

# Expressive Power of Temporal Message Passing

Przemysław Andrzej Wałęga<sup>1,2</sup>, Michael Rawson<sup>3</sup>

<sup>1</sup>Queen Mary University of London, UK

<sup>2</sup>University of Oxford, UK

<sup>3</sup>University of Southampton, UK

p.walega@qmul.ac.uk, michael@rawsons.uk

## Abstract

Graph neural networks (GNNs) have recently been adapted to temporal settings, often employing temporal versions of the message-passing mechanism known from GNNs. We divide temporal message passing mechanisms from literature into two main types: global and local, and establish Weisfeiler-Leman characterisations for both. This allows us to formally analyse expressive power of temporal message-passing models. We show that global and local temporal message-passing mechanisms have incomparable expressive power when applied to arbitrary temporal graphs. However, the local mechanism is strictly more expressive than the global mechanism when applied to colour-persistent temporal graphs, whose node colours are initially the same in all time points. Our theoretical findings are supported by experimental evidence, underlining practical implications of our analysis.

## Introduction

*Message-passing graph neural networks* (or MP-GNNs) (Gilmer et al. 2017) are prominent models for graph learning, which have achieved state-of-the-art performance in tasks of link prediction as well as in node and graph classification. Importantly, they proved successful in a number of real-world applications including social networks, protein-protein interactions, and knowledge graphs (Zhou et al. 2020).

In recent years, there has been growing interest in adapting MP-GNNs to process *temporal graphs* (for an example of a temporal graph see the left part of Figure 1) which are particularly well-suited for dynamic applications such as recommender systems (Wu et al. 2023), traffic forecasting (Yu, Yin, and Zhu 2018), finance networks (Pareja et al. 2020), and modelling the spread of diseases (Kapoor et al. 2020). Research in this direction gave rise to various temporal MP-GNNs (MP-TGNNs) obtained by introducing temporal variants of the message-passing mechanism (Longa et al. 2023; Skarding, Gabrys, and Musial 2021; Gao and Ribeiro 2022). This can be obtained by assigning to graph nodes different embeddings (feature vectors) for different time points and then passing messages between timestamped nodes. Depending on the routes of messages-passing between timestamped nodes and on the encoding of the tempo-

ral component in the messages, we arrive at various temporal message-passing mechanisms. In this paper we distinguish two main groups of MP-TGNNs: *global*, where messages can be passed between nodes stamped with different times and *local*, where messages are passed only between nodes stamped with the same time, while information about other times is encoded within messages.

Although several variants of global (Longa et al. 2023; Xu et al. 2020; Luo and Li 2022) and local (Rossi et al. 2020; Qu et al. 2020) MP-TGNNs have been designed and successfully applied, we still lack a good understanding of modelling capabilities dictated by their temporal message-passing mechanisms, and do not have answers to the following fundamental questions. What tools can we use to analyse the expressive power of global and local MP-TGNNs? Are there limits to the expressive power of either type? Which type can express more? How does the difference in expressiveness affect practical performance? Answers to these questions are key when choosing an appropriate temporal message-passing mechanism for a particular task and when designing new MP-TGNNs. The importance of answering such questions has been clearly shown by research on expressive power of static MP-GNNs, which equipped us with powerful tools and gave rise to a whole new research direction (Morris et al. 2019; Xu et al. 2019; Cai, Fürer, and Immerman 1992; Grohe 2023; Barceló et al. 2020). In the temporal setting, however, such an analysis is still missing. We aim to fill this urgent gap.

**Contributions.** Our main contributions are as follows:

- We formalise the two main types of MP-TGNNs, global and local, depending on the form of the adopted temporal message-passing mechanism.
- We characterise the expressive power of both types. To determine which temporal nodes can be distinguished by MP-TGNNs, we construct a knowledge graph and then apply the 1-dimensional Weisfeiler-Leman test (1-WL). As depicted in Figure 1, our construction of the knowledge graph is different for global and local MP-TGNNs, but in both cases this approach allows us to precisely capture the expressive power of MP-TGNNs. For example, given the temporal graph  $TG$  in Figure 1, global and local MP-TGNNs distinguish the same nodes, since the colourings in the rightmost graphs are the same.

