Neural Networks for attributed and dynamic graphs[☆]

Silvia Beddar-Wiesing^{a,*}, Giuseppe Alessio D’Inverno^{b,1}, Caterina Graziani^{b,1},
Veronica Lachi^{b,1}, Alice Moallem-Oureh^{a,1}, Franco Scarselli^b, Josephine Maria Thomas^a

^a *Graphs in Artificial Intelligence and Neural Networks (GAIN), University of Kassel, Germany*

^b *Siena Artificial Intelligence Lab (SAILab), University of Siena, Italy*

ARTICLE INFO

Dataset link: <https://github.com/AleDinve/dyn-gnn>

Keywords:

Graph Neural Network
Dynamic graphs
GNN expressivity
Unfolding trees
Weisfeiler–Lehman test

ABSTRACT

Graph Neural Networks (GNNs) are a large class of relational models for graph processing. Recent theoretical studies on the expressive power of GNNs have focused on two issues. On the one hand, it has been proven that GNNs are as powerful as the Weisfeiler–Lehman test (1-WL) in their ability to distinguish graphs. Moreover, it has been shown that the equivalence enforced by 1-WL equals unfolding equivalence. On the other hand, GNNs turned out to be universal approximators on graphs modulo the constraints enforced by 1-WL/unfolding equivalence. However, these results only apply to Static Attributed Undirected Homogeneous Graphs (SAUHG) with node attributes. In contrast, real-life applications often involve a much larger variety of graph types. In this paper, we conduct a theoretical analysis of the expressive power of GNNs for two other graph domains that are particularly interesting in practical applications, namely dynamic graphs and SAUHG with edge attributes. Dynamic graphs are widely used in modern applications; hence, the study of the expressive capability of GNNs in this domain is essential for practical reasons and, in addition, it requires a new analyzing approach due to the difference in the architecture of dynamic GNNs compared to static ones. On the other hand, the examination of SAUHG is of particular relevance since they act as a standard form for all graph types: it has been shown that all graph types can be transformed without loss of information to SAUHG with both attributes on nodes and edges. This paper considers generic GNN models and appropriate 1-WL tests for those domains. Then, the known results on the expressive power of GNNs are extended to the mentioned domains: it is proven that GNNs have the same capability as the 1-WL test, the 1-WL equivalence equals unfolding equivalence and that GNNs are universal approximators modulo 1-WL/unfolding equivalence. Moreover, the proof of the approximation capability is mostly constructive and allows us to deduce hints on the architecture of GNNs that can achieve the desired approximation.

1. Introduction

Graph data is becoming pervasive in many application domains, such as biology, physics, and social network analysis (Kazemi et al., 2020; Skardinga, Gabrys, & Musial, 2021). Graphs are handy for complex data since they allow for naturally encoding information about

entities, their links, and their attributes. In modern applications, several different types of graphs are commonly used and possibly combined: graphs can be homogeneous or heterogeneous, directed or undirected, have attributes on nodes and/or edges, and be static or dynamic hyper- or multigraphs (Thomas, Beddar-Wiesing, & Moallem-Oureh, 2021). Considering the diversity of graph types, it has recently

[☆] This work was partially supported by the Ministry of Education and Research Germany (BMBF, grant number 01IS20047A) and partially by INdAM GNCS group.

* Corresponding author.

E-mail addresses: s.beddarwiesing@uni-kassel.de (S. Beddar-Wiesing), giuseppealessio.d@student.unisi.it (G.A. D’Inverno), caterina.graziani@student.unisi.it (C. Graziani), veronica.lachi@student.unisi.it (V. Lachi), amoallem@uni-kassel.de (A. Moallem-Oureh), franco@diism.unisi.it (F. Scarselli), jthomas@uni-kassel.de (J.M. Thomas).

URLs: <http://www.gain-group.de> (S. Beddar-Wiesing), <https://sailab.diism.unisi.it/people/giuseppe-alessio-dinverno/> (G.A. D’Inverno), <https://sailab.diism.unisi.it/people/caterina-graziani/> (C. Graziani), <https://sailab.diism.unisi.it/people/veronica-lachi/> (V. Lachi), <http://www.gain-group.de> (A. Moallem-Oureh), <https://sailab.diism.unisi.it/people/franco-scarselli/> (F. Scarselli), <http://www.gain-group.de> (J.M. Thomas).

¹ These authors contributed equally.

<https://doi.org/10.1016/j.neunet.2024.106213>

Received 27 January 2023; Received in revised form 5 February 2024; Accepted 23 February 2024

Available online 28 February 2024

0893-6080/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

been shown that Static Attributed Undirected Homogenous Graphs (SAUHG) with both attributes on nodes and edges can act as a standard form for graph representation, namely that all the common graph types can be transformed into those SAUHG without losing their encoded information (Thomas et al., 2021).

In the field of graph learning, Graph Neural Networks (GNNs) have become a prominent class of models used to process graphs and address different learning tasks directly. For this purpose, most GNN models adopt a computational scheme based on a local aggregation mechanism. It recursively updates the local information of a node stored in an attribute vector by aggregating the attributes of neighboring nodes. After several iterations, the attributes of a node capture the local structural information received from its k -hop neighborhood. At the end of the iterative process, the obtained node attributes can address different graph-related tasks by applying a suitable readout function.

The Graph Neural Network model of Scarselli, Gori, Tsoi, Hagenbuchner, and Monfardini (2009), which is called Original GNN (OGNN) in this paper, was the first model capable of facing both node/edge and graph-focused tasks utilizing suitable aggregation and readout functions, respectively. In the current research, a large number of new applications and models have been proposed, including Neural Networks for graphs (Micheli, 2009), Gated Sequence Graph Neural Networks (Li, Tarlow, Brockschmidt, & Zemel, 2016), Spectral Networks (Bruna, Zaremba, Szlam, & LeCun, 2014), Graph Convolutional Neural Networks (Kipf & Welling, 2017), GraphSAGE (Hamilton, Ying, & Leskovec, 2017a), Graph Attention Networks (Veličković et al., 2018), and Graph Networks (Battaglia et al., 2018). Another extension of GNNs consisted of the proposal of several new models capable of dealing with dynamic graphs (Kazemi et al., 2020), which can be used, e.g., to classify sequences of graphs, sequences of nodes in a dynamic graph or to predict the appearance of an edge in a dynamic graph.

Recently, a great effort has been dedicated to studying the theoretical properties of GNNs (Jegelka, 2022). Such a trend is motivated by the attempt to derive the foundational knowledge required to design reasonable solutions efficiently for many possible applications of GNNs (Zhou et al., 2020). A particular interest lies in the study of the expressive power of GNN models since, in many application domains, the performance of a GNN depends on its capability to distinguish different graphs. For example, in Bioinformatics, the properties of a chemical molecule may depend on the presence or absence of small substructures; in social network analysis, the count of triangles allows us to evaluate the presence and the maturity of communities. Similar examples exist in dynamic domains: the distinction of substructures contributes significantly to the successful analysis of molecular conformations (Luo, Shi, Xu, & Tang, 2021); in studies of the evolution of social networks (Deng, Rangwala, & Ning, 2019), substructures in the form of contextual knowledge can help to improve performance.

Formally, a central result on the expressive power of GNNs has shown that GNNs are at most as powerful as the Weisfeiler–Lehman graph isomorphism test (1-WL) (Leman & Weisfeiler, 1968; Shervashidze, Schweitzer, Van Leeuwen, Mehlhorn, & Borgwardt, 2011; Xu, Hu, Leskovec, & Jegelka, 2019). The 1-WL test iteratively divides graphs into groups of possibly isomorphic² graphs using a local aggregation schema. Therefore, the 1-WL test exactly defines the classes of graphs that GNNs can recognize as non-isomorphic.

Furthermore, it has been proven that the equivalence classes induced by the 1-WL test are equivalent to the ones obtained from the unfolding trees of nodes on two graphs (D’Inverno, Bianchini, Sampoli, & Scarselli, 2021; Krebs & Verbitsky, 2015), and hence, 1-WL and the unfolding equivalences can be used interchangeably. An unfolding tree rooted at a node is constructed by starting at the root node and unrolling the graph along the neighboring nodes until it reaches a

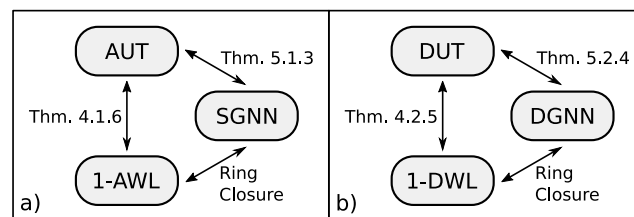


Fig. 1. (a) In Theorem 4.1.6, we prove the equivalence of the attributed unfolding tree equivalence (AUT) and the attributed 1-WL equivalence (1-AWL) for SAUHG. Afterward in Theorem 5.1.3, we show a result on the approximation capability of static GNNs for SAUHG (SGNN) using the AUT equivalence. (b) Analogously to the attributed case, we show similar results for Dynamic GNNs (DGNN) which can be used on temporal graphs.

certain depth. If the unfolding trees of two nodes are equal in the limit, the nodes are called unfolding tree equivalent.³ Fig. 1 visualizes the relations between the 1-WL, the unfolding tree equivalences, and the GNN expressiveness.

Another research goal of the expressive power of GNNs is to study the GNN approximation capability. Formally, it has been proven in Scarselli, Gori, Tsoi, Hagenbuchner, and Monfardini (2008) that OGNNs can approximate in probability, up to any degree of precision, any measurable function $\tau(G, v) \rightarrow \mathbb{R}^m$ that respects the unfolding equivalence. Such a result has been recently extended to a large class of GNNs (D’Inverno et al., 2021) called message-passing GNNs, including most contemporary architectures.

Despite the availability of the mentioned results on the expressive power of GNNs, their application is still limited to undirected static graphs with attributes on nodes. This limitation is particularly restrictive since modern applications usually involve more complex data structures, such as heterogeneous, directed, dynamic graphs and multigraphs. In particular, the ability to process dynamic graphs is progressively gaining significance in many fields such as social network analysis (Deng et al., 2019), recommender systems (Rossi et al., 2020; Xu, Ruan, Korpeoglu, Kumar, & Achan, 2020), traffic forecasting (Yu, Yin, & Zhu, 2017) and knowledge graph completion (Trivedi, Dai, Wang, & Song, 2017; Wu, Cao, Cheung, & Hamilton, 2020). Several surveys discuss the usage of dynamic graphs in other application domains (Barros, Mendonça, Vieira, & Ziviani, 2021; Longa et al., 2023; Skardinga et al., 2021; Thomas, Moallem-Oureh, Beddar-Wiesing, & Holzhueter, 2022; Xue et al., 2022).

Although GNNs are considered universal approximators on the extended domains, it is uncertain which GNN architectures contribute to such universality. Moreover, an open question is how the definition of the 1-WL test has to be modified to cope with novel data structures and whether the universal results fail for particular graph types.

In this paper, we propose a study on the expressive power of GNNs for two domains of particular interest, namely dynamic graphs and static attributed undirected homogeneous graphs (SAUHG) with node and edge attributes. On the one hand, dynamic graphs are interesting from a practical and a theoretical point of view and are used in several application domains (Kazemi et al., 2020). Moreover, dynamic GNN models are structurally different from GNNs for static graphs, and the results and methodology required to analyze their expressive power cannot directly be deduced from existing literature. On the other hand, SAUHG with node and edge attributes are interesting because, as mentioned above, they act as a standard form for several other types of graphs that can all be transformed to SAUHG (Thomas et al., 2021).

² It is worth noting that the 1-WL test is inconclusive since there exist pairs of graphs that the test recognizes as isomorphic even if they are not.

³ Currently, the concept underlying unfolding trees is widely used to study the GNN expressiveness, even if they are mostly called *computation graphs* (Garg, Jegelka, & Jaakkola, 2020).

First, we introduce appropriate versions of the 1-WL test and the unfolding equivalence to construct the fundamental theory for the domains of SAUHGs and dynamic graphs and discuss their relation afterward. Then, we consider generic GNN models that can operate on both domains and prove their universal approximation capability modulo the aforementioned 1-WL/unfolding equivalences. More precisely, the main contributions of this paper are as follows.

- We present new versions of the 1-WL test and of the unfolding equivalence appropriate for dynamic graphs and SAUHGs with node and edge attributes, and we show that they induce the same equivalences on nodes. Such a result makes it possible to use them interchangeably to study the expressiveness of GNNs.
- We show that generic GNN models for dynamic graphs and SAUHGs with node and edge attributes are capable of approximating, in probability and up to any precision, any measurable function on graphs that respects the 1-WL/unfolding equivalence.
- The result on approximation capability holds for graphs with unconstrained attributes of reals and target functions. Thus, most of the domains used in practical applications are included.
- Moreover, the proof is based on space partitioning, which allows us to deduce information about the GNN architecture that can achieve the desired approximation.
- We validate our theoretical results conducting an experimental validation. Our setups show that (1) sufficiently powerful DGNNs can approximate well dynamic systems that preserve the unfolding equivalence and (2) using non-universal architectures can lead to poor performances.

The rest of the paper is organized as follows. Section 2 illustrates the related literature. In Section 3, the notation used throughout the paper is described, and the main definitions are introduced. In Section 4, we introduce the 1-WL test and unfolding equivalences suitable for dynamic graphs and SAUHGs with node and edge attributes and prove that those equivalences are equal. In Section 5, the approximation theorems for GNNs on both graph types are presented. We support our theoretical findings by setting up synthetic experiments in 6. Finally, Section 7 includes our conclusions and future matter of research. All of the proofs are collected in Appendix A.

2. Related work

In the seminal work (Xu et al., 2019), it has been proven that standard GNNs have the same expressive power as the 1-WL test. To overcome such a limitation, new GNN models and variants of the WL test have been proposed. For example, in You, Gomes-Selman, Ying, and Leskovec (2021), a model is introduced where node identities are directly injected into the aggregate functions. In Morris et al. (2019), the k -WL test has been taken into account to develop a more powerful GNN model, given its greater capability to distinguish non-isomorphic graphs. In Bodnar, Frasca, Wang, et al. (2021), a simplicial-wise Weisfeiler–Lehman test is introduced, and it is proven to be strictly more powerful than the 1-WL test and not less powerful than the 3-WL test; a similar test (called cellular WL test), proposed in Bodnar, Frasca, Otter, et al. (2021), is proven to be more powerful than the simplicial version. Nevertheless, all these tests do not deal with edge-attributed graphs and not with dynamic graphs.

In Barcelo, Galkin, Morris, and Orth (2022), the authors consider a GNN model for multi-relational graphs where edges are labeled with types: it is proven that such a model has the same computation capability as a corresponding WL-test. The result is similar to our result on SAUGH, but the way we aggregate the message on the edges is different. Additionally, our work extends (Barcelo et al., 2022) from several points of view: the approximation capability of GNNs is studied, a relationship between the 1-WL test and unfolding trees is established, and edge attributes include more general vectors of reals. Moreover, all those studies are extended to dynamic graphs.

Moreover, the WL test mechanism applied to GNNs has also been studied within the paradigm of *unfolding trees* (Sato, 2020; Zhang & Li, 2021). An equivalence between the two concepts has been established by Krebs and Verbitsky (2015), but it is limited to static graphs without edge attributes.