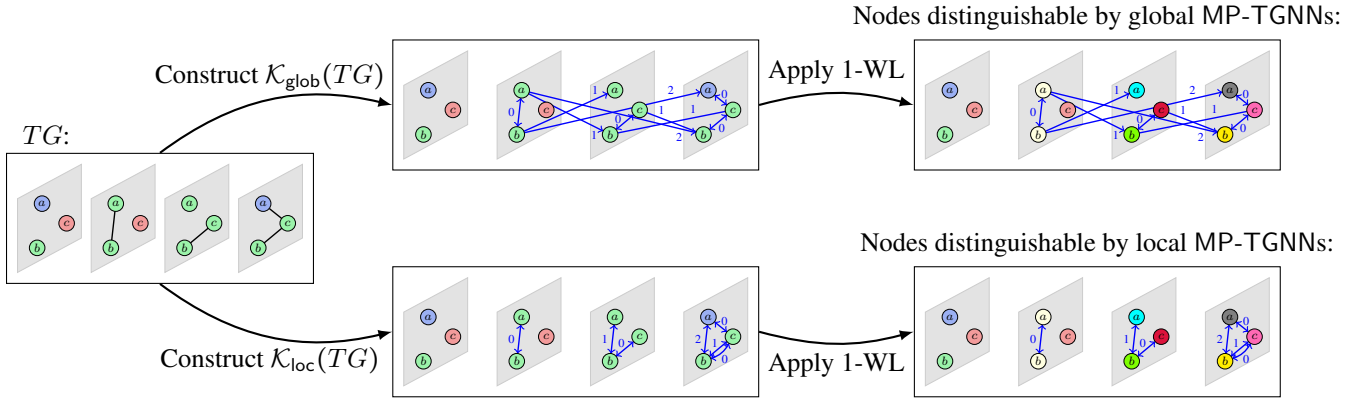


Figure 1: Our approach to determine which nodes in a temporal graph  $TG$  are distinguishable by MP-TGNN: we construct of knowledge graphs  $\mathcal{K}_{\text{glob}}(TG)$  and  $\mathcal{K}_{\text{loc}}(TG)$ , and then apply 1-WL

- We use the above characterisation to show that, quite surprisingly, both global and local MP-TGNNs can distinguish nodes which are pointwise isomorphic. This leads us to introduce a stronger *timewise isomorphism*, well-suited for characterisation of MP-TGNNs.
- The Weisfeiler-Leman characterisation also allows us to show that global and local MP-TGNNs have incomparable expressive power: each of the types can distinguish nodes which are indistinguishable by the other type. However, if the input temporal graph is *colour-persistent* (initial embedding of each node is the same at all time points), local MP-TGNNs are more expressive than global MP-TGNNs. We can extend these results to a complete expressiveness classification as in Figure 2.

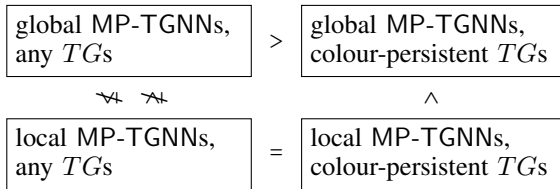


Figure 2: Relative expressive power of MP-TGNNs

- Finally, we experimentally validate our theoretical results by constructing proof-of-concept global and local models. We show that, indeed, on colour-persistent graphs local models outperform global models, when compared on the temporal link-prediction and TGB 2.0 benchmark suite (Gastinger et al. 2024). This is the case when models use the same number of layers, and the difference in performance increases further if we choose optimal number of layers for each type of model separately.

## Background

**Temporal graphs.** We focus on temporal graphs in the so-called *snapshot* representation (Longa et al. 2023; Gao and Ribeiro 2022; Skarding, Gabrys, and Musial 2021) shown in Figure 3. A *temporal graph* is a finite sequence  $TG = (G_1, t_1), \dots, (G_n, t_n)$  of undirected node-coloured graphs

$G_i = (V_i, E_i, c_i)$ , where  $t_1 < \dots < t_n$  are real-valued *time points*, constituting the temporal domain  $\text{time}(TG)$ . Each  $V_i$  is a finite set of nodes, each  $E_i \subseteq \{\{u, v\} \subseteq V_i \mid u \neq v\}$  is a set of undirected edges, and  $c_i : V_i \rightarrow D$  assigns nodes colours from some set  $D$ , which could be real feature vectors. Following standard notation, we sometimes use  $\mathbf{x}_v(t_i)$  instead of  $c_i(v)$ . We represent  $c_i$  using different colours for nodes in figures. We assume that the domain of nodes does not change over time, so  $V_1 = \dots = V_n = V(TG)$ . We call a pair of a node  $v \in V(TG)$  and a time point  $t \in \text{time}(TG)$  a *timestamped node*  $(v, t)$  and we let  $t\text{-nodes}(TG)$  be the set of all timestamped nodes in  $TG$ . For the sake of a clear presentation we assume that edges are not labelled.

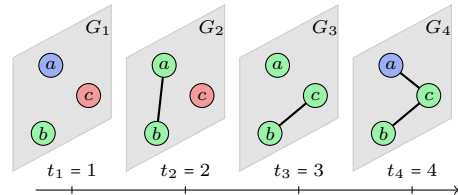


Figure 3: A temporal graph in the snapshot representation

We say that a temporal graph is *colour-persistent* if initial colours of nodes do not change in time, so  $c_i(v) = c_j(v)$  for each node  $v$  and all  $i, j \in \{1, \dots, n\}$ , see Figure 4 (a). Colour-persistent graphs can be also represented as static edge-labelled multi graphs, called *aggregated* form (Gao and Ribeiro 2022), as depicted in Figure 4 (b).

**Temporal graph neural networks with message-passing.** We let a *message-passing temporal graph neural network* (MP-TGNN) be a model  $\mathcal{A}$  which, given a temporal graph  $TG$ , computes embeddings for all timestamped nodes by implementing a temporal variant of message-passing. Embeddings are then used to predict links or classify nodes and graphs. Some models (e.g. TGAT, NAT, and TGN) apply temporal message-passing to arbitrary temporal graphs, whereas others (e.g. TDGNN) are applicable to colour-persistent temporal graphs (or equivalently, to the aggregated representation) only. Below we present a general form

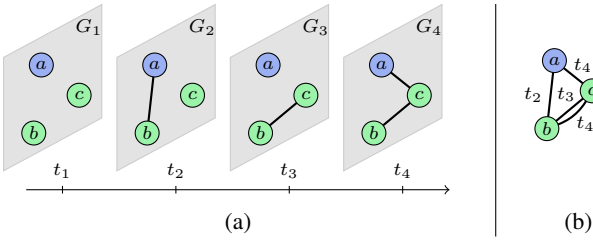


Figure 4: A colour-persistent temporal graph (a) and its aggregated representation (b)

of an MP-TGNN model  $\mathcal{A}$  with  $L$  layers, which subsumes a number of message-passing mechanisms. Given a temporal graph  $TG = (G_1, t_1), \dots, (G_n, t_n)$ , a model  $\mathcal{A}$  computes for each node  $v$ , time point  $t$ , and layer  $\ell \in \{0, \dots, L\}$  an embedding  $\mathbf{h}_v^{(\ell)}(t)$  as follows:

$$\mathbf{h}_v^{(0)}(t) = \mathbf{x}_v(t), \quad (1)$$

$$\mathbf{h}_v^{(\ell)}(t) = \text{COM}^{(\ell)}\left(\mathbf{h}_v^{(\ell-1)}(t), \text{AGG}^{(\ell)}\left(\left\{\left\{\star, g(t-t')\right\} \mid (u, t') \in \mathcal{N}(v, t)\right\}\right)\right), \quad (2)$$

where:

- $\text{COM}^{(\ell)}$  and  $\text{AGG}^{(\ell)}$  are *combination* and *aggregation* functions in layer  $\ell$ ;  $\text{COM}^{(\ell)}$  maps a pair of vectors into a single vector, whereas  $\text{AGG}^{(\ell)}$  maps a multiset, represented as  $\{\{\dots\}\}$ , into a single vector,
- $g$  maps a time duration into a vector or scalar quantity,
- $\mathcal{N}(v, t)$  is the temporal neighbourhood of  $(v, t)$ , defined as follows (Souza et al. 2022):

$$\mathcal{N}(v, t) = \left\{ (u, t') \mid t' = t_i \text{ and } \{u, v\} \in E_i, \right. \\ \left. \text{for some } (G_i, t_i) \in TG \text{ with } t_i \leq t \right\}. \quad (3)$$

Hence,  $\mathcal{N}(v, t)$  is the set of timestamped nodes  $(u, t')$  such that there is an edge between  $u$  and  $v$  at  $t' \leq t$ .

- $\star$  is either  $\mathbf{h}_u^{(\ell-1)}(t')$  or  $\mathbf{h}_u^{(\ell-1)}(t)$ . If  $\star = \mathbf{h}_u^{(\ell-1)}(t')$  we say that  $\mathcal{A}$  is a *global* (in time) TGNN, as computation of  $\mathbf{h}_v^{(\ell)}(t)$  requires aggregation of embeddings in all past time points  $t'$ ; *global* TGNNs are also called *Temporal Embedding* TGNNs (Longa et al. 2023) and include TGAT and NAT. If  $\star = \mathbf{h}_u^{(\ell-1)}(t)$  we say that  $\mathcal{A}$  is *local*, as only embeddings from the current time point  $t$  are aggregated; *local* TGNNs include TGN and TDGNN.

**Weisfeiler-Leman algorithm.** An *isomorphism* between undirected node-coloured graphs  $G_1 = (V_1, E_1, c_1)$  and  $G_2 = (V_2, E_2, c_2)$  is any bijection  $f : V_1 \rightarrow V_2$ , satisfying for any  $u, v$ : (i)  $c_1(u) = c_2(f(u))$  and (ii)  $\{u, v\} \in E_1$  if and only if  $\{f(u), f(v)\} \in E_2$ . The *1-dimensional Weisfeiler-Leman algorithm* (1-WL) (Weisfeiler and Leman 1968) is a powerful heuristic for graph isomorphism (Babai and Kucera 1979), which has the same expressive power as MP-GNNs with injective aggregation and combination (Morris et al. 2019; Xu et al. 2019).