Some studies have been dedicated to the approximation and generalization properties of Graph Neural Networks. In Scarselli et al. (2008), the authors proved the universal approximation properties of the original Graph Neural Network model, modulo the unfolding equivalence. The universal approximation is shown for GNNs with random node initialization in Abboud, Ceylan, Grohe, and Lukasiewicz (2021), while, in Xu et al. (2019), GNNs are shown to be able to encode any graph with countable input features. Moreover, the authors of Dasoulas, Santos, Scaman, and Virmaux (2020) proved that GNNs, provided with a colored local iterative procedure (CLIP), can be universal approximators for functions on graphs with node attributes. The approximation property has also been extended to Folklore Graph Neural Networks in Maron, Ben-Hamu, Serviansky, and Lipman (2019) and Linear Graph Neural Networks, and general GNNs in Azizian and Lelarge (2021) and Maron, Ben-Hamu, Shamir, and Lipman (2019), both in the invariant and equivariant case. Recently, the universal approximation theorem has been proved for modern message-passing graph neural networks by D’Inverno et al. (2021) giving a hint on the properties of the network architecture (e.g., the number of layers, the characterization of the aggregation function). A relation between the graph diameter and the computational power of GNNs has been established in Loukas (2020), where the GNNs are assimilated to the so-called LOCAL models and it is proved that a GNN with a number of layers larger than the diameter of the graph can compute any Turing function of the graph. Nevertheless, no information on the aggregation function characterization is given.

Despite the availability of universal approximation theorems for static graphs with node attributes, the theory lacks results about the approximation capability of other types of graphs, such as dynamic graphs and graphs with attributes on both nodes and edges. Therefore, this paper aims to extend the results of the expressive power of GNNs for dynamic graphs and SAUHGs with node and edge attributes.

3. Notation and preliminaries

Before extending the work about the expressive power of GNNs to dynamic and edge-attributed graph domains, the mathematical notation and preliminary definitions are given in this chapter. In this paper, only finite graphs are under consideration.

\mathbb{N}	natural numbers
\mathbb{N}_0	natural numbers starting at 0
\mathbb{R}	real numbers
\mathbb{R}^k	\mathbb{R} vector space of dimension k
\mathbb{N}^k	\mathbb{N} vector space of dimension k
\mathbb{Z}^k	\mathbb{Z} vector space of dimension k
$\mathbf{0} = (0, \dots, 0)^\top$	zero vector of corresponding size
$ a $	absolute value of a real a
$\ \cdot\ $	norm on \mathbb{R}
$\ \cdot\ _p$	p -norm on \mathbb{R}
$\ \cdot\ _\infty$	∞ -norm on \mathbb{R}
$ M $	number of elements of a set M
$[n], n \in \mathbb{N}$	sequence $1, 2, \dots, n$
$[n]_0, n \in \mathbb{N}_0$	sequence $0, 1, \dots, n$
\emptyset	empty set
\perp	undefined; non-existent element
$\{\cdot\}$	set
$\{\!\cdot\!\}$	multiset, i.e. set allowing multiple appearances of entries
$(x_i)_{i \in I}$	vector of elements x_i for indices in set I
$[v w]$	stacking of vectors v, w

\wedge	conjunction
\cup	union of two (multi)sets
\subset	sub(multi)set
\subsetneq	proper sub(multi)set
\times	factor set of two sets

The following definition introduces a static, node/edge attributed, undirected and homogeneous graph called SAUHG. The reason for defining and using it comes from [Thomas et al. \(2021\)](#). Here, it is shown that every graph type is bijectively transformable into each other. Therefore, SAUHGs will be used as a standard form for all structurally different graph types as directed or undirected simple graphs, hypergraphs, multigraphs, heterogeneous or attributed graphs, and any composition of those. For a detailed introduction, see [Thomas et al. \(2021\)](#).

Definition 3.1 (Static Attributed Undirected Homogeneous Graphs). G' is called **static, node/edge attributed, undirected, homogeneous graph (SAUHG)** if $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega')$, with $\mathcal{V}' \subset \mathbb{N}$ is a finite set of nodes, $\mathcal{E}' \subset \{\{u, v\} \mid \forall u, v \in \mathcal{V}'\}$ is a finite set of edges and node and edge attributes are determined by the mappings $\alpha' : \mathcal{V}' \rightarrow A$, $\omega' : \mathcal{E}' \rightarrow B$ that map into the arbitrary node attribute set A , and edge attribute set B . The domain of SAUHGs will be denoted as \mathcal{G}' .

Remark 3.2 (Attribute sets). In the above definition of the SAUHG, the node and edge attribute sets A and B can be arbitrary. However, without loss of generality, we can assume that the attribute sets are equal because they can be arbitrarily extended to $A' = A \cup B$. Additionally, any arbitrary attribute set A' can be embedded into the k -dimensional vector space \mathbb{R}^k . Since the attribute sets in general do not matter for the theories in this paper and to support a better readability, in what follows we consider $\alpha' : \mathcal{V}' \rightarrow \mathbb{R}^k$, $\omega' : \mathcal{E}' \rightarrow \mathbb{R}^k$ for every SAUHG.

All the aforementioned graph types are static, but temporal changes play an essential role in learning on graphs representing real-world applications; thus, dynamic graphs are defined in the following. In particular, the dynamic graph definition here consists of a discrete-time representation.

Definition 3.3 (Dynamic Graph). Let $I = [0, \dots, l] \subsetneq \mathbb{N}_0$ be a set of timesteps. Then a **(discrete) dynamic graph** can be considered as a vector of static graph snapshots, i.e., $G = (G_t)_{t \in I}$, where $G_t = (\mathcal{V}_t, \mathcal{E}_t) \forall t \in I$. Furthermore,

$$\alpha_v(t) := \alpha(v, t), \quad v \in \mathcal{V}_t \quad \text{and}$$

$$\omega_{\{u,v\}}(t) := \omega(\{u, v\}, t), \quad \{u, v\} \in \mathcal{E}_t$$

where $\alpha : \mathcal{V} \times I \rightarrow A$ and $\omega : \mathcal{E} \times I \rightarrow B$ define the vector of **dynamic node/edge attributes**. Further, let $\mathcal{V} := \bigcup_{t \in I} \mathcal{V}_t$ and $\mathcal{E} := \bigcup_{t \in I} \mathcal{E}_t$ the total node and edge set of the dynamic graph.

In particular, when a node v does not exist, its attributes and neighborhood are empty, respectively. Moreover, let

$$\Omega_{ne_v}(t) = \left(\omega_{\{v,x_1\}}(t), \dots, \omega_{\{v,x_{|ne_v(t)|}\}}(t) \right)_{t \in I}$$

be the sequence of **dynamic edge attributes of the neighborhood** corresponding node at each timestep. Note that, as in [Remark 3.2](#), in what follows, we assume the attribute sets to be equal and corresponding to \mathbb{R}^k .

To prove the approximation theorems for SAUHGs and dynamic graphs, we need to specify the GNN architectures capable of handling those graph types. The standard Message-Passing GNN for static node-attributed undirected homogeneous graphs is given in [Xu et al. \(2019\)](#). Here, the node attributes are used as the initial representation and

input to the GNN. The update is executed by aggregation over the representations of the neighboring nodes (see [Fig. 2](#)).

Given that a SAUHG acts as a standard form for all graph types, the ordinary GNN architecture will be extended to take also edge attributes into account. This can be done by analogously including the edge attributes in the first iteration to the processing of the node information in the general GNN framework as follows.

Definition 3.4 (SGNN). For a SAUHG $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega')$ let $u, v \in \mathcal{V}'$ and $e = \{u, v\}$. The SGNN propagation scheme for iteration $i \in [l]$, $l > 0$ is defined as

$$\mathbf{h}_v^{(i)} = \text{COMBINE}^{(i)} \left(\mathbf{h}_v^{(i-1)}, \text{AGGREGATE}^{(i)} \left(\{\mathbf{h}_u^{(i-1)}\}_{u \in ne_v}, \{\omega_{\{u,v\}}\}_{u \in ne_v} \right) \right)$$

The output for a node-specific learning problem after the last iteration l respectively is given by

$$\mathbf{z}_v = \text{READOUT}(\mathbf{h}_v^{(l)}),$$

using a selected aggregation scheme and a suitable READOUT function, and the output for a graph-specific learning problem is determined by

$$\mathbf{z} = \text{READOUT}(\{\mathbf{h}_v^{(l)} \mid v \in \mathcal{V}'\}).$$

For the dynamic case, we chose a widely used GNN model that is consistent with the theory we built. Based on [Skardinga et al. \(2021\)](#), the discrete dynamic graph neural network (DGNN) uses a GNN to encode each graph snapshot. Here, the model is modified by using the previously defined SGNN in place of the standard one.

Definition 3.5 (Discrete DGNN). Given a discrete dynamic graph $G = (G_t)_{t \in I}$, a **discrete DGNN** using a continuously differentiable recursive function f for temporal modeling can be expressed as:

$$\begin{aligned} \mathbf{h}_1(t), \dots, \mathbf{h}_n(t) &:= \text{SGNN}(G_t) \quad \forall t \geq 0 \\ \mathbf{q}_1(0), \dots, \mathbf{q}_n(0) &:= \text{SGNN}(G_0) \\ \mathbf{q}_v(t) &:= f(\mathbf{q}_v(t-1), \mathbf{h}_v(t)) \quad \forall v \in \mathcal{V} \end{aligned} \quad (1)$$

where $\mathbf{h}_v(t) \in \mathbb{R}^r$ is the hidden representation of node v at time t of dimension r and $\mathbf{q}_v(t) \in \mathbb{R}^s$ is an s -dimensional hidden representation of node v produced by f , and $f : \mathbb{R}^s \times \mathbb{R}^r \rightarrow \mathbb{R}^s$ is a neural architecture for temporal modeling (in the methods surveyed in [Skardinga et al. \(2021\)](#), f is almost always an RNN or an LSTM).

The stacked version of the discrete DGNN is then:

$$\begin{aligned} H(t) &= \text{SGNN}(G_t) \\ Q(0) &= H(0) = \text{SGNN}(G_0) \\ Q(t) &= F(Q(t-1), H(t)) \end{aligned} \quad (2)$$

where $H(t) \in \mathbb{R}^{n \times r}$, $Q(t) \in \mathbb{R}^{n \times s}$, $F : \mathbb{R}^{n \times s} \times \mathbb{R}^{n \times r} \rightarrow \mathbb{R}^{n \times s}$, being n the number of nodes, r and s the dimensions of the hidden representation of a node produced respectively by the SGNN and by the f . Applying F corresponds to component-wise applying f for each node ([Skardinga et al., 2021](#)).

To conclude, a function $\text{READOUT}_{\text{dyn}}$ will take as input $Q(t)$ and gives a suitable output for the considered task, so that altogether the DGNN will be described as

$$\varphi(t, G, v) = \text{READOUT}_{\text{dyn}}(Q(t)).$$

Remark 3.6. The DGNN is a Message-Passing model because the SGNN is one, by definition.

The GNNs expressivity is studied in terms of their capability to distinguish two non-isomorphic graphs.

Definition 3.7 (Graph Isomorphism). Let G_1 and G_2 be two static graphs, then $G_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2)$ are **isomorphic** to each other $G_1 \approx G_2$, if and only if there exists a bijective function $\phi : \mathcal{V}_1 \rightarrow \mathcal{V}_2$, with

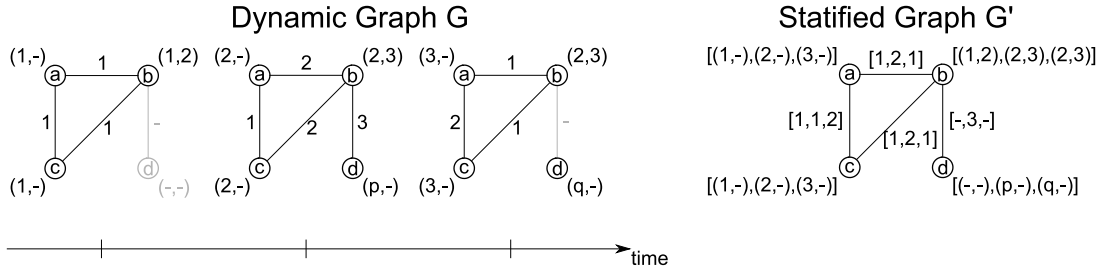


Fig. 2. Illustration of the statification of a dynamic graph. On the left, the temporal evolution of a graph, including non-existent nodes and edges (gray), is given, and on the right, the corresponding statified graph with the total amount of nodes and edges together with the concatenated attributes is shown.

1. $v_1 \in \mathcal{V}_1 \Leftrightarrow \phi(v_1) \in \mathcal{V}_2 \quad \forall v_1 \in \mathcal{V}_1$,
2. $\{v_1, v_2\} \in \mathcal{E}_1 \Leftrightarrow \{\phi(v_1), \phi(v_2)\} \in \mathcal{E}_2 \quad \forall \{v_1, v_2\} \in \mathcal{E}_1$

In case the two graphs are **attributed**, i.e., $G_1 = (\mathcal{V}_1, \mathcal{E}_1, \alpha_1, \omega_1)$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2, \alpha_2, \omega_2)$, then $G_1 \approx G_2$ if and only if additionally there exist bijective functions $\varphi_\alpha : A_1 \rightarrow A_2$ and $\varphi_\omega : B_1 \rightarrow B_2$ with images $A_i := \text{im}(\alpha_i)$ and $B_i := \text{im}(\omega_i)$, $i = 1, 2$.

1. $\varphi_\alpha(\alpha_1(v_1)) = \alpha_2(\phi(v_1)) \quad \forall v_1 \in \mathcal{V}_1$,
2. $\varphi_\omega(\omega_1(\{u_1, v_1\})) = \omega_2(\{\phi(u_1), \phi(v_1)\}) \quad \forall \{u_1, v_1\} \in \mathcal{E}_1$.

If the two graphs are **dynamic**, they are called to be isomorphic if and only if the static graph snapshots of each timestep are isomorphic.

Graph isomorphism (GI) gained prominence in the theory community when it emerged as one of the few natural problems in the complexity class NP that could neither be classified as being hard (NP-complete) nor shown to be solvable with an efficient algorithm (that is, a polynomial-time algorithm) (Grohe & Schweitzer, 2020). Indeed it lies in the class NP-Intermediate. However, in practice, the so-called Weisfeiler–Lehman (WL) test is used to at least recognize non-isomorphic graphs (Sato, 2020). If the WL test outputs two graphs as isomorphic, the isomorphism is likely but not given for sure.

The expressive power of GNNs can also be approached from the point of view of their approximation capability.

It generally analyzes the capacity of different GNN models to approximate arbitrary functions (Bronstein, Bruna, Cohen, & Velickovic, 2021). Different universal approximation theorems can be defined depending on the model, the considered input data, and the sets of functions. This paper will focus on the set of functions that preserve the unfolding tree equivalence, defined as follows.

Since the results in this section hold for undirected and unattributed graphs, we aim to extend the universal approximation theorem to GNNs working on SAUHGs (cf. Definition 3.1) and dynamic graphs (cf. Definition 3.3). For this purpose, in the next sections, we introduce a static attributed and a dynamic version of both the WL test and the unfolding trees to show that the graph equivalences regarding the attributed/dynamic WL test and attributed/dynamic unfolding trees are equivalent. With these notions, we define the set of functions that are attributed/dynamic unfolding tree preserving and reformulate the universal approximation theorem to the attributed and dynamic cases (cf. Theorem 5.1.3 and Theorem 5.2.4).

4. Weisfeiler–Lehman and unfolding trees

There are many different extensions of the WL test, e.g., n-dim. WL test, n-dim folklore WL test, or set n-dim. WL test (Sato, 2020). In general, the 1-WL test is based on a graph coloring algorithm. The coloring is applied in parallel on the nodes of the two input graphs, and at the end, the number of colors used per each graph is counted. Then, two graphs are detected as non-isomorphic if these numbers do not coincide, whereas when the numbers are equal, the graphs are possibly isomorphic (WL equivalent). Another way to check the isomorphism of two graphs and therefore compare the GNN expressivity is to consider

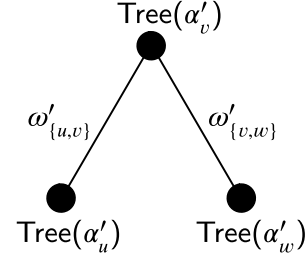


Fig. 3. Unfolding tree recursive construction.

the so-called unfolding trees of all their nodes. An unfolding tree consists of a tree constructed by a breadth-first visit of the graph, starting from a given node. Two graphs are possibly isomorphic if all their unfolding trees are equal. From Krebs and Verbitsky (2015), it is known that for static undirected and unattributed graphs, both the unfolding tree and the Weisfeiler–Lehman approach for testing the isomorphism of two graphs are equivalent.

In this section, we extend the unfolding tree (UT) and the Weisfeiler–Lehman (WL) tests to the domains of SAUHGs and dynamic graphs respectively. Using these we show that also the extended versions of UT-equivalence and WL-equivalence are equivalent.

4.1. Equivalence for attributed static graphs

The extended result on SAUHGs is formalized and proven in Theorem 4.1.6. The original WL test and unfolding tree notions cover all graph properties except edge attributes. Thus, the notion of unfolding trees and the Weisfeiler–Lehman test have to be extended to an attributed version.

Definition 4.1.1 (Attributed Unfolding Tree). The **attributed unfolding tree** \mathbf{T}_v^d in graph $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega')$ of node $v \in \mathcal{V}'$ **up to depth** $d \in \mathbb{N}_0$ is defined as

$$\mathbf{T}_v^d = \begin{cases} \text{Tree}(\alpha'_v), & \text{if } d = 0 \\ \text{Tree}(\alpha'_v, \Omega'_{ne_v}, \mathbf{T}_{ne_v}^{d-1}) & \text{if } d > 0, \end{cases}$$

where $\text{Tree}(\alpha'_v)$ is a tree constituted of node v with attribute α'_v . $\text{Tree}(\alpha'_v, \Omega'_{ne_v}, \mathbf{T}_{ne_v}^{d-1})$ is the tree consisting of the root node v and subtrees $\mathbf{T}_{ne_v}^{d-1} = \{\{\mathbf{T}_{u_1}^{d-1}, \dots, \mathbf{T}_{u_{|ne_v|}}^{d-1}\}\}$ of depth $d-1$, that are connected by the corresponding edge attributes $\Omega'_{ne_v} = \{\{\omega'_{\{v, u_1\}}, \dots, \omega'_{\{v, u_{|ne_v|}\}}\}\}$ of the neighbors of v (see Fig. 3).

Moreover, the **attributed unfolding tree** of v determined by $\mathbf{T}_v = \lim_{d \rightarrow \infty} \mathbf{T}_v^d$ is obtained by merging all unfolding trees \mathbf{T}_v^d of any depth d .

Definition 4.1.2 (Attributed Unfolding Equivalence). Let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, \alpha_1, \omega_1)$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2, \alpha_2, \omega_2)$ be two SAUHGs. Then G_1 and G_2 are **attributed unfolding tree equivalent**, noted by $G_1 \sim_{AUT} G_2$, if and only if $\{\{\mathbf{T}_u \mid u \in \mathcal{V}_1\}\} = \{\{\mathbf{T}_v \mid v \in \mathcal{V}_2\}\}$. Analogously, two nodes

$u \in \mathcal{V}_1, v \in \mathcal{V}_2$ are unfolding tree equivalent, noted by $u \sim_{AUT} v$ if and only if $\mathbf{T}_u = \mathbf{T}_v$.

Using the definition of the attributed unfolding equivalence on graphs, the 1-Weisfeiler Lehman (1-WL) test provided in [D'Inverno et al. \(2021\)](#) is extended to attributed graphs. This allows for the definition of the attributed 1-WL equivalence on graphs and the subsequent [Lemma 4.1.5](#) and the resulting [Theorem 4.1.6](#) that pose the relation between the attributed unfolding equivalence and the attributed 1-WL test.

Definition 4.1.3 (Attributed 1-WL test). Let HASH be a bijective function that codes every possible node attribute with a color from a color set C and $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega')$. The **attributed 1-WL (1-AWL) test** is defined recursively through the following.

- At iteration $i = 0$, the color is set to the hashed node attribute:

$$c_v^{(0)} = \text{HASH}(\alpha'_v)$$
- At iteration $i > 0$, the HASH function is extended to the edge weights:

$$c_v^{(i)} = \text{HASH}\left(\left(c_v^{(i-1)}, \Omega'_{ne_v}, c_{ne_v}^{(i-1)}\right)\right)$$

In the following, the 1-WL equivalence of graphs and nodes is extended by using the attributed version of the 1-WL test (cf. [Definition 4.1.3](#)).

Definition 4.1.4 (Attributed 1-WL Equivalence). Two nodes u, v are attributed WL-equivalent, noted by $u \sim_{AWL} v$, if and only if $c_u = c_v$.

Analogously, let $G_1 = (\mathcal{V}_1, \mathcal{E}_1, \alpha_1, \omega_1)$ and $G_2 = (\mathcal{V}_2, \mathcal{E}_2, \alpha_2, \omega_2)$ be two SAUHGs. Then, $G_1 \sim_{AWL} G_2$, if and only if for all nodes $v_1 \in \mathcal{V}_1$ there exists a corresponding node $v_2 \in \mathcal{V}_2$ such that $v_1 \sim_{AWL} v_2$.

Finally, to complete the derivation of the equivalence between the attributed unfolding equivalence (cf. [Definition 4.1.2](#)) and the attributed WL-equivalence (cf. [Definition 4.1.4](#)), the following helping lemma [Lemma 4.1.5](#) is given, which directly leads to [Theorem 4.1.6](#). The lemma states the equivalence between the attributed unfolding tree equivalence of nodes and the equality of their attributed unfolding trees up to a specific depth. In [Scarselli et al. \(2008\)](#), it has been shown that the unfolding trees of infinite depth are not necessary to consider for this equivalence. Instead, the larger number of nodes of both graphs under consideration is sufficient for the depth of the unfolding trees, which is finite since the graphs are bounded. The following lemma determines the equivalence between the attributed unfolding trees of two nodes and their colors resulting from the attributed 1-WL test.

Lemma 4.1.5. Consider $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega')$ as the SAUHG resulting from a transformation of an arbitrary static graph $G = (\mathcal{V}, \mathcal{E}, \alpha, \omega)$ with nodes $u, v \in \mathcal{V}$ and corresponding attributes α_u, α_v . Then it holds

$$\forall d \in \mathbb{N}_0 : \mathbf{T}_u^d = \mathbf{T}_v^d \iff c_u^{(d)} = c_v^{(d)}.$$

The proof can be found in [Appendix A.2](#)

Directly from [Lemma 4.1.5](#), the equivalence of the attributed unfolding tree equivalence and the attributed 1-WL equivalence of two nodes belonging to the same graph can be formalized.

Theorem 4.1.6. Consider G' as in [Lemma 4.1.5](#). Then, it holds

$$\forall u, v \in \mathcal{V}' : u \sim_{AUT} v \iff u \sim_{AWL} v.$$

Proof. The proof follows from the proof of [Lemma 4.1.5](#). \square

4.2. Equivalence for dynamic graphs

In this section, the previously introduced concepts of unfolding tree and WL equivalences are extended to the dynamic case. Note that

[Lemma 4.1.5](#) and, therefore, [Theorem 4.1.6](#) also hold in case G' is the SAUHG (cf. [Definition 3.1](#)) resulting from a transformation of a dynamic graph $G = (G_t)_{t \in I}$ to its static attributed version. However, the GNNs working on dynamic graphs usually use a significantly different architecture than those that work on static attributed graphs. Therefore, the following includes the derivation of the various equivalences on dynamic graphs separately.

First, dynamic unfolding trees are introduced as a sequence of unfolding trees for each graph snapshot respectively. Afterward, the equivalence of two dynamic graphs regarding their dynamic unfolding trees is presented.

Definition 4.2.1 (Dynamic Unfolding Tree). Let $G = (G_t)_{t \in I}$ with $G_t = (\mathcal{V}_t, \mathcal{E}_t, \alpha_t, \omega_t)$ be a dynamic graph. The **dynamic unfolding tree** $\mathbf{T}_v^d(t)$ at time $t \in I$ of node $v \in \mathcal{V}$ up to depth $d \in \mathbb{N}_0$ is defined as

$$\mathbf{T}_v^d(t) = \begin{cases} \text{Tree}(\alpha_v(t)), & \text{if } d = 0 \\ \text{Tree}(\alpha_v(t), \Omega_{ne_v(t)}, \mathbf{T}_{ne_v(t)}^{d-1}(t)) & \text{if } d > 0, \end{cases}$$

where $\text{Tree}(\alpha_v(t))$ is a tree constituted of node v with attribute $\alpha_v(t)$. Furthermore, $\text{Tree}(\alpha_v(t), \Omega_{ne_v(t)}, \mathbf{T}_{ne_v(t)}^{d-1}(t))$ is the tree with root node v with attribute $\alpha_v(t)$. Additionally, $\mathbf{T}_{ne_v(t)}^{d-1}(t) = \{\{\mathbf{T}_{u_1}^{d-1}(t), \dots, \mathbf{T}_{u_{|ne_v(t)|}}^{d-1}(t)\}\}$ are corresponding subtrees with edge attributes $\Omega_{ne_v(t)}$. If the node v does not exist at time t , the corresponding tree is empty, there is no tree of depth $d > 0$ for this timestep and v does not occur in any neighborhood of other nodes.

In total, the **dynamic unfolding tree of v at time t** , $\mathbf{T}_v(t) = \lim_{d \rightarrow \infty} \mathbf{T}_v^d(t)$ is obtained by merging, i.e., all the unfolding trees $\mathbf{T}_v^d(t)$ for any d .

Definition 4.2.2. Two nodes $u, v \in \mathcal{V}$ are said to be **dynamic unfolding equivalent** $u \sim_{DUT} v$ if $\mathbf{T}_u(t) = \mathbf{T}_v(t)$ for every timestep t . Analogously, two dynamic graphs G_1, G_2 are said to be **dynamic unfolding equivalent** $G_1 \sim_{DUT} G_2$, if there exists a bijection between the nodes of the graphs that respects the partition induced by the unfolding equivalence on the nodes.

Definition 4.2.3 (Dynamic 1-WL Test). Let $G = (G_t)_{t \in I}$ with $G_t = (\mathcal{V}, \mathcal{E}, \alpha_t, \omega_t)$ be a dynamic graph. Let HASH_t^0 be a bijective function encoding every node attribute of G_t with a color from a color set C . The **dynamic 1-WL test (1-DWL)** generates a vector of color sets one for each timestep $t \in I$ by:

- At iteration $i = 0$ the color is set to the hashed node attribute or a fixed color for non-existent nodes:

$$c_v^{(0)}(t) = \begin{cases} \text{HASH}_t^0(\alpha_v(t)) & \text{if } v \in \mathcal{V}_t, \\ c^\perp & \text{otherwise.} \end{cases}$$

- Then, the aggregation mechanism is defined by the bijective function HASH_t for $i > 0$:

$$c_v^{(i)}(t) = \text{HASH}_t\left(\left(c_v^{(i-1)}(t), \Omega_{ne_v(t)}, c_{ne_v(t)}^{(i-1)}(t)\right)\right)$$

Note that for $i > 0$, $c_v^{(i-1)}(t) = c^\perp$ holds for a non-existent node at time t . Further, the neighborset is empty so the other inputs of HASH_t are empty, and together with c^\perp it will always give the same color for non-existent nodes.

Definition 4.2.4 (Dynamic 1-WL Equivalence). Two nodes $u, v \in \mathcal{V}$ in a dynamic graph G are said to be **dynamic WL equivalent**, noted by $u \sim_{DWL} v$, if their colors resulting from the WL test are pairwise equal per timestep. Analogously, let G_1 and G_2 be dynamic graphs. Then $G_1 \sim_{DWL} G_2$, if and only if for all nodes $v_1 \in \mathcal{V}_1^{(1)}$ there exists a corresponding node $v_2 \in \mathcal{V}_2^{(2)}$ with $c_{v_1}(t) = c_{v_2}(t)$ for all $t \in I$.

Theorem 4.2.5 (Equivalence of Dynamic WL Equivalence and Dynamic UT Equivalence for Nodes). Let $G = (G_t)_{t \in I}$ be a dynamic graph and $u, v \in \mathcal{V}$. Then, it holds

$$u \sim_{DUT} v \iff u \sim_{DWL} v.$$

Proof. Two nodes are dynamic unfolding tree equivalent iff they are attributed unfolding tree equivalent at each timestep $u \sim_{AUT} v \quad \forall t \in I$ 4.2.2. Further, as consequence of Theorem 4.1.6, it holds that for all $t \in I$ the two nodes are attributed WL equivalent $u \sim_{AWL} v$ and thus, the two nodes are dynamic WL equivalent by Definition 4.2.4. In case of the non-existence of u at a certain timestep t , the Theorem still holds. \square

5. Approximation capability of GNN's

In this section, the results from Sections 4.1 and 4.2 are brought together in the formulation of a universal approximation theorem for GNNs working on SAUHG and dynamic graphs and the set of functions that preserve the attributed or dynamic unfolding equivalence, respectively.

5.1. GNNs for attributed static graphs

Since the goal is to show the attributed extension of the universal approximation theorem, it is necessary to define the corresponding family of attributed unfolding equivalence-preserving functions. A function preserves the attributed unfolding equivalence if the output of the function is equal when two nodes are attributed unfolding equivalent.

Definition 5.1.1. Let \mathcal{G}' be the domain of bounded SAUHG, $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega') \in \mathcal{G}'$ a SAUHG and $u, v \in \mathcal{V}'$ two nodes. Then a function $f : \mathcal{G}' \rightarrow \mathbb{R}^m$ is said to **preserve the attributed unfolding equivalence** on \mathcal{G}' if

$$v \sim_{AUT} u \Rightarrow f(G', v) = f(G', u).$$

All functions that preserve the attributed unfolding equivalence are collected in the set $\mathcal{F}(\mathcal{G}')$.

Analogous to the argumentation in Scarselli et al. (2008), there exists a relation between the unfolding equivalence preserving functions and the unfolding trees for attributed graphs, as follows.

Proposition 5.1.2 (Functions of Attributed Unfolding Trees). A function f belongs to $\mathcal{F}(\mathcal{G}')$ if and only if there exists a function κ defined on trees such that for any graph $G' \in \mathcal{G}'$ it holds $f(G', v) = \kappa(\mathbf{T}_v)$, for any node $v \in G'$.

The proof works analogously to the proof of the unattributed version presented in Scarselli et al. (2008) and can be found in Appendix A.1.

Considering the previously defined concepts and statements for SAUHG in Section 4.1, finally, the following theorem states the universal approximation capability of the SGNNs on bounded SAUHG.