Recently 1-WL has been applied to *knowledge graphs*  $KG = (V, E, R, c)$  where  $V$  are nodes,  $E \subseteq R \times V \times V$  are directed edges with labels from  $R$ , and  $c : V \rightarrow D$  colours nodes (Huang et al. 2023; Barceló et al. 2022). An *isomorphism* between knowledge graphs  $KG_1 = (V_1, E_1, R_1, c_1)$  and  $KG_2 = (V_2, E_2, R_2, c_2)$  is any bijection  $f : V_1 \rightarrow V_2$  such that, for all  $u, v \in V_1$  and  $r \in R_1$ : (i)  $c_1(u) = c_2(f(u))$  and (ii)  $(r, u, v) \in E_1$  if and only if  $(r, f(u), f(v)) \in E_2$ . A *relational local 1-WL algorithm* (conventionally  $\text{rwl}_1$ , but we write  $\text{rwl}$ ) is a natural extension of 1-WL to the case of knowledge graphs (Huang et al. 2023). Given a knowledge graph  $KG = (V, E, R, c)$ , the algorithm computes iteratively, for all  $v \in V$  and  $\ell \in \mathbb{N}$ , values  $\text{rwl}^{(\ell)}(v)$  as follows:

$$\text{rwl}^{(0)}(v) = c(v), \\ \text{rwl}^{(\ell)}(v) = \tau\left(\text{rwl}^{(\ell-1)}(v), \left\{\left\{\text{rwl}^{(\ell-1)}(u, r) \mid u \in \mathcal{N}_r(v), r \in R\right\}\right\}\right),$$

where  $\mathcal{N}_r(v) = \{u \mid (r, u, v) \in E\}$  is the  $r$ -neighbourhood of  $v$ , and  $\tau$  is an injective function. It is shown that  $\text{rwl}$  has the same expressive power as R-MPNNs, that is, MP-GNNs processing knowledge graphs (Huang et al. 2023).

## Related Work

There is recently an increasing interest in temporal and dynamic graph neural networks (Longa et al. 2023; Qin and Yeung 2024; Skarding, Gabrys, and Musial 2021; Kazemi et al. 2020). Pertinent models include TGN (Rossi et al. 2020), TGAT (Xu et al. 2020), TDGNN (Qu et al. 2020), and NAT (Luo and Li 2022), which are all based on temporal message-passing mechanisms.

Expressive power results for temporal models are very limited. Souza et al. (2022) compared expressive power of temporal graph neural networks exploiting temporal walks, with those based on local message passing combined with recurrent memory modules. Gao and Ribeiro (2022) compared time-and-graph with time-then-graph models, which are obtained by different combinations of static graph neural networks and recurrent neural networks. In the context of temporal knowledge graphs, expressive power of similar models was recently considered by Chen and Wang (2023).

More mature results have been established for models processing edge-labelled graphs. Such graphs are closely related to temporal graphs, since the aggregated representation of a temporal graph (Gao and Ribeiro 2022), Figure 4 (b), is a multigraph with edges labelled by time points. However, since the aggregated representation does not allow us to assign different colours to the same node in different time points, not all temporal graphs can be directly transformed into multigraphs. Barceló et al. (2022) introduced 1-WL for models processing undirected multi-relational graphs, whereas Beddar-Wiesing et al. (2024) introduced 1-WL for dynamic graphs. Huang et al. (2023) proposed 1-WL for models processing directed multi-relational graphs (i.e. knowledge graphs), namely for relational message passing neural networks (R-MPNNs), which encompass several known models such as RGCN (Schlichtkrull et al. 2018) and CompGCN (Vashishth et al. 2020).

Temporal graphs can be also given in the *event-based* representation (Longa et al. 2023), as a sequence of timestamped events that add/delete edges or modify feature vectors of nodes. Since temporal graphs in the aggregated and event-based representations can be transformed into the snapshot representation (Gao and Ribeiro 2022; Longa et al. 2023), we focus on the snapshot representation in the paper.

### Temporal Weisfeiler-Leman Characterisation

We provide a general approach for establishing expressive power of MP-TGNNs using standard 1-WL. To do so, we transform a temporal graph  $TG$  into a knowledge graph  $KG$  such that MP-TGNNs can distinguish exactly those nodes in  $TG$  whose counterparts in  $KG$  can be distinguished by the standard 1-WL. This contrasts with approaches studying expressive power by modifying 1-WL for particular types of temporal graph neural networks (Souza et al. 2022; Gao and Ribeiro 2022). Note that our results concern distinguishability of nodes, not graphs. Node distinguishability is likely of more practical interest and can be used to distinguish graphs.

We transform  $TG$  into two knowledge graphs:  $\mathcal{K}_{\text{glob}}(TG)$  and  $\mathcal{K}_{\text{loc}}(TG)$ , suitable for analysing, respectively, global and local MP-TGNNs. We first introduce  $\mathcal{K}_{\text{glob}}(TG)$ , whose edges correspond to temporal message-passing in global MP-TGNNs (Figure 5). Intuitively,  $\mathcal{K}_{\text{glob}}(TG)$  contains a separate node  $(v, t)$  for each timestamped node in  $TG$  and an edge between  $(v, t)$  and  $(u, t')$  labelled by  $t - t'$  if  $(u, t')$  is in the temporal neighbourhood of  $(v, t)$ .

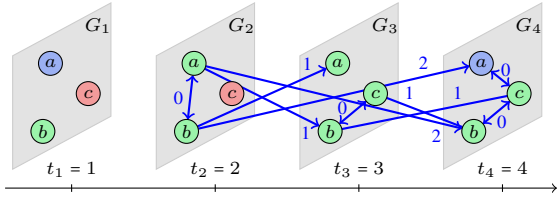


Figure 5:  $\mathcal{K}_{\text{glob}}(TG)$  constructed for  $TG$  from Figure 3

**Definition 1.** Let  $TG = (G_1, t_1), \dots, (G_n, t_n)$  be a temporal graph with  $G_i = (V_i, E_i, c_i)$ . We define a knowledge graph  $\mathcal{K}_{\text{glob}}(TG) = (V, E, R, c)$  with components:

- $V = t\text{-nodes}(TG)$ ,
- $E = \{(t_j - t_i, (v, t_i), (u, t_j)) \mid i \leq j \text{ and } \{u, v\} \in E_i\}$ ,
- $R = \{0, \dots, n - 1\}$ ,
- $c: V \rightarrow R$  satisfies<sup>1</sup>  $c(v, t_i) = c_i(v)$ , for all  $(v, t_i) \in V$ .

We use  $\mathcal{K}_{\text{glob}}(TG)$  to bridge the expressive power of global MP-TGNNs and 1-WL. First we show that global MP-TGNNs cannot distinguish more timestamped nodes over  $TG$  than 1-WL over  $\mathcal{K}_{\text{glob}}(TG)$ .

**Theorem 2.** For any temporal graph  $TG$ , any timestamped nodes  $(v, t)$  and  $(u, t')$  in  $TG$ , and any  $\ell \in \mathbb{N}$ :

- If  $\text{rw}^{(\ell)}(v, t) = \text{rw}^{(\ell)}(u, t')$  in  $\mathcal{K}_{\text{glob}}(TG)$ ,
- then  $\mathbf{h}_v^{(\ell)}(t) = \mathbf{h}_u^{(\ell)}(t')$  in any global TGNN.

<sup>1</sup>for brevity we will drop double brackets, e.g. from  $c((v, t_i))$

*Proof sketch.* By induction on  $\ell$ . The base case holds since  $\text{rw}^{(0)}(v, t_i) = c(v, t_i) = c_i(v) = \mathbf{h}_v^{(0)}(t_i)$ , for each  $(v, t_i)$  in  $TG$ . The inductive step proceeds in a similar way to GNNs (Morris et al. 2019), but additionally exploits the following key property of  $\mathcal{K}_{\text{glob}}(TG)$ :  $(u, t') \in \mathcal{N}_r(v, t)$  in  $\mathcal{K}_{\text{glob}}(TG)$  iff  $r = t - t'$  and  $(u, t') \in \mathcal{N}(v, t)$  in  $TG$ , for all timestamped nodes  $(v, t)$  and  $(u, t')$  in  $TG$ .  $\square$

Moreover, we can show the opposite direction: there is a global MP-TGNN distinguishing exactly the same nodes over  $TG$  as 1-WL over  $\mathcal{K}_{\text{glob}}(TG)$ . By Theorem 2 each global MP-TGNN is not more expressive than 1-WL over  $\mathcal{K}_{\text{glob}}(TG)$ , so for the next theorem it suffices to construct an MP-TGNN at least as expressive as 1-WL over  $\mathcal{K}_{\text{glob}}(TG)$ .

**Theorem 3.** For any temporal graph  $TG$  and any  $L \in \mathbb{N}$ , there exists a global MP-TGNN  $\mathcal{A}$  with  $L$  layers such that for all timestamped nodes  $(v, t)$ ,  $(u, t')$  in  $TG$  and all  $\ell \leq L$  the following are equivalent:

- $\mathbf{h}_v^{(\ell)}(t) = \mathbf{h}_u^{(\ell)}(t')$  in  $\mathcal{A}$ ,
- $\text{rw}^{(\ell)}(v, t) = \text{rw}^{(\ell)}(u, t')$  in  $\mathcal{K}_{\text{glob}}(TG)$ .