Theorem 5.1.3 (Universal Approximation Theorem by SGNN). Let \mathcal{G}' be the domain of bounded SAUHG with the maximal number of nodes $N = \max_{G' \in \mathcal{G}'} |G'|$. For any measurable function $f \in \mathcal{F}(\mathcal{G}')$ preserving the attributed unfolding equivalence (cf. Definition 5.1.1), any norm $\|\cdot\|$ on \mathbb{R} , any probability measure P on \mathcal{G}' , for any reals ϵ, λ where $\epsilon, \lambda > 0$, there exists a SGNN defined by the continuously differentiable functions $\text{COMBINE}^{(i)}$, $\text{AGGREGATE}^{(i)}$, at iteration $i \leq 2N - 1$, and by the function READOUT , with hidden dimension $r = 1$, i.e., $h_v^i \in \mathbb{R} \quad \forall i$, such that the function φ (realized by the GNN) computed after $2N - 1$ steps for all $G' \in \mathcal{G}'$ satisfies the condition

$$P(\|f(G', v) - \varphi(G', v)\| \leq \epsilon) \geq 1 - \lambda.$$

The corresponding proof can be found in A.4.

As in Scarselli et al. (2008) we want now to study the case when the employed components (COMBINE, AGGREGATE, READOUT) are sufficiently general to be able to approximate any function preserving the unfolding equivalence. We call this class of networks, \mathcal{Q}_S , SGNN models with universal components. To simplify our discussion, we introduce the transition function $f^{(i)}$ to indicate the concatenation of the $\text{AGGREGATE}^{(i)}$ and $\text{COMBINE}^{(i)}$, i.e.,

$$f^{(i)}(\mathbf{h}_v^i, \{\mathbf{h}_u^{i-1}, u \in ne[v]\} \{\omega_{\{u,v\}}\}_{u \in ne_v}) = \text{COMBINE}^{(i)}\left(\mathbf{h}_v^{(i-1)}, \text{AGGREGATE}^{(i)}\left(\{\mathbf{h}_u^{(i-1)}\}_{u \in ne_v}, \{\omega_{\{u,v\}}\}_{u \in ne_v}\right)\right)$$

Then, we can formally define the class \mathcal{Q}_S .

Definition 5.1.4. A class \mathcal{Q}_S of SGNN models is said to have *universal components* if, for any $\epsilon > 0$ and any continuous target functions $\overline{\text{COMBINE}}^{(i)}$, $\overline{\text{AGGREGATE}}^{(i)}$, $\overline{\text{READOUT}}$, there exists an SGNN belonging to \mathcal{Q}_S , with functions $\text{COMBINE}_\theta^{(i)}$, $\text{AGGREGATE}_\theta^{(i)}$, READOUT_θ and parameters θ such that

$$\|\overline{f}^{(i)}(\mathbf{h}, \{\mathbf{h}_1, \dots, \mathbf{h}_n\}) - f_\theta^{(i)}(\mathbf{h}, \{\mathbf{h}_1, \dots, \mathbf{h}_n\})\|_\infty \leq \epsilon$$

$$\|\overline{\text{READOUT}}(\mathbf{q}) - \text{READOUT}_\theta(\mathbf{q})\|_\infty \leq \epsilon,$$

holds, for any vectors $\mathbf{h}, \mathbf{h}_1, \dots, \mathbf{h}_n \in \mathbb{R}^r$, $\mathbf{q} \in \mathbb{R}^s$. The transition functions $\overline{f}^{(i)}$ and $f_\theta^{(i)}$ correspond to the target function and the SGNN, respectively.

The following result shows that Theorem 5.1.3 still holds even for SGNNs with universal components.

Theorem 5.1.5 (Approximation by Neural Networks). Assuming that the hypotheses of Theorem 5.1.3 are fulfilled and \mathcal{Q}_S is a class of SGNNs with universal components. Then, there exists a parameter set θ and some functions $\text{COMBINE}_\theta^{(i)}$, $\text{AGGREGATE}_\theta^{(i)}$, READOUT_θ , implemented by Neural Networks in \mathcal{Q}_S , such that the thesis of Theorem 5.1.3 holds. \square

Proof. The proof is identical to the one contained in Scarselli et al. (2008); to give a hint on the methodology, we refer to the more complex proof of the analogous Theorem 5.2.6 for DGNN. \square

5.2. GNNs for dynamic graphs

Suitable functions that preserve the unfolding equivalence on dynamic graphs are dynamic systems. Before this statement is formalized and proven in Proposition 5.2.3, dynamic systems and their property to preserve the dynamic unfolding equivalence are defined in the following.

Definition 5.2.1 (Dynamic System). Let \mathcal{G} be a domain of dynamic graphs and let $\mathcal{V} = \bigcup_t \mathcal{V}_t$. A **dynamic system** is defined as a function $\text{dyn} : D := I \times \mathcal{G} \times \mathcal{V} \rightarrow \mathbb{R}^m$ formalized for $G = (G_t)_{t \in I} \in \mathcal{G}$, and $v \in \mathcal{V}_t$ by

$$\text{dyn}(t, G, v) := g(x_v(t)). \quad (3)$$

Here, $g : \mathbb{R}^r \rightarrow \mathbb{R}^m$ is an output function, and the *state function* $x_v(t)$ is determined by

$$x_v(t) = \begin{cases} a(t, G, v) & \text{if } t = 0 \\ f(x_v(t-1), a(t-1, G, v)) & \text{if } t > 0, \end{cases}$$

for $v \in \mathcal{V}_t$, where $a : I \times \mathcal{G} \times \mathcal{V} \rightarrow \mathbb{R}^r$ is a function that processes the graph snapshot at time t and provides an r -dimensional internal state representation for each node v . Finally, $f : \mathbb{R}^r \times \mathbb{R}^r \rightarrow \mathbb{R}^r$ is a recursive function, that is called *state update function*.

Definition 5.2.2. A dynamic system $\text{dyn}(\cdot, \cdot, \cdot)$ **preserves the dynamic unfolding tree equivalence** on \mathcal{G} if and only if for any input graph sequences $G_1, G_2 \in \mathcal{G}$, and two nodes $u, v \in \mathcal{V}$ it holds

$$v \sim_{DUT} u \implies \text{dyn}(t, G_1, v) = \text{dyn}(t, G_2, u) \quad \forall t.$$

The class of dynamic systems that preserve the unfolding equivalence on \mathcal{D} will be denoted with $\mathcal{F}(\mathcal{D})$. A characterization of $\mathcal{F}(\mathcal{D})$ is given by the following result (following the work in Scarselli et al. (2008)).

Proposition 5.2.3 (Functions of Dynamic Unfolding Trees). A dynamic system dyn belongs to $\mathcal{F}(\mathcal{D})$ if and only if there exists a function κ defined on attributed trees such that for all $(t, G, v) \in \mathcal{D}$ it holds

$$\text{dyn}(t, G, v) = \kappa\left(\left(\mathbf{T}_v(i)\right)_{i \in [t]}\right).$$

The proof can be found in Appendix A.3.

Finally, the universal approximation of the Message-Passing GNN for dynamic graphs is determined as follows.

Theorem 5.2.4 (Universal Approximation Theorem by DGNN). Let $G = (G_t)_{t \in I}$ be a discrete dynamic graph in the graph domain \mathcal{G} and $N = \max_{G \in \mathcal{G}} |G|$ be the maximal number of nodes in the domain. Let $\text{dyn}(t, G, v) \in \mathcal{F}(\mathcal{D})$ be any measurable dynamical system preserving the unfolding equivalence, $\|\cdot\|$ be a norm on \mathbb{R} , P be any probability measure on \mathcal{D} and ϵ, λ be any real numbers where $\epsilon, \lambda > 0$. Then, there exists a DGNN composed by SGNNs with $2N - 1$ layers and hidden dimension $r = 1$, and Recurrent Neural Network state dimension $s = 1$ such that the function φ realized by this model satisfies

$$P(\|\text{dyn}(t, G, v) - \varphi(t, G, v)\| \leq \epsilon) \geq 1 - \lambda \quad \forall t \in I.$$

The proof can be found in Appendix A.5.

Theorem 5.2.4 intuitively states that, given a dynamical system dyn , there is a DGNN that approximates it. The functions which the DGNN is a composition of (such as the dynamical function f , COMBINE⁽ⁱ⁾, AGGREGATE⁽ⁱ⁾, etc.) are supposed to be continuously differentiable but still generic, while can be generic and completely unconstrained. This situation does not correspond to practical cases where the DGNN adopts particular architectures, and those functions are Neural Networks, or more generally, parametric models — for example, made of layers of sum, max, average, etc. Thus, it is of fundamental interest to clarify whether the theorem still holds when the components of the DGNN are parametric models.

Definition 5.2.5. A class \mathcal{Q}_D of discrete DGNN models is said to have *universal components* if the employed SGNNs have universal components as defined in Definition 5.1.4 and the employed recurrent model is designed such that for any $\epsilon_1, \epsilon_2 > 0$ and any continuously differentiable target functions $\bar{f}, \overline{\text{READOUT}}_{\text{dyn}}$ there is a discrete DGNN in the class \mathcal{Q}_D , with functions $f_\theta, \text{READOUT}_{\text{dyn}, \theta}$ and parameters θ such that, for any input vectors $\mathbf{h} \in \mathbb{R}^r, \mathbf{q}, \mathbf{q}^* \in \mathbb{R}^s$, it holds

$$\begin{aligned} \left\| \bar{f}(\mathbf{q}, \mathbf{h}) - f_\theta(\mathbf{q}, \mathbf{h}) \right\|_\infty &\leq \epsilon_1, \\ \left\| \overline{\text{READOUT}}_{\text{dyn}}(\mathbf{q}^*) - \text{READOUT}_{\text{dyn}, \theta}(\mathbf{q}^*) \right\|_\infty &\leq \epsilon_2. \end{aligned}$$

The following result shows that Theorem 5.2.4 still holds even for discrete DGNNs with universal components.

Theorem 5.2.6 (Approximation by Neural Networks). Assume that the hypotheses of Theorem 5.2.4 are fulfilled and \mathcal{Q}_D is a class of discrete DGNNs with universal components. Then, there exists a parameter set θ , and the functions $\bar{f}, \overline{\text{READOUT}}_{\text{dyn}}$, implemented by Neural Networks in \mathcal{Q}_D , such that Theorem 5.2.4 holds.

The proof can be found in Appendix A.6

5.3. Discussion

The following remarks may further help to understand the results proven in the previous paragraphs:

- Theorem 5.1.3 suggests an alternative approach to process several graph domains with a universal SGNN model. Actually, almost all the graphs, including, e.g., hypergraphs, multigraphs, directed graphs, etc., can be transformed to SAUGHs with node and edge attributes (Thomas et al., 2021). Then, we can use a universal GNN model on such a domain using sufficiently expressive AGGREGATE and COMBINE functions.
- The proofs of Theorems 5.1.3 and 5.2.4 are based on space partitioning reasoning. Differently from technique based on Stone–Weierstrass theorem (Azizian & Lelarge, 2021), which are existential in nature, such an approach allows us to deduce information about the characteristics of networks that reach the desired approximation. Actually, the theorems point out that the approximation can be obtained with a minimal hidden dimension $r = 1$ both in SGNNs and DGNNs and with a state dimension $s = 1$ in the Recurrent Neural Network of DGNNs. Such a result may appear surprising, but the proofs show that GNNs can encode unfolding trees with a single real number.
- Moreover, Theorems 5.1.3 and 5.2.4 specify that GNNs can obtain the approximation with $2N - 1$ layers. We could incorrectly presume that the maximum number of layers required to reach a desired approximation depends on the diameter $\text{diam}(G)$ of the graph, which can be smaller than the number of nodes N since a GNN can move the information from one node to another in $\text{diam}(G)$ iterations. However, $\text{diam}(G)$ layers are not always sufficient to distinguish all the nodes of a graph. In fact, it has been proven that $N - 1$ is a lower bound on the number of iterations that the 1-WL algorithm has to carry out to be able to distinguish any pairs of 1-WL distinguishable graphs (Kiefer & McKay, 2020), and $2N - 1$ is a lower bound for 1-WL algorithm to distinguish pairs of nodes in two different graphs (Krebs & Verbitsky, 2015). So overall, $2N - 1$ is also the lower bound for the GNN computation time to approximate any function for either graph- or node-focused tasks (see D’Inverno et al. (2021) for a detailed discussion).
- Theorems 5.1.3 and 5.2.4 specify that the approximation is modulo unfolding equivalence, or, correspondingly, modulo WL equivalence. It can be observed that in the dynamic case, only a part of the architecture affects the equivalence. Actually, a dynamic GNN contains two modules: the first one, an SGNN, produces an embedding of the input graph at each time instance; the second component contains a Recurrent Neural Network that processes the sequence of the embeddings. The dynamic unfolding equivalence is defined by sequences of unfolding trees, which are built independently for each node and time instance by the SGNN. Similarly, the dynamic WL equivalence is defined by sequences of colors defined independently at each time step. Intuitively, the Recurrent Neural Network does not affect the equivalence, since Recurrent Neural Networks can be universal approximators and implement any function of the sequence without introducing other constraints beyond those already introduced by the SGNN.
- Theorem 5.2.4 does not hold for any Dynamic GNN, as we take into account a discrete recurrent model working on graph snapshots (also known as Stacked DGNN). Nevertheless, several DGNNs of this kind are listed in Kazemi et al. (2020), such as GCRN-M1 (Seo, Defferrard, Vandergheynst, & Bresson, 2018), RgCNN (Narayan & Roe, 2018), PATCHY-SAN (Niepert, Ahmed, & Kutzkov, 2016), DyGGNN (Taheri, Gimpel, & Berger-Wolf, 2019), and others. Still, the approximation capability depends on the functions AGGREGATE and COMBINE designed for each GNN working on the single snapshot and the implemented Recurrent Neural Network. For example, the most general model, the original RNN, has been proven to be a universal approximator (Hammer, 2000).

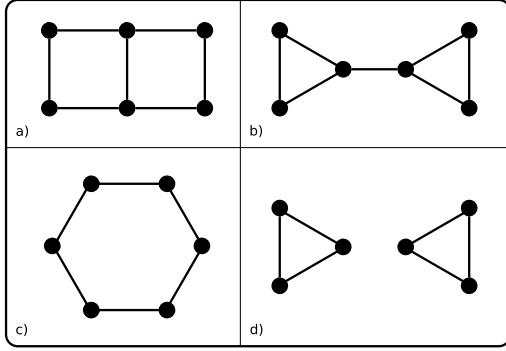


Fig. 4. The four static graphs are used as components to generate the synthetic dataset. Graphs (a) and (b) are equivalent under the static 1-WL test; the same holds for (c) and (d).

6. Experimental validation

In this Section, we support our theoretical findings with an experimental study. For this purpose, we carry out two sets of experiments described as follows:

- E1. We show that a DGNN with universal components can approximate a function $F_{DWL} : \mathcal{G} \rightarrow \mathbb{N}$ that models the 1-DWL test. The function F_{DWL} assigns a target label to the input graph that represents the class of 1-DWL equivalence;
- E2. In the same approximation task, we compare DGNNs with different GNN modules from the literature to show how the universality of the components affects the approximation capability.

We focus on the ability of the DGNN to approximate F_{DWL} , so only training performances are considered, i.e., we do not investigate the generalization capabilities over a test set. Since the 1-DWL test provides the finest partition of graphs reachable by a DGNN, the mentioned tasks experimentally evaluate the expressive power of DGNNs.

Dataset. The dataset consists of dynamic graphs, i.e., vectors of static graph snapshots of fixed length T . Each static snapshot is one of the graphs in Fig. 4. Since the dataset is composed of all the possible combinations of the four graphs, it contains 4^T dynamic graphs. Given that the graphs in Fig. 4 are pairwise 1-WL equivalent ((a) is 1-WL equivalent to (b) and (c) is 1-WL equivalent to (d)), the number of classes is 2^T , with $\frac{4^T}{2^T} = 2^T$ graphs in each class. For each dynamic graph, the target is the corresponding 1-DWL output, represented as a natural number. For training purposes, the targets are normalized between 0 and 1 and uniformly spaced in the interval $[0, 1]$. Therefore, the distance between each class label is $d = 1/2^T$. A dynamic graph G with target y_G will be said to be correctly classified if, given out = DGNN(G), we have $|\text{out} - y_G| < d/2$.

Experimental setup.

- E1 For the first set of experiments, the Dynamic Graph Neural Network is composed of two modules: A Graph Isomorphism Network (GIN) (Xu et al., 2019) and a Recurrent Neural Network (RNN), which implement the static GNN and the temporal Network f of Eq. (1), respectively. Since it has been proven that the GIN is a universal architecture (Xu et al., 2019) and the RNNs are universal approximators for dynamical systems on vector sequences (Hammer, 2000), the architecture used in the experiments fits the hypothesis of Theorem 5.2.4. Thus, it can approximate any dynamical system on the temporal graph domain.

The model hyperparameters for the experiments are set as follows. The GIN includes $n_{\max} = 6$ layers.⁴ The MLP in the GIN network contains one hidden layer with a hyperbolic tangent activation function and batch normalization. Hidden layers of different sizes, i.e., $h_{\text{gin}} \in \{1, 4, 8\}$, have been tested. For sake of simplicity, the output network has one hidden layer with the same number of neurons as the MLP in the GIN. Furthermore, $h_{\text{rnn}} = 8$ is the size of the hidden state of the RNN.

- E2 In the other set of experiments, we test DGNNs composed by different GNN static modules and an RNN module (analogously to E1.). In particular, we compare DGNNs with the GNN module taken from the following list:

- GIN as mentioned before;
- Graph Convolutional Network (GCN) (Kipf & Welling, 2017);
- GNN presented in Hamilton, Ying, and Leskovec (2017b) (see also Morris et al. (2019)) where the aggregation function is the *sum* of the hidden features of the neighbors; it will be called `gconv_add`;
- GNN presented in Hamilton et al. (2017b) with *mean* of the hidden features of the neighbors as aggregation function, called `gconv_mean` here;
- GAT (Veličković et al., 2018).

Here, the used hyperparameters are hidden dimension is $h = 8$, the number of layers $L = 4$, and the time length $L = T = 5$.

In both the experimental cases, the model is trained over 300 epochs using the Adam optimizer with a learning rate of $\lambda = 10^{-3}$. Each configuration is evaluated over 10 runs. The overall training is then performed on an Intel(R) Core(TM) i7-9800X processor running at 3.80 GHz using 31 GB of RAM and a GeForce GTX 1080 Ti GPU unit. The code used to run the experiments can be found at <https://github.com/AleDinve/dyn-gnn>.

Results. The results of the experiments confirm our theoretical statements. More precisely, the DGNNs performed as follows during training.

- E1 In Fig. 5, the evolution of the training accuracy over the epochs is presented for different GIN hidden layer sizes h_{gin} and for dynamic graphs up to time lengths $T = 4$ (Fig. 5(a)) and $T = 5$ (Fig. 5(b)). All the architectures reach 100% accuracy for experiments on both time lengths. Even setting $h_{\text{gin}} = 1$ leads to a perfect classification at a slower rate. It may appear surprising that, even with a hidden representation of size 1, the DGNN can well approximate the function F_{DWL} . However, as we already pointed out in Section 5.3, the possibility of reaching the universal approximation with a feature of dimension 1 is confirmed by Theorem 5.2.4.
- E2 The DGNN with the GIN module achieve the best performance in terms of learning accuracy and speed of decreasing, as illustrated in Fig. 6. The DGNN with the `gconv_add` module is able to learn the task, although learning is unstable (see Fig. 6 (b)). This is not surprising since this module has been proven to match the expressive power of the 1-WL test (Morris et al., 2019). The other DGNNs are incapable to learn the objective function. This is a consequence of their weaker expressive power, widely investigated in literature (D’Inverno et al., 2021; Xu et al., 2019).

Thus, overall, our theoretical expectations were met by both experiments.

⁴ As investigated in Section 5.3, for graph-focused tasks, it is sufficient to perform the message-passing convolution for several times equal to the maximum number of nodes over the graphs in the dataset domain.

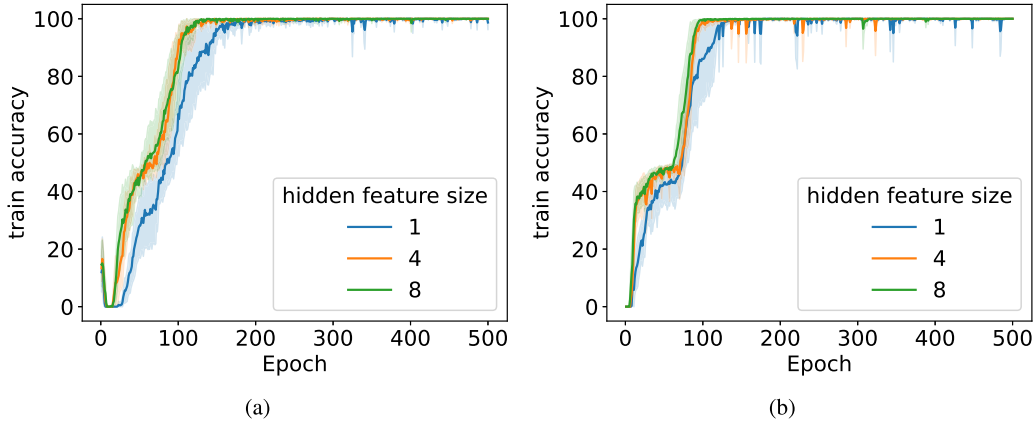


Fig. 5. Experimental Framework E1. Training accuracy over the epochs for a DGNN trained on the dataset containing dynamic graphs up to time length $T = 4$ (a) and $T = 5$ (b).

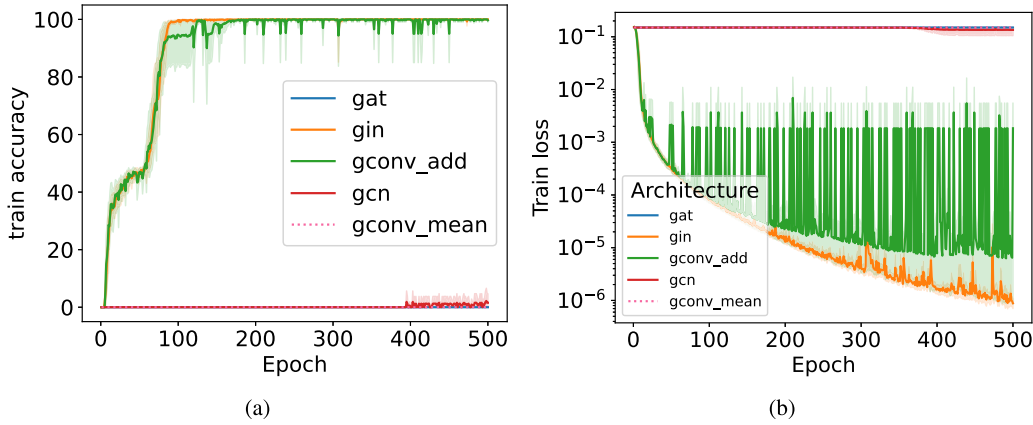


Fig. 6. Experimental framework E2. Training accuracy (a) and training loss (b) over the epochs for several DGNNs trained on the dataset containing dynamic graphs up to time length $T = 5$. Figure (b) is in logarithmic scale.

7. Conclusion and future work

This paper provides two extensions of the 1-WL isomorphism test and the definition of unfolding trees of nodes and graphs. First, we introduced WL test notions to attributed and dynamic graphs, and second, we introduced extended concepts for unfolding trees on attributed and dynamic graphs. Further, we proved the existence of a bijection between the dynamic 1-WL equivalence of dynamic graphs and the attributed 1-WL equivalence of their corresponding static versions that are bijective regarding their encoded information. The same result we proved w.r.t. the unfolding tree equivalence. Moreover, we extended the strong connection between unfolding trees and the (dynamic/attributed) 1-WL tests — proving that they give rise to the same equivalence between nodes for the attributed and the dynamic case, respectively. Note that the GNNs working on static graphs usually have another architecture than those working on dynamic graphs. Therefore, we have proved that both the different GNN types can approximate any function that preserves the (attributed/dynamic) unfolding equivalence (i.e., the (attributed/dynamic) 1-WL equivalence).

Note that the dynamic GNN considered here in this paper is given in discrete-time representation, i.e., as a sequence of static graph snapshots without actual timestamps. Thus, [Theorem 5.2.4](#) does not hold for any Dynamic GNN, as we take into account a discrete recurrent model working on graph snapshots (also known as Stacked DGNN). Nevertheless, several DGNNs of this kind are listed in [Kazemi et al. \(2020\)](#), such as GCRN-M1 ([Seo et al., 2018](#)), RgCNN ([Narayan & Roe, 2018](#)), PATCHY-SAN ([Niepert et al., 2016](#)), DyGGNN ([Taheri et al., 2019](#)), and others. Still, the approximation capability depends on the functions AGGREGATE and COMBINE designed for each GNN working

on the single snapshot and the implemented Recurrent Neural Network. For example, the most general model, the original RNN, has been proven to be a universal approximator ([Hammer, 2000](#)).

As future work, extending all our results for graphs in continuous-time representation would be interesting. One difficulty in this context is deciding in which sense two continuous-time dynamic graphs are called WL equivalent since there are many possibilities for dealing with the given timestamps. The investigation of equivalences of dynamic graphs requires determining the handling of dynamic graphs that are equal in their structure but differ in their temporal occurrence, i.e., dependent on the commitment of the WL equivalence or the unfolding tree equivalence, it is required to decide whether the concepts need to be **time-invariant**. For time-invariant equivalence, the following concepts hold as they are. In case two graphs with the same structure should be distinguished when they appear at different times, the node and edge attributes can be extended by an additional dimension carrying the exact timestamp. Thereby, the unfolding trees of two (structural) equal nodes would be different, having different timestamps in their attributes. Then, all dynamic graphs $G^{(j)} \in \mathcal{G}$ are defined over the same time interval I . Without loss of generality, this assumption can be made since the set of timestamps of $G^{(j)}$ noted by $I_{G^{(j)}}$ can be padded by including missing timestamps t_q and $G^{(j)}$ can be padded by empty graphs $G_q^{(j)} = \mathcal{V}_q^{(j)} = \emptyset$, $\mathcal{E}_q^{(j)} = \emptyset$, $\alpha_q(\emptyset) = \emptyset$, $\omega_q(\emptyset) = \emptyset$.

Furthermore, this paper considers extensions of the usual 1-WL test and the commonly known unfolding trees. Further future work could be to investigate extensions, for example, the n-dim attributed/dynamic WL test or other versions of unfolding trees, covering GNN models not considered by the frameworks used in this paper. These extensions

might result in a more exemplary classification of the expressive power of different GNN architectures.

Moreover, the proposed results mainly focus on the expressive power of GNNs. However, GNNs with the same expressive power may differ for other fundamental properties, e.g., the computational and memory complexity and the generalization capability. Understanding how the architecture of AGGREGATE⁽ⁱ⁾, COMBINE⁽ⁱ⁾, and READOUT impact those properties is of fundamental importance for practical applications of GNNs.

CRedit authorship contribution statement

Silvia Beddar-Wiesing: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Giuseppe Alessio D’Inverno:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. **Caterina Graziani:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Veronica Lachi:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Alice Moallem-Oureh:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Franco Scarselli:** Conceptualization, Formal analysis, Supervision, Writing – review & editing. **Josephine Maria Thomas:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The code is available on github: <https://github.com/AleDinve/dyn-gnn>.

Acknowledgments

We want to acknowledge Monica Bianchini and Maria Lucia Sampoli for their proofreading contribution and fruitful discussions.

During the preparation of this work the authors used grammarly in order to improve the grammar of the work. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Appendix A

A.1. Proof of Proposition 5.1.2

A function f belongs to $\mathcal{F}(G')$ if and only if there exists a function κ defined on trees such that for any graph $G' \in \mathcal{G}'$ it holds $f(G', v) = \kappa(\mathbf{T}_v)$, for any node $v \in G'$.

Proof. We prove by showing both equivalence directions:

- ⇐ If there exists a function κ on attributed unfolding trees such that $f(G', v) = \kappa(\mathbf{T}_v)$ for all $v \in G'$, then $u \sim_{AUT} v$ for $u, v \in G$ implies $f(G', u) = \kappa(\mathbf{T}_u) = \kappa(\mathbf{T}_v) = f(G', v)$.
- ⇒ If f preserves the attributed unfolding equivalence, then a function κ on the attributed unfolding tree of an arbitrary node v can be defined as $\kappa(\mathbf{T}_v) := f(G', v)$. Then, if \mathbf{T}_u and \mathbf{T}_v are two attributed unfolding trees, $\mathbf{T}_u = \mathbf{T}_v$ implies $f(G', u) = f(G', v)$ and κ is uniquely defined. \square

A.2. Proof of Lemma 4.1.5

Consider $G' = (\mathcal{V}', \mathcal{E}', \alpha', \omega')$ as the SAUHG resulting from a transformation of an arbitrary static graph $G = (\mathcal{V}, \mathcal{E}, \alpha, \omega)$ with nodes $u, v \in \mathcal{V}$ and corresponding attributes α_u, α_v . Then it holds

$$\forall d \in \mathbb{N}_0 : \mathbf{T}_u^d = \mathbf{T}_v^d \iff c_u^{(d)} = c_v^{(d)} \quad (4)$$

Proof. The proof is carried out by induction on d , which represents both the depth of the unfolding trees and the iteration step in the WL coloring.

$d = 0$: It holds

$$\mathbf{T}_u^0 = Tree(\alpha'_u) = Tree(\alpha'_v) = \mathbf{T}_v^0$$

$$\iff \alpha'_u = \alpha'_v \text{ and } c_u^{(0)} = HASH(\alpha'_u) = HASH(\alpha'_v) = c_v^{(0)}.$$

$d > 0$: Suppose that Eq. (4) holds for $d - 1$, and prove that it holds also for d .