*Proof sketch.* The important part of the proof is for the forward implication, as the other implication follows from Theorem 2. We use the result of Huang et al. (2023)[Theorem A.1] showing that for any knowledge graph and in particular  $\mathcal{K}_{\text{glob}}(TG)$ , there is a relational message-passing neural network (R-MPNN) model  $\mathcal{B}$  such that if two nodes  $(v, t)$  and  $(u, t')$  of  $\mathcal{K}_{\text{glob}}(TG)$  have the same embeddings at a layer  $\ell$ , we get  $\text{rw}^{(\ell)}(v, t) = \text{rw}^{(\ell)}(u, t')$ . Huang et al.’s model  $\mathcal{B}$  computes  $\bar{\mathbf{h}}_{(v,t)}^{(\ell)}$  as follows:  $\bar{\mathbf{h}}_{(v,t)}^{(0)} = c(v, t)$  and  $\bar{\mathbf{h}}_{(v,t)}^{(\ell)} = \text{sign}(\mathbf{W}^{(\ell)}(\bar{\mathbf{h}}_{(v,t)}^{(\ell-1)} + \sum_{r \in R} \sum_{(u,t') \in \mathcal{N}_r(v,t)} \alpha_r \bar{\mathbf{h}}_{(u,t')}^{(\ell-1)}) - \mathbf{b}$ ). To finish the proof, we construct a global MP-TGNN such that  $\mathbf{h}_v^{(\ell)}(t)$  computed by  $\mathcal{A}$  on  $TG$  coincide with  $\bar{\mathbf{h}}_{(v,t)}^{(\ell)}$  computed by  $\mathcal{B}$  on  $\mathcal{K}_{\text{glob}}(TG)$ . We obtain it by setting in Equation (2) functions  $\text{AGG}^{(\ell)}$  to the sum and  $\text{COM}^{(\ell)}$  to the sign of a particular linear combination.  $\square$

Next, we show that we can also construct a knowledge graph representing message-passing in local MP-TGNNs. In contrast to  $\mathcal{K}_{\text{glob}}(TG)$ , edges of the new knowledge graph  $\mathcal{K}_{\text{loc}}(TG)$  are bidirectional and hold only between nodes stamped with the same time. Such a knowledge graph is presented in Figure 6 and formally defined below.

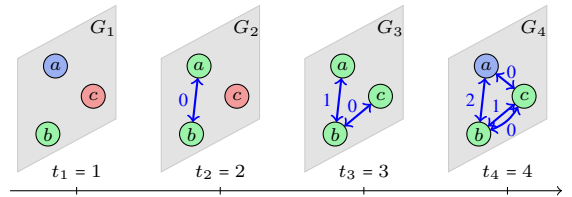


Figure 6: Knowledge graph  $\mathcal{K}_{\text{loc}}(TG)$  for  $TG$  from Figure 3

**Definition 4.** Let  $TG = (G_1, t_1), \dots, (G_n, t_n)$  be a temporal graph with  $G_i = (V_i, E_i, c_i)$ . We define a knowledge graph  $\mathcal{K}_{\text{loc}}(TG) = (V, E, R, c)$  with:

- $V = t\text{-nodes}(TG)$ ,
- $E = \{(t_j - t_i, (v, t_j), (u, t_j)) \mid i \leq j \text{ and } \{u, v\} \in E_i\}$ ,
- $R = \{0, \dots, n-1\}$ ,
- $c: V \rightarrow R$  satisfies  $c(v, t_i) = c_i(v)$ , for all  $(v, t_i) \in V$ .

We can show that local MP-TGNNs can distinguish exactly the same nodes in  $TG$  as  $\text{rwl}$  can distinguish in  $\mathcal{K}_{\text{loc}}(TG)$ , as formally stated in the following two theorems.

**Theorem 5.** For any temporal graph  $TG$ , any timestamped nodes  $(v, t)$  and  $(u, t')$  in  $TG$ , and any  $\ell \in \mathbb{N}$ :

- If  $\text{rwl}^{(\ell)}(v, t) = \text{rwl}^{(\ell)}(u, t')$  in  $\mathcal{K}_{\text{loc}}(TG)$ ,
- then  $\mathbf{h}_v^{(\ell)}(t) = \mathbf{h}_u^{(\ell)}(t')$  in any local MP-TGNN.

**Theorem 6.** For any temporal graph  $TG$  and any  $L \in \mathbb{N}$ , there exists a local MP-TGNN  $\mathcal{A}$  with  $L$  layers such that for all timestamped nodes  $(v, t), (u, t')$  in  $TG$  and all  $\ell \leq L$ , the following are equivalent:

- $\mathbf{h}_v^{(\ell)}(t) = \mathbf{h}_u^{(\ell)}(t')$  in  $\mathcal{A}$ ,
- $\text{rwl}^{(\ell)}(v, t) = \text{rwl}^{(\ell)}(u, t')$  in  $\mathcal{K}_{\text{loc}}(TG)$ .

The Weisfeiler-Leman characterisation of global and local MP-TGNNs established in the above theorems provides us with a versatile tool for analysing expressive power, which we will intensively apply in the following parts of the paper.

## Timewise Isomorphism

While message-passing GNNs (corresponding to 1-WL) provide us with a heuristic for graph isomorphism, their temporal extensions can be seen as heuristics for isomorphism between temporal graphs. However, in the temporal setting it is not clear what notion of isomorphism we should use to obtain an analogous correspondence. We use the characterisation from the previous section to show, quite surprisingly, that both global and local MP-TGNNs can distinguish nodes which are *pointwise isomorphic*—called isomorphic by Beddar-Wiesing et al. (2024). This observation leads us to definition of *timewise isomorphism* as a suitable notion for node indistinguishability in temporal graphs.