- By definition, $\mathbf{T}_u^d = \mathbf{T}_v^d$ is equivalent to

$$\begin{cases} \mathbf{T}_u^{d-1} = \mathbf{T}_v^{d-1} & \text{and} \\ Tree(\alpha'_u, \Omega'_{ne_u}, \mathbf{T}_{ne_u}^{d-1}) = Tree(\alpha'_v, \Omega'_{ne_v}, \mathbf{T}_{ne_v}^{d-1}). \end{cases} \quad (5)$$

- Applying the induction hypothesis, it holds that

$$\mathbf{T}_u^{d-1} = \mathbf{T}_v^{d-1} \iff c_u^{(d-1)} = c_v^{(d-1)}. \quad (6)$$

- Eq. (5) is equivalent to the following:

$$\alpha'_u = \alpha'_v, \quad \Omega'_{ne_u} = \Omega'_{ne_v} \quad \text{and} \quad \mathbf{T}_{ne_u}^{d-1} = \mathbf{T}_{ne_v}^{d-1}.$$

Given the definition of the unfolding trees and their construction, this is equivalent to

$$\begin{cases} \omega'_{\{u, u_i\}} = \omega'_{\{v, v_i\}} & \forall u_i \in ne_u, v_i \in ne_v \text{ and} \\ \mathbf{T}_{u_i}^{d-1} = \mathbf{T}_{v_i}^{d-1} & \forall u_i \in ne_u, v_i \in ne_v. \end{cases} \quad (7)$$

- By the induction hypothesis, Eq. is equivalent to

$$c_{ne_u}^{d-1} = c_{ne_v}^{d-1}, \text{ i.e.,}$$

$$\{\{c_{u_i}^{(d-1)} \mid u_i \in ne_u\}\} = \{\{c_{v_i}^{(d-1)} \mid v_i \in ne_v\}\}.$$

- Putting together Eq. (6), (7), and the fact that the HASH function is bijective, we obtain:

$$\begin{aligned} & HASH\left(\left(c_u^{(d-1)}, \Omega'_{ne_u}, \{\{c_{u_i}^{(d-1)} \mid u_i \in ne_u\}\}\right)\right) \\ &= HASH\left(\left(c_v^{(d-1)}, \Omega'_{ne_v}, \{\{c_{v_i}^{(d-1)} \mid v_i \in ne_v\}\}\right)\right) \end{aligned}$$

which, by definition, is equivalent to $c_u^{(d)} = c_v^{(d)}$. \square

A.3. Proof of Proposition 5.2.3

A dynamic system dyn belongs to $\mathcal{F}(D)$ if and only if there exists a function κ defined on attributed trees such that for all $(t, G, v) \in D$ it holds

$$\text{dyn}(t, G, v) = \kappa\left(\left(\mathbf{T}_v(i)\right)_{i \in [t]}\right).$$

Proof. We show the proposition by proving both directions of the equivalence relation:

- ⇒: If there exists κ such that $\text{dyn}(t, G, v) = \kappa\left(\left(\mathbf{T}_v(i)\right)_{i \in [t]}\right)$ for all triplets $(t, G, v) \in D$, then for any pair of nodes $u \in G_1, v \in G_2$ with $u \sim_{DUT} v$ it holds

$$\text{dyn}(t, G_1, u) = \kappa\left(\left(\mathbf{T}_u(i)\right)_{i \in [t]}\right) = \kappa\left(\left(\mathbf{T}_v(i)\right)_{i \in [t]}\right) = \text{dyn}(t, G_2, v).$$

⇐: On the other hand, if dyn preserves the unfolding equivalence, then we can define κ as

$$\kappa\left(\left(\mathbf{T}_v(i)\right)_{i \in [l]}\right) = \text{dyn}(t, G, v).$$

Note that the above equality is a correct specification for a function. In fact, if

$$\kappa\left(\left(\mathbf{T}_v(i)\right)_{i \in [l]}\right) = \kappa\left(\left(\mathbf{T}_u(i)\right)_{i \in [l]}\right)$$

implies $\text{dyn}(t, G, u) = \text{dyn}(t, G, v)$, then κ is uniquely defined. \square

A.4. Sketch of the proof of Theorem 5.1.3

Let \mathcal{G}' be the domain of bounded SAUHGs with the maximal number of nodes $N = \max_{G' \in \mathcal{G}'} |G'|$. For any measurable function $f \in \mathcal{F}(\mathcal{G}')$ preserving the attributed unfolding equivalence (cf. Definition 5.1.1), any norm $\|\cdot\|$ on \mathbb{R} , any probability measure P on \mathcal{G}' , for any reals ϵ, λ where $\epsilon, \lambda > 0$, there exists a SGNN defined by the continuously differentiable functions $\text{COMBINE}^{(i)}$, $\text{AGGREGATE}^{(i)}$, at iteration $i \leq 2N-1$, and by the function READOUT , with hidden dimension $r = 1$, i.e., $h_v^i \in \mathbb{R} \forall i$, such that the function φ (realized by the SGNN) computed after $2N-1$ steps for all $G' \in \mathcal{G}'$ satisfies the condition

$$P(\|f(G', v) - \varphi(G', v)\| \leq \epsilon) \geq 1 - \lambda.$$

Proof. Since the proof proceeds analogously to the one in D'Inverno et al. (2021), we will only sketch the proof idea here and refer to the original paper for further details. First, we need a preliminary lemma which is an extension of Scarselli et al. (2008, Lem. 1) to the domain of SAUHGs \mathcal{G}' . Intuitively, this lemma suggests that a domain of SAUGH graphs with continuous attributes can be partitioned into small subsets so that the attributes of the graphs are almost constant in each partition. Moreover, in probability, a finite number of partitions is sufficient to cover a large part of the domain.

Lemma A.4.1.

For any probability measure P on \mathcal{D}' , and any reals λ, δ , where $\lambda > 0$, $\delta \geq 0$, there exists a real $\bar{b} > 0$, which is independent of δ , a set $\bar{\mathcal{D}}' \subseteq \mathcal{D}'$, and a finite number of partitions $\bar{\mathcal{D}}'_1, \dots, \bar{\mathcal{D}}'_p$ of $\bar{\mathcal{D}}'$, where $\bar{\mathcal{D}}'_j = \mathcal{G}'_j \times \{v_j\}$, with $\mathcal{G}'_j \subseteq \mathcal{G}'$ and $v_j \in \mathcal{G}'_j$, such that:

1. $P(\bar{\mathcal{D}}') \geq 1 - \lambda$ holds;
2. for each j , all the graphs in \mathcal{G}'_j have the same structure, i.e., they differ only in the values of their attributes;
3. for each set $\bar{\mathcal{D}}'_j$, there exists a hypercube $\mathcal{H}_j \subset \mathbb{R}^{NM2k}$ such that $\gamma_{G'} \in \mathcal{H}_j$ holds for any graph $G' \in \mathcal{G}'_j$ with $N = \max_{G' \in \mathcal{G}'} |\mathcal{V}'|$ and $M = \max_{G' \in \mathcal{G}'} |\mathcal{E}'|$. Here, $\gamma_{G'}$ denotes the vector obtained by concatenating all the attribute vectors of both nodes and edges of G' , namely $\gamma_{G'} = [A_{G'} | \Omega_{G'}]$, where $A_{G'}$ is the concatenation of all the node attributes and $\Omega_{G'}$ is the concatenation of all edge attributes;
4. for any two different sets $\mathcal{G}'_i, \mathcal{G}'_j$, $i \neq j$, their graphs have different structures, or their hypercubes $\mathcal{H}_i, \mathcal{H}_j$ are disjoint, i.e., $\mathcal{H}_i \cap \mathcal{H}_j = \emptyset$;
5. for each j and each pair of graphs $G_1, G_2 \in \mathcal{G}'_j$, the inequality $\|\gamma_{G_1} - \gamma_{G_2}\|_\infty \leq \delta$ holds;
6. for each graph $G' \in \bar{\mathcal{D}}'$, the inequality $\|\gamma_{G'}\|_\infty \leq \bar{b}$ holds.

Proof. The proof is similar to the one contained in Scarselli et al. (2008). The only remark needed here is that we can consider the whole concatenating of all attributes from both nodes and edges without loss of generality; indeed, if we were considering the node and the edge attributes separately, we would need conditions on the hypercubes, s.t.:

$$\|A_{G_1} - A_{G_2}\|_\infty \leq \delta^A, \delta^A > 0,$$

$$\text{and } \|\Omega_{G_1} - \Omega_{G_2}\|_\infty \leq \delta^\Omega, \delta^\Omega > 0.$$

Then we can stack those attribute vectors, as in the statement, s.t.:

$$\begin{aligned} \|\gamma_{G_1} - \gamma_{G_2}\|_\infty &= \|[A_{G_1} | \Omega_{G_1}] - [A_{G_2} | \Omega_{G_2}]\|_\infty \\ &\leq \|A_{G_1} - A_{G_2}\|_\infty + \|\Omega_{G_1} - \Omega_{G_2}\|_\infty \\ &\leq \delta^A + \delta^\Omega := \delta \end{aligned}$$

which allows us to exploit the same proof contained in Scarselli et al. (2008). \square

The following theorem, where the domain contains a finite number of graphs and the attributes are integers, is equivalent to Theorem 5.1.3.

Theorem A.4.2. For any finite set of p patterns $\{(G'_j, v) \mid G'_j \in \mathcal{G}', v \in \mathcal{V}'_j, j \in [p]\}$, with the maximal number of nodes in the domain $N = \max_{G' \in \mathcal{G}'} |G'|$, for any function which preserves the attributed unfolding equivalence, and for any real $\epsilon > 0$, there exist continuously differentiable functions $\text{AGGREGATE}^{(i)}$, $\text{COMBINE}^{(i)}$, $\forall i \leq 2N-1$, s.t.

$$\mathbf{h}_v^i = \text{COMBINE}^{(i)}\left(\mathbf{h}_v^{(i-1)}, \text{AGGREGATE}^{(i)}\left(\{\mathbf{h}_u^{(i-1)}\}_{u \in ne_v}, \{\omega_{\{u,v\}}\}_{u \in ne_v}\right)\right)$$

and a function READOUT , with hidden dimension $r = 1$, i.e., $\mathbf{h}_v^i \in \mathbb{R}$, so that the function φ (realized by the SGNN), computed after N steps, satisfies the condition

$$|\tau(G'_j, v) - \varphi(G'_j, v)| \leq \epsilon \quad \text{for any } v \in \mathcal{V}'_j. \quad (8)$$

Sketch of the Proof. The idea of the proof is designing a GNN that can approximate any function τ that preserves the attributed unfolding equivalence. According to Proposition 5.1.2 there exists a function κ , s.t.

$$\tau(G'_j, v) = \kappa(T_v).$$

Therefore, the GNN has to encode the attributed unfolding tree into the node attributes, i.e., for each node v , we want to have $\mathbf{h}_v = \nabla(\mathbf{T}_v)$, where ∇ is an encoding function that maps attributed unfolding trees into real numbers. The existence and injectiveness of ∇ are ensured by construction. More precisely, the encodings are constructed recursively by the $\text{AGGREGATE}^{(i)}$ and the $\text{COMBINE}^{(i)}$ functions using the neighborhood information, i.e., the node and edge attributes.

Consequently, the theorem can be proven given that there exist appropriate functions ∇ , $\text{AGGREGATE}^{(i)}$, $\text{COMBINE}^{(i)}$ and READOUT . For this purpose, the functions $\text{AGGREGATE}^{(i)}$ and $\text{COMBINE}^{(i)}$ must satisfy $\forall i \leq 2N-1$:

$$\begin{aligned} \nabla(\mathbf{T}_v^i) &= \mathbf{h}_v^i \\ &= \text{COMBINE}^{(i)}\left(\mathbf{h}_v^{(i-1)}, \text{AGGREGATE}^{(i)}\left(\{\mathbf{h}_u^{(i-1)}\}_{u \in ne_v}, \{\omega'_{\{u,v\}}\}_{u \in ne_v}\right)\right) \\ &= \text{COMBINE}^{(i)}\left(\nabla(\mathbf{T}_v^{i-1}), \text{AGGREGATE}^{(i)}\left(\{\nabla(\mathbf{T}_u^{i-1})\}_{u \in ne_v}, \Omega'_{ne_v}\right)\right). \end{aligned}$$

In a simple solution, $\text{AGGREGATE}^{(i)}$ decodes the attributed trees of the neighbors u of v , \mathbf{T}_u^{i-1} , and stores them into a data structure to be accessed by $\text{COMBINE}^{(i)}$. The detailed construction of the appropriate functions is given in D'Inverno et al. (2021). \square

Adopting an argument similar to that in Scarselli et al. (2008), it is proven that the previous theorem is equivalent to Theorem 5.1.3 and this concludes the proof. \square

A.5. Proof of Theorem 5.2.4

Let $G = (G_t)_{t \in I}$ be a discrete dynamic graph in the graph domain \mathcal{G} and $N = \max_{G \in \mathcal{G}} |G|$ be the maximal number of nodes in the domain. Let $\text{dyn}(t, G, v) \in \mathcal{F}(\mathcal{D})$ be any measurable dynamical system preserving the unfolding equivalence, $\|\cdot\|$ be a norm on \mathbb{R} , P be any probability measure on \mathcal{D} and ϵ, λ be any real numbers where $\epsilon, \lambda > 0$. Then, there exists a DGNN composed by SGNNs with $2N-1$ layers and hidden dimension $r = 1$, and Recurrent Neural Network state dimension $s = 1$ such that the function φ realized by this model satisfies

$$P(\|\text{dyn}(t, G, v) - \varphi(t, G, v)\| \leq \epsilon) \geq 1 - \lambda \quad \forall t \in I.$$

Proof. To prove the theorem above, we need some preliminary results. Using the same argument used for SAUHGs in 5.1.3, we need, as a preliminary lemma, the extension of Scarselli et al. (2008, Lem. 1) to the domain of dynamic graphs D , analogously to the extension to the domain of SAUHGs in A.4.1.

Lemma A.5.1. *Lemma A.4.1 holds for the domain of dynamic graphs D .*

Proof. Indeed, taking into account the argument in Thomas et al. (2021), one can establish a bijection between the domain of dynamic graphs and the domain of SAUHGs; on the latter, we can directly apply A.4.1. \square

Theorem 5.2.4 is equivalent to the following, where the domain contains a finite number of elements in D and the attributes are integers.

Theorem A.5.2. *For any finite set of p patterns*

$$\{(t^{(j)}, G^{(j)}, v^{(j)}) \mid (t^{(j)}, G^{(j)}, v^{(j)}) \in D, j \in [p]\}$$

with the maximal number of nodes $N = \max_{G \in \mathcal{G}} |G|$ and with graphs having integer features, for any measurable dynamical system preserving the unfolding equivalence, $\|\cdot\|$ be a norm on \mathbb{R} , P be any probability measure on D and ϵ be any real number where $\epsilon > 0$. Then, there exists a DGNN as defined in Definition 3.5 such that the function φ (realized by this model) computed after N steps satisfies the condition

$$\|dyn(t^{(j)}, G^{(j)}, v^{(j)}) - \varphi(t^{(j)}, G^{(j)}, v^{(j)})\| \leq \epsilon \quad (9)$$

$\forall j \in [p]$ where $t^{(j)} \in I$.

The equivalence between Theorems 5.2.4 and A.5.2 is formally proven by the following lemma.

Lemma A.5.3. *Theorem 5.2.4 holds if and only if Theorem A.5.2 holds.*

Proof. The proof is similar to the one contained in Scarselli et al. (2008). Nevertheless, we want to highlight that in this case, patterns are taken from $D := I \times \mathcal{G} \times \mathcal{V}$, as we are proving it in the context of the dynamic graphs. \square

Now, we can proceed to prove Theorem A.5.2.

Proof of Theorem A.5.2. The proof of this theorem involves assuming that the output dimension is $m = 1$, i.e., $dyn(t, G, v) \in \mathbb{R}$, but the result can be extended to the general case with $m \in \mathbb{N}$ by concatenating the corresponding results. As a result of Proposition 5.2.3, there exists a function κ , s.t. $dyn(t, G, v) = g(x_v(t)) = \kappa((\mathbf{T}_v(i))_{i \in [t]})$ where $\mathbf{T}_v(i)$ is an attributed unfolding tree. Given N_t as the number of nodes of the graph at timestep t , in order to store the graph information, an attributed unfolding tree of depth $2N_t - 1$ is required for each node, in such a way that κ can satisfy

$$dyn(t, G, v) = \kappa((\mathbf{T}_v(i))_{i \in [t]}) = \kappa((\mathbf{T}_v^{N_t}(i))_{i \in [t]}).$$

The required depth is a straight consequence of Theorem 4.1.3 in D'Inverno et al. (2021), which we briefly report here.