Pointwise isomorphism requires that pairs of corresponding snapshots in two temporal graphs are isomorphic. For example  $(a, t_2)$  in  $TG$  and  $(a', t_2)$  in  $TG'$  from Figure 7 are pointwise isomorphic since  $f_1$  with  $f_1(a) = b'$ ,  $f_1(b) = c'$ , and  $f_1(c) = a'$  is an isomorphism between  $G_1$  and  $G'_1$ , and  $f_2$  with  $f_2(a) = a'$ ,  $f_2(b) = b'$ , and  $f_2(c) = c'$  is an isomorphism between  $G_2$  and  $G'_2$ . A formal definition is below.

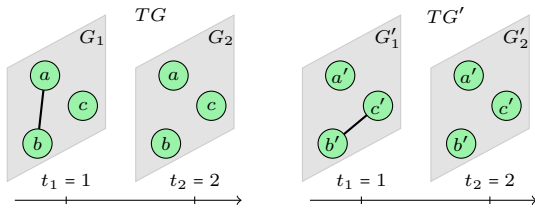


Figure 7: Pointwise isomorphic  $(a, t_2)$  and  $(a', t_2)$

**Definition 7.** Temporal graphs  $TG = (G_1, t_1), \dots, (G_n, t_n)$  and  $TG' = (G'_1, t'_1), \dots, (G'_m, t'_m)$  are pointwise isomorphic if both of the following hold:

- $\text{time}(TG) = \text{time}(TG')$  (so  $n = m$  and  $t_i = t'_i$  for all  $i \in \{1, \dots, n\}$ )
- for every  $i \in \{1, \dots, n\}$  there exists an isomorphism  $f_i$  between  $G_i$  and  $G'_i$ .

If this is the case and  $f_i(v) = u$ , we say that  $(v, t_i)$  and  $(u, t_i)$  are pointwise isomorphic.

It turns out that both global and local MP-TGNNs can distinguish pointwise isomorphic nodes. In particular, they can distinguish  $(a, t_2)$  and  $(a', t_2)$  from Figure 7, as we show below using Theorem 3 and Theorem 6.

**Theorem 8.** There are temporal graphs  $TG$  and  $TG'$  with pointwise isomorphic  $(v, t)$  and  $(u, t')$  such that  $\mathbf{h}_v^{(1)}(t) \neq \mathbf{h}_u^{(1)}(t')$  for some global and local MP-TGNNs.

*Proof sketch.* Consider  $TG$  and  $TG'$  from Figure 7, where  $(a, t_2)$  is pointwise isomorphic to  $(a', t_2)$ . If we apply  $\text{rwl}$  to  $\mathcal{K}_{\text{glob}}(TG)$  and  $\mathcal{K}_{\text{glob}}(TG')$ ,  $\text{rwl}^{(1)}(a, t_2) \neq \text{rwl}^{(1)}(a', t_2)$ , because  $(a, t_2)$  has one incoming edge in  $\mathcal{K}_{\text{glob}}(TG)$ , but  $(a', t_2)$  has no incoming edges in  $\mathcal{K}_{\text{glob}}(TG')$ . The same holds if we apply  $\text{rwl}$  to  $\mathcal{K}_{\text{glob}}(TG')$  and  $\mathcal{K}_{\text{loc}}(TG')$ . So, by Theorem 3 and Theorem 6, there are global and local MP-TGNNs in which  $\mathbf{h}_a^{(1)}(t_2) \neq \mathbf{h}_{a'}^{(1)}(t_2)$ .  $\square$

Theorem 8 shows that pointwise isomorphism is unsuitable for detecting node indistinguishability in MP-TGNNs. We obtain an adequate isomorphism notion by, on the one hand, requiring additionally that all  $f_i$  mentioned in the definition of pointwise isomorphism coincide but, on the other hand, relaxing the requirement  $\text{time}(TG) = \text{time}(TG')$ .

**Definition 9.** Temporal graphs  $TG = (G_1, t_1), \dots, (G_n, t_n)$  and  $TG' = (G'_1, t'_1), \dots, (G'_m, t'_m)$  are timewise isomorphic if both of the following hold:

- $n = m$  and  $t_{i+1} - t_i = t'_{i+1} - t'_i$ , for every  $i \in \{1, \dots, n-1\}$ ,
- there exists a function  $f$  which is an isomorphism between  $G_i$  and  $G'_i$ , for every  $i \in \{1, \dots, n\}$ .

If  $f(v) = u$ , we say that  $(v, t_i)$  and  $(u, t'_i)$  are timewise isomorphic, for any  $t_i \in \text{time}(TG)$ .