Theorem A.5.4. *D'Inverno et al., 2021, Theorem 4.1.3 The following statements hold for graphs with at most N nodes.*

1. Let \mathbf{G} and \mathbf{H} be connected graphs and x, y be nodes of \mathbf{G} and \mathbf{H} , respectively. The infinite unfolding trees $\mathbf{T}_x, \mathbf{T}_y$ are equal if and only if they are equal up to depth $2N - 1$, i.e., $\mathbf{T}_x = \mathbf{T}_y$ iff $\mathbf{T}_x^{2N-1} = \mathbf{T}_y^{2N-1}$.
2. For any N , there exist two graphs \mathbf{G} and \mathbf{H} with nodes x, y , respectively, such that the infinite unfolding trees $\mathbf{T}_x, \mathbf{T}_y$ are different, but they are equal up to depth $2N - 16\sqrt{N}$, i.e., $\mathbf{T}_x \neq \mathbf{T}_y$ and $\mathbf{T}_x^i = \mathbf{T}_y^i$ for $i \leq 2N - 16\sqrt{N}$. \square

The Theorem is tightly connected to the results previously displayed in Krebs and Verbisky (2015).

The main idea behind the proof of Theorem A.5.2 is to design a DGNN that can encode the sequence of attributed unfolding trees $(\mathbf{T}_v(i))_{i \in [t]}$ into the node attributes at each timestep t , i.e., $\mathbf{q}_v(t) = \#_t((\mathbf{T}_v(i))_{i \in [t]})$. This is achieved by using a coding function that maps sequences of $t + 1$ attributed trees into real numbers. To implement the encoding that could fit the definition of the DGNN, two coding functions are needed: the ∇ function, which encodes the attributed unfolding trees, and the family of coding functions $\#_t$. The composition of these functions is used to define the node's attributes, and the DGNN can produce the desired output by using this encoded information as follows:

$$\begin{aligned} \mathbf{q}_v(0) &= \mathbf{h}_v(0) = \#_0(\nabla^{-1}(\mathbf{h}_v(0))) \\ \mathbf{q}_v(t) &= \#_t(\text{APPEND}_t(\#_{t-1}^{-1}(\mathbf{q}_v(t-1)), \nabla^{-1}(\mathbf{h}_v(t)))) \end{aligned} \quad (10)$$

where the auxiliary function APPEND_t and the $\nabla, \#_t$ coding functions are defined in the following.

APPEND_t

Let $\mathcal{T}^d(v)$ be the domain of the attributed unfolding trees with root v , up to a certain depth d . The function $\text{APPEND}_t : \{(\mathbf{T}_v^d(i))_{i \in [t-1]}\} \cup \emptyset \times \mathcal{T}^d(v) \rightarrow \{(\mathbf{T}_v^d(i))_{i \in [t]}\}$ is defined as follows:

$$\begin{aligned} \text{APPEND}_0(\emptyset, \mathbf{T}_v^d(0)) &:= \mathbf{T}_v^d(0) \\ \text{APPEND}_t((\mathbf{T}_v^d(0), \dots, \mathbf{T}_v^d(t-1)), \mathbf{T}_v^d(t)) &:= (\mathbf{T}_v^d(0), \dots, \mathbf{T}_v^d(t-1), \mathbf{T}_v^d(t)) \end{aligned}$$

Intuitively, this function appends the unfolding tree snapshot of the node v at time t to the sequence of the unfolding trees of that node at the previous $t - 1$ timesteps.

In the following, the coding functions are defined; their existence and injectiveness are provided by construction.

The ∇ Coding Function

Let $\nabla := \mu_{\nabla} \circ \nu_{\nabla}$ be a composition of any two injective functions μ_{∇} and ν_{∇} with the following properties:

- μ_{∇} is an injective function from the domain of static unfolding trees, calculated on the nodes in the graph G_t , to the Cartesian product $\mathbb{N} \times \mathbb{N}^P \times \mathbb{Z}^A = \mathbb{N}^{P+1} \times \mathbb{Z}^A$, where P is the maximum number of nodes a tree could have.
- Intuitively, in the Cartesian product, \mathbb{N} represents the tree structure, \mathbb{N}^P denotes the node numbering, while, for each node, an integer vector in \mathbb{Z}^A is used to encode the node attributes. Notice that μ_{∇} exists and is injective since the maximal information contained in an unfolding tree is given by the union of all its node attributes and all its structural information, which just equals the dimension of the codomain of μ_{∇} .
- ν_{∇} is an injective function from $\mathbb{N}^{P+1} \times \mathbb{Z}^A$ to \mathbb{R} , whose existence is guaranteed by the cardinality theory, since the two sets have the same cardinality.

Since μ_{∇} and ν_{∇} are injective, also the existence and the injectiveness of ∇ is ensured.

The $\#_t$ Coding Family

Similarly to ∇ , the functions $\#_t := \mu_{\#_t} \circ \nu_{\#_t}$ are composed by two functions $\mu_{\#_t}$ and $\nu_{\#_t}$ with the following properties:

- $\mu_{\#_t}$ is an injective function from the domain of the dynamic unfolding trees $\mathcal{T}_t^d(v) := \{(\mathbf{T}_v^d(i))_{i \in [t]}\}$ to the Cartesian product $\mathbb{N}^t \times \mathbb{N}^{tP_t} \times \mathbb{Z}^{tA} = \mathbb{N}^{t(P_t+1)} \times \mathbb{Z}^{tA}$, where P_t is the maximum number of nodes a tree could have at time t .
- $\nu_{\#_t}$ is an injective function from $\mathbb{N}^{t(P_t+1)} \times \mathbb{Z}^{tA}$ to \mathbb{R} , whose existence is guaranteed by the cardinality theory, since the two sets have the same cardinality.

Since $\mu_{\#_i}$ and $\nu_{\#_i}$ are injective, also the existence and the injectiveness of $\#_i$ are ensured.

The recursive function f , $\text{AGGREGATE}_t^{(i)}$, $\text{COMBINE}_t^{(i)}$

The recursive function f has to satisfy

$$f(\mathbf{q}_v(t-1), \mathbf{h}_v(t)) = \#_i((\mathbf{T}_v(i))_{i \in [t]}) = \mathbf{q}_v(t),$$

where the $\mathbf{h}_v(t)$ is the hidden representation of node v at time t extracted from the t th SGNN, i.e., $\mathbf{h}_v(t) = \text{SGNN}(G_t, v)$. In particular, at each iteration i , we have

$$\mathbf{h}_v^i(t) = \text{COMBINE}_t^{(i)}(\mathbf{h}_v^{i-1}(t), \text{AGGREGATE}_t^{(i)}(\{\mathbf{h}_u^{i-1}(t)\}_{u \in ne_v(t)}, \{\omega_{\{u,v\}}(t)\}_{u \in ne_v(t)}))$$

Further, the functions $\text{AGGREGATE}_t^{(i)}$ and $\text{COMBINE}_t^{(i)}$ – following the proof in D’Inverno et al. (2021) – must satisfy

$$\begin{aligned} \nabla(\mathbf{T}_v^i(t)) &= \mathbf{h}_v^i(t) = \\ & \text{COMBINE}_t^{(i)}(\mathbf{h}_v^{i-1}(t), \text{AGGREGATE}_t^{(i)}(\{\mathbf{h}_u^{i-1}(t)\}_{u \in ne_v(t)}, \{\omega_{\{u,v\}}(t)\}_{u \in ne_v(t)})) \\ &= \text{COMBINE}_t^{(i)}(\nabla(\mathbf{T}_v^{i-1}(t)), \text{AGGREGATE}_t^{(i)}(\{\nabla(\mathbf{T}_u^{i-1}(t))\}_{u \in ne_v(t)})) \end{aligned}$$

$\forall i \leq 2N - 1$ and $\forall t \in I$.

For example, the trees can be collected into the coding of a new tree, i.e.,

$$\text{AGGREGATE}_t^{(i)}(\nabla(\mathbf{T}_u^{i-1}(t)), u \in ne_v(t)) = \nabla(\cup_{u \in ne_v(t)} \nabla^{-1}(\nabla(\mathbf{T}_u^{i-1}(t))))$$

where $\cup_{u \in ne_v(t)}$ denotes an operator that constructs a tree with a root having void attributes from a set of subtrees (see Fig. 7). Then, $\text{COMBINE}_t^{(i)}$ assigns the correct attributes to the root by extracting them from $\mathbf{T}_v^{i-1}(t)$, i.e.,

$$\text{COMBINE}_t^{(i)}(\nabla(\mathbf{T}_v^{i-1}(t)), b) = \nabla(\text{ATTACH}(\nabla^{-1}(\nabla(\mathbf{T}_v^{i-1}(t))), \nabla^{-1}(b))),$$

where ATTACH is an operator that returns a tree constructed by replacing the attributes of the root in the latter tree with those of the former tree and b is the result of the $\text{AGGREGATE}_t^{(i)}$ function.

Now, notice that, with this definition, $\text{AGGREGATE}_t^{(i)}$, $\text{COMBINE}_t^{(i)}$, and $\text{READOUT}_{\text{dyn}}$ may not be differentiable. Nevertheless, Eq. (9) has to be satisfied only for a finite number of graphs, namely G_j . Thus, we can specify other functions $\overline{\text{AGGREGATE}}_t^{(i)}$, $\overline{\text{COMBINE}}_t^{(i)}$, and $\overline{\text{READOUT}}$, which produce exactly the same computations when they are applied on the graphs G_j , but that can be extended to the rest of their domain so that they are continuously differentiable. Obviously, such an extension exists since those functions are only constrained to interpolate a finite number of points.⁵

The $\text{READOUT}_{\text{dyn}}$ function

Eventually, $\text{READOUT}_{\text{dyn}}$ must satisfy:

$$\kappa(\cdot) := \text{READOUT}_{\text{dyn}}(\#_i(\cdot))$$

so that, ultimately,

$$\text{dyn}(t, G, v) =$$

$$\text{READOUT}_{\text{dyn}}(\#_i(\text{APPEND}_t(\#_{t-1}^{-1}(\mathbf{q}_v(t-1)), \nabla^{-1}(\mathbf{h}_v(t)))) \quad \square$$

This concludes the proof via Lemma A.5.3. \square

A.6. Proof of Theorem 5.2.6

Assume that the hypotheses of Theorem 5.2.4 are fulfilled and \mathcal{Q}_D is a class of discrete DGNNs with universal components. Then, there exists a parameter set θ , and the functions $\text{SGNN}(t)_\theta$, $f_{0,\theta}$, f_θ , implemented by Neural Networks in \mathcal{Q}_D , such that the thesis of Theorem 5.2.4 holds.

⁵ Notice that a similar extension can also be applied to the coding function ∇ and to the decoding function ∇^{-1} . In this case, the coding function is not injective on the whole domain, but only on the graphs mentioned in the theorem.

Attach

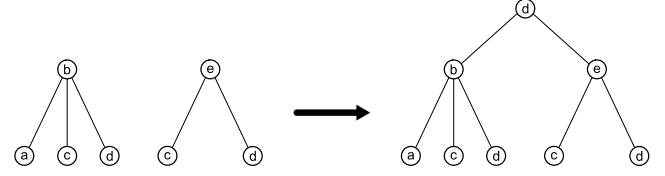


Fig. 7. The ATTACH operator on trees.

Proof. The idea of the proof follows from the same reasoning adopted in D’Inverno et al. (2021). Intuitively, since the discrete DGNN of Theorem 5.2.4 is implemented by continuously differentiable functions, its output depends continuously on the possible changes in the DGNN implementation: small changes in the function implementation cause small changes in the DGNN outputs. Therefore, the functions of the DGNN of Theorem 5.2.4 can be replaced by Neural Networks, provided that those networks are suitable approximators.

As in the proof of the dynamic version of the approximation theorem, cf. Theorem 5.2.4, without loss of generality, we will assume that the attribute dimension is $n = 1$.⁶

First of all, note that Theorem 5.2.4 ensures that we can find continuously differentiable functions \bar{f} , $\overline{\text{READOUT}}_{\text{dyn}}$ such that, for the corresponding function $\bar{\varphi}$ implemented by the DGNN it holds:

$$P(\|\text{dyn}(t, G, v) - \bar{\varphi}(t, G, v)\| \leq \frac{\epsilon}{2}) \geq 1 - \lambda \quad \forall t \in I, \epsilon, \lambda > 0. \quad (11)$$

Considering that the theorem has to hold only in probability, we can also assume that the domain is bounded to a finite set of patterns $\{(t^{(i)}, (G_t)_{t \in I}, v^{(i)}) \mid i = 1, \dots, p\}$ (as in Theorem A.5.2). As a result, the functions, \bar{f} and $\overline{\text{READOUT}}_{\text{dyn}}$ are bounded and have a bounded Jacobian. We can take the maximum of these Jacobians, which we will denote as B .

Moreover, let f_θ , $\text{READOUT}_{\text{dyn},\theta}$ be universal components for DGNN, as in Definition 5.2.5, that approximate \bar{f} , $\overline{\text{READOUT}}_{\text{dyn}}$, respectively. Further, let $\epsilon_1, \epsilon_2, > 0$ be the corresponding approximation errors, i.e.,

$$\begin{aligned} \|\bar{f}(\mathbf{q}, \mathbf{h}) - f_\theta(\mathbf{q}, \mathbf{h})\|_\infty &\leq \epsilon_1, \text{ and} \\ \|\overline{\text{READOUT}}_{\text{dyn}}(Q(t)) - \text{READOUT}_{\text{dyn},\theta}(Q(t))\|_\infty &\leq \epsilon_2 \end{aligned} \quad (12)$$

hold $\forall t \in I$.

Now, from the proof of Theorem 5.1.5 we know that

$$P(\|\overline{\text{SGNN}}_i(G, v) - \text{SGNN}_{\theta,i}(G, v)\| \leq \epsilon_s) \geq 1 - \lambda_i$$

for $i \in [t]$, $\epsilon_s > 0$ and for any norm. Then we can take every λ_i small enough, s.t.

$$\|\overline{\text{SGNN}}_i(G, v) - \text{SGNN}_{\theta,i}(G, v)\|_\infty \leq \epsilon_s$$

holds on a finite set of patterns large enough to include those ones of the i th timestep of each patterns of dynamic graphs on which Eq. (11) holds.

Therefore, if we define $\bar{\mathbf{h}}(t) := \overline{\text{SGNN}}_i(G_t)$ and $\mathbf{h}_\theta(t) := \text{SGNN}_{\theta,i}(G_t)$ we have

$$\|\bar{\mathbf{h}}(t) - \mathbf{h}_\theta(t)\|_\infty = \|\overline{\text{SGNN}}_i(G_t) - \text{SGNN}_{\theta,i}(G_t)\|_\infty \leq \epsilon_s.$$

In addition, let $\bar{H}(t)$ and $H_\theta(t)$ be the internal representations produced by $\overline{\text{SGNN}}$ and SGNN_θ , stacked over all the nodes of the input graph. Then it holds

$$\|\bar{H}(t) - H_\theta(t)\|_\infty \leq N\epsilon_s, \quad \forall t \in I, \quad (13)$$

where $N = \max_{G \in \mathcal{G}} |G|$ is the maximum number of nodes of the static graphs input in the bounded domain. Let again $\bar{Q}(0) := \bar{H}(0)$ and

⁶ A GNN can theoretically be modeled with multiple components by stacking Neural Networks for each dimension, respectively.

$\bar{Q}(t) := \bar{F}(\bar{Q}(t-1), \bar{H}(t))$ be the stacking of the internal states produced by DGNN's internal recursive function \bar{f} . Analogously, let $Q_\theta(0) := H_\theta(0)$ and $Q_\theta(t) := F_\theta(Q_\theta(t-1), H_\theta(t))$ be the output produced by the corresponding function of the parameterized DGNN.

Then it holds:

$$\|\bar{Q}(0) - Q_\theta(0)\|_\infty = \|\bar{H}(0) - H_\theta(0)\|_\infty \leq N\epsilon_s \quad (14)$$

and

$$\begin{aligned} \|\bar{f}(\bar{Q}(0), \cdot) - \bar{f}(Q_\theta(0), \cdot)\|_\infty &\leq B\|\bar{Q}(0) - Q_\theta(0)\|_\infty \\ \|\bar{f}(\cdot, \bar{H}(1)) - \bar{f}(\cdot, H_\theta(1))\|_\infty &\leq B\|\bar{H}(1) - H_\theta(1)\|_\infty \end{aligned}$$

for a bound B on the Jacobian of $\bar{f}(\mathbf{q}, \mathbf{h}) \forall t \in I$ and $\forall \mathbf{q}$, which, along with Eqs. (13) and (14) gives

$$\begin{aligned} \|\bar{f}(\bar{Q}(0), \cdot) - \bar{f}(Q_\theta(0), \cdot)\|_\infty &\leq N\epsilon_s B \\ \|\bar{f}(\cdot, \bar{H}(1)) - \bar{f}(\cdot, H_\theta(1))\|_\infty &\leq N\epsilon_s B \end{aligned} \quad (15)$$

Therefore, we have that:

$t = 1$:

$$\begin{aligned} &\|\bar{Q}(1) - Q_\theta(1)\|_\infty \\ &= \|\bar{f}(\bar{Q}(0), \bar{H}(1)) - f_\theta(Q_\theta(0), H_\theta(1))\|_\infty \\ &\stackrel{\text{add } 0}{=} \|\bar{f}(\bar{Q}(0), \bar{H}(1)) - \bar{f}(Q_\theta(0), \bar{H}(1)) \\ &\quad + \bar{f}(Q_\theta(0), \bar{H}(1)) - \bar{f}(Q_\theta(0), H_\theta(1)) \\ &\quad + \bar{f}(Q_\theta(0), H_\theta(1)) - f_\theta(Q_\theta(0), H_\theta(1))\|_\infty \\ &\stackrel{\Delta\text{-ineq.}}{\leq} \|\bar{f}(\bar{Q}(0), \bar{H}(1)) - \bar{f}(Q_\theta(0), \bar{H}(1))\|_\infty \\ &\quad + \|\bar{f}(Q_\theta(0), \bar{H}(1)) - \bar{f}(Q_\theta(0), H_\theta(1))\|_\infty \\ &\quad + \|\bar{f}(Q_\theta(0), H_\theta(1)) - f_\theta(Q_\theta(0), H_\theta(1))\|_\infty \\ &\stackrel{(15)}{\leq} 2N\epsilon_s B + N\epsilon_1 \\ &:= \lambda_1(\epsilon_s, \epsilon_1). \end{aligned}$$

$t > 0$: Analogously, it follows for $t > 1$ that

$$\begin{aligned} &\|\bar{Q}(t) - Q_\theta(t)\|_\infty \\ &= \|\bar{f}(\bar{Q}(t-1), \bar{H}(t)) - f_\theta(Q_\theta(t-1), H_\theta(t))\|_\infty \\ &= \|\bar{f}(\bar{Q}(t-1), \bar{H}(t)) - \bar{f}(Q_\theta(t-1), \bar{H}(t)) \\ &\quad + \bar{f}(Q_\theta(t-1), \bar{H}(t)) - \bar{f}(Q_\theta(t-1), H_\theta(t)) \\ &\quad + \bar{f}(Q_\theta(t-1), H_\theta(t)) - f_\theta(Q_\theta(t-1), H_\theta(t))\|_\infty \\ &\leq \|\bar{f}(\bar{Q}(t-1), \bar{H}(t)) - \bar{f}(Q_\theta(t-1), \bar{H}(t))\|_\infty \\ &\quad + \|\bar{f}(Q_\theta(t-1), \bar{H}(t)) - \bar{f}(Q_\theta(t-1), H_\theta(t))\|_\infty \\ &\quad + \|\bar{f}(Q_\theta(t-1), H_\theta(t)) - f_\theta(Q_\theta(t-1), H_\theta(t))\|_\infty \\ &\leq N\lambda_0 B + N\epsilon_s B + N\epsilon_1 \\ &:= \lambda_1(\epsilon_s, \epsilon_1). \end{aligned}$$

The above reasoning can then be applied recursively to prove that

$$\|\bar{Q}(t) - Q_\theta(t)\|_\infty \leq \lambda_t(\epsilon_s, \epsilon_1),$$

where $\lambda_t(\epsilon_s, \epsilon_1)$ could be found as little as possible, according to ϵ_s, ϵ_1 . Finally, let $\epsilon_2 > 0$, so that

$$\begin{aligned} &\|\bar{\varphi}(t, G, v) - \varphi_\theta(t, G, v)\|_\infty \\ &= \|\overline{\text{READOUT}}_{\text{dyn}}(\bar{Q}(t)) - \overline{\text{READOUT}}_{\text{dyn}, \theta}(Q_\theta(t))\|_\infty \\ &\leq \|\overline{\text{READOUT}}_{\text{dyn}}(\bar{Q}(t)) - \overline{\text{READOUT}}_{\text{dyn}}(Q_\theta(t))\|_\infty \\ &\quad + \|\overline{\text{READOUT}}_{\text{dyn}}(Q_\theta(t)) - \overline{\text{READOUT}}_{\text{dyn}, \theta}(Q_\theta(t))\|_\infty \\ &\leq \lambda_t B + \epsilon_2 = \lambda(\epsilon_s, \epsilon_1, \epsilon_2). \end{aligned}$$

Thus, we choose $\epsilon_s, \epsilon_1, \epsilon_2$, s.t. $\lambda \leq \frac{\epsilon}{2}$; going back in probability, we obtain

$$P(\|\bar{\varphi}(t, G, v) - \varphi_\theta(t, G, v)\| \leq \frac{\epsilon}{2}) \geq 1 - \lambda \quad \forall t \in I,$$

which, along with Eq. (11), proves the result. \square

References

- Abboud, R., Ceylan, İ. İ., Grohe, M., & Lukasiewicz, T. (2021). The surprising power of graph neural networks with random node initialization. In Z. Zhou (Ed.), *Proceedings of the thirtieth international joint conference on artificial intelligence* (pp. 2112–2118). ijcai.org.
- Azizian, W., & Lelarge, M. (2021). Expressive power of invariant and equivariant graph neural networks. In *9th International conference on learning representations*. OpenReview.net.
- Barcelo, P., Galkin, M., Morris, C., & Orth, M. R. (2022). Weisfeiler and leman go relational. In *Learning on graphs conference* (pp. 46:1–46:26). PMLR.
- Barros, C. D., Mendonça, M. R., Vieira, A. B., & Ziviani, A. (2021). A survey on embedding dynamic graphs. *ACM Computing Surveys*, 55(1), 1–37.
- Battaglia, P. W., Hamrick, J. B., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V. F., Malinowski, M., et al. (2018). Relational inductive biases, deep learning, and graph networks. arXiv preprint arXiv:1806.01261.
- Bodnar, C., Frasca, F., Otter, N., Wang, Y. G., Liò, P., Montufar, G. F., et al. (2021). Weisfeiler and Lehman Go Cellular: CW Networks. *Advances in Neural Information Processing Systems*, 34.
- Bodnar, C., Frasca, F., Wang, Y., Otter, N., Montufar, G. F., Lio, P., et al. (2021). Weisfeiler and Lehman Go Topological: Message Passing Simplicial Networks. In *International conference on machine learning* (pp. 1026–1037). PMLR.
- Bronstein, M. M., Bruna, J., Cohen, T., & Velickovic, P. (2021). Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. CoRR abs/2104.13478.
- Bruna, J., Zaremba, W., Szlam, A., & LeCun, Y. (2014). Spectral networks and locally connected networks on graphs. In *ICLR 2014*.
- Dasoulas, G., Santos, L. D., Scaman, K., & Virmaux, A. (2020). Coloring graph neural networks for node disambiguation. In C. Bessiere (Ed.), *Proceedings of the twenty-ninth international joint conference on artificial intelligence* (pp. 2126–2132). ijcai.org, <http://dx.doi.org/10.24963/ijcai.2020/294>.
- Deng, S., Rangwala, H., & Ning, Y. (2019). Learning dynamic context graphs for predicting social events. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1007–1016).
- D'Inverno, G. A., Bianchini, M., Sampoli, M. L., & Scarselli, F. (2021). On the approximation capability of GNNs in node classification/regression tasks. arXiv preprint arXiv:2106.08992.
- Garg, V., Jegelka, S., & Jaakkola, T. (2020). Generalization and representational limits of graph neural networks. In *International conference on machine learning* (pp. 3419–3430). PMLR.
- Grohe, M., & Schweitzer, P. (2020). The graph isomorphism problem. *Communications of the ACM*, 63(11), 128–134.
- Hamilton, W., Ying, Z., & Leskovec, J. (2017a). Inductive representation learning on large graphs. In *Advances in neural information processing systems* (pp. 1024–1034).
- Hamilton, W. L., Ying, R., & Leskovec, J. (2017b). Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584.
- Hammer, B. (2000). On the approximation capability of recurrent neural networks. *Neurocomputing*, 31(1–4), 107–123.
- Jegelka, S. (2022). Theory of graph neural networks: Representation and learning. arXiv preprint arXiv:2204.07697.
- Kazemi, S. M., Goel, R., Jain, K., Kobyzev, I., Sethi, A., Forsyth, P., et al. (2020). Representation learning for dynamic graphs: A survey. *Journal of Machine Learning Research*, 21, 70:1–70:73.
- Kiefer, S., & McKay, B. D. (2020). The iteration number of colour refinement. In *47th International colloquium on automata, languages, and programming*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.
- Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In *ICLR 2017*.
- Krebs, A., & Verbitsky, O. (2015). Universal covers, color refinement, and two-variable counting logic: Lower bounds for the depth. In *2015 30th annual ACM/IEEE symposium on logic in computer science* (pp. 689–700). IEEE.
- Leman, A., & Weisfeiler, B. (1968). A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Tekhnicheskaya Informatsiya*, 2(9), 12–16.
- Li, Y., Tarlow, D., Brockschmidt, M., & Zemel, R. (2016). Gated graph sequence neural networks. In Y. Bengio, & Y. LeCun (Eds.), *4th International conference on learning representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, conference track proceedings*.
- Longa, A., Lachi, V., Santini, G., Bianchini, M., Lepri, B., Lio, P., et al. (2023). Graph neural networks for temporal graphs: State of the art, open challenges, and opportunities. *Transactions on Machine Learning Research*.
- Loukas, A. (2020). What graph neural networks cannot learn: depth vs width. In *8th International conference on learning representations*. OpenReview.net.
- Luo, S., Shi, C., Xu, M., & Tang, J. (2021). Predicting molecular conformation via dynamic graph score matching. *Advances in Neural Information Processing Systems*, 34, 19784–19795.
- Maron, H., Ben-Hamu, H., Serviansky, H., & Lipman, Y. (2019). Provably powerful graph networks. In *Advances in neural information processing systems: vol. 32*.
- Maron, H., Ben-Hamu, H., Shamir, N., & Lipman, Y. (2019). Invariant and equivariant graph networks. In *7th International conference on learning representations*. OpenReview.net.

- Micheli, A. (2009). Neural network for graphs: A contextual constructive approach. *IEEE Transactions on Neural Networks*, 20(3), 498–511.
- Morris, C., Ritzert, M., Fey, M., Hamilton, W. L., Lenssen, J. E., Rattan, G., et al. (2019). Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks. In *The thirty-third AAAI conference on artificial intelligence, AAAI 2019, the thirty-first innovative applications of artificial intelligence conference, IAAI 2019, the ninth AAAI symposium on educational advances in artificial intelligence* (pp. 4602–4609). AAAI Press, <http://dx.doi.org/10.1609/aaai.v33i01.33014602>.
- Narayan, A., & Roe, P. H. (2018). Learning graph dynamics using deep neural networks. *IFAC-PapersOnLine*, 51(2), 433–438.
- Niepert, M., Ahmed, M., & Kutzkov, K. (2016). Learning convolutional neural networks for graphs. In *International conference on machine learning* (pp. 2014–2023). PMLR.
- Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., & Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. arXiv preprint arXiv:2006.10637.
- Sato, R. (2020). A survey on the expressive power of graph neural networks. ArXiv abs/2003.04078.
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2008). Computational capabilities of graph neural networks. *IEEE Transactions on Neural Networks*, 20(1), 81–102.
- Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., & Monfardini, G. (2009). The graph neural network model. *IEEE Transactions on Neural Networks*, 20, 61–80.
- Seo, Y., Defferrard, M., Vandergheynst, P., & Bresson, X. (2018). Structured sequence modeling with graph convolutional recurrent networks. In *International conference on neural information processing* (pp. 362–373). Springer.
- Shervashidze, N., Schweitzer, P., Van Leeuwen, E. J., Mehlhorn, K., & Borgwardt, K. M. (2011). Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*.
- Skardinga, J., Gabrys, B., & Musial, K. (2021). Foundations and modelling of dynamic networks using dynamic graph neural networks: A survey. *IEEE Access*.
- Taheri, A., Gimpel, K., & Berger-Wolf, T. (2019). Learning to represent the evolution of dynamic graphs with recurrent models. In *Companion proceedings of the 2019 world wide web conference* (pp. 301–307).
- Thomas, J. M., Beddar-Wiesing, S., & Moallem-Oureh, A. (2021). Graph type expressivity and transformations. arXiv:2109.10708.
- Thomas, J., Moallem-Oureh, A., Beddar-Wiesing, S., & Holzhüter, C. (2022). Graph neural networks designed for different graph types: A survey. *Transactions on Machine Learning Research*.
- Trivedi, R., Dai, H., Wang, Y., & Song, L. (2017). Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In *International conference on machine learning* (pp. 3462–3471). PMLR.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In *ICLR 2018*.
- Wu, J., Cao, M., Cheung, J. C. K., & Hamilton, W. L. (2020). Temp: Temporal message passing for temporal knowledge graph completion. arXiv preprint arXiv:2010.03526.
- Xu, K., Hu, W., Leskovec, J., & Jegelka, S. (2019). How powerful are graph neural networks? In *7th international conference on learning representations*. OpenReview.net.
- Xu, D., Ruan, C., Korpeoglu, E., Kumar, S., & Achan, K. (2020). Inductive representation learning on temporal graphs. arXiv preprint arXiv:2002.07962.
- Xue, G., Zhong, M., Li, J., Chen, J., Zhai, C., & Kong, R. (2022). Dynamic network embedding survey. *Neurocomputing*, 472, 212–223.
- You, J., Gomes-Selman, J., Ying, R., & Leskovec, J. (2021). Identity-aware graph neural networks. In *Thirty-Fifth AAAI conference on artificial intelligence, AAAI 2021, thirty-third conference on innovative applications of artificial intelligence, IAAI 2021, the eleventh symposium on educational advances in artificial intelligence* (pp. 10737–10745). AAAI Press.
- Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. arXiv preprint arXiv:1709.04875.
- Zhang, M., & Li, P. (2021). Nested graph neural networks. *Advances in Neural Information Processing Systems*, 34, 15734–15747.
- Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81.