Next we show that the timewise isomorphism is an adequate notion for timestamped nodes indistinguishability since timestamped nodes which are timewise isomorphic cannot be distinguished by any (global or local) MP-TGNN.

**Theorem 10.** If  $(v, t)$  and  $(u, t')$  are timewise isomorphic, then  $\mathbf{h}_v^{(\ell)}(t) = \mathbf{h}_u^{(\ell)}(t')$  in any MP-TGNN and any  $\ell \in \mathbb{N}$ .

*Proof sketch.* Assume that  $(v, t)$  from  $TG$  and  $(u, t')$  from  $TG'$  are timewise isomorphic. Hence, by Definition 9,  $TG$  and  $TG'$  are of the forms  $TG = (G_1, t_1), \dots, (G_n, t_n)$  and  $TG' = (G'_1, t'_1), \dots, (G'_n, t'_n)$ , as well as  $t = t_i$  and  $t' = t'_i$  for some  $i \in \{1, \dots, n\}$ . Moreover,  $f(v) = u$  for some  $f: V(TG) \rightarrow V(TG')$  satisfying requirements in Definition 9. We define  $f': t\text{-nodes}(TG) \rightarrow t\text{-nodes}(TG')$  such that  $f'(w, t_j) = (f(w), t'_j)$  for all  $(w, t_j) \in t\text{-nodes}(TG)$ . We can show that  $f'$  is an isomorphism between knowledge graphs  $\mathcal{K}_{\text{glob}}(TG)$  and  $\mathcal{K}_{\text{glob}}(TG')$ , as well as between  $\mathcal{K}_{\text{loc}}(TG)$  and  $\mathcal{K}_{\text{loc}}(TG')$ . Hence, in both cases  $f'(v, t) = (u, t')$  implies  $\text{rwl}^{(\ell)}(v, t) = \text{rwl}^{(\ell)}(u, t')$ , for all

$\ell \in \mathbb{N}$ . Thus, by Theorem 2 and Theorem 5, we obtain that  $\mathbf{h}_v^{(\ell)}(t) = \mathbf{h}_u^{(\ell)}(t')$  for any global and local MP-TGNNs.  $\square$

## Relative Expressiveness of Temporal Message Passing Mechanisms

In this section we will use temporal Weisfeiler-Leman characterisation to prove expressive power results summarised in Figure 2. Our results are on the *discriminative* (also called separating) power, which aims to determine if a given type of models is able to distinguish two timestamped nodes. Formally, we say that a model *distinguishes* a timestamped node  $(v, t)$  in a temporal graph  $TG$  from  $(u, t')$  in  $TG'$  if this model computes different embedding for  $(v, t)$  and  $(u, t')$  at some layer  $\ell$ , that is,  $\mathbf{h}_v^{(\ell)}(t) \neq \mathbf{h}_u^{(\ell)}(t')$ . We say that a type of models (e.g. global or local MP-TGNNs) *can distinguish*  $(v, t)$  from  $(u, t')$ , if some model of this type distinguishes  $(v, t)$  from  $(u, t')$ .

We start by showing that global MP-TGNNs can distinguish timestamped nodes which are indistinguishable by local MP-TGNNs. The reason is that in a global MP-TGNN an embedding of  $(v, t)$  can depend on embeddings at  $t' < t$ , but this cannot happen in a local MP-TGNN.

**Theorem 11.** *There are timestamped nodes that can be distinguished by global, but not by local MP-TGNNs.*

*Proof sketch.* Consider  $(b, t_4)$  from  $TG$  in Figure 3 and  $(b, t_4)$  from  $TG'$  in Figure 4 (a). We can show that, for any  $\ell \in \mathbb{N}$ , application of  $\text{rwl}^{(\ell)}$  to  $\mathcal{K}_{\text{loc}}(TG)$  and  $\mathcal{K}_{\text{loc}}(TG')$  assigns the same labels to these timestamped nodes. Hence, by Theorem 5, local MP-TGNNs cannot distinguish these nodes. On the other hand, for any  $\ell \geq 1$ , application of  $\text{rwl}^{(\ell)}$  to  $\mathcal{K}_{\text{glob}}(TG)$  and  $\mathcal{K}_{\text{glob}}(TG')$  assigns different labels to these nodes. Therefore, by Theorem 3, global MP-TGNNs can distinguish these nodes.  $\square$

Based on the observation from Theorem 11, one could expect that global MP-TGNNs are strictly more expressive than local MP-TGNN. Surprisingly, this is not the case. Indeed, as we show next, there are timestamped nodes which can be distinguished by local, but not by global MP-TGNNs.

**Theorem 12.** *There are timestamped nodes that can be distinguished by local, but not by global MP-TGNNs. This holds true even for colour-persistent temporal graphs.*

*Proof sketch.* Consider  $(a, t_2)$  from  $TG$  and  $(a', t_2)$  from  $TG'$  in Figure 8. Observe that  $(a, t_2)$  in  $\mathcal{K}_{\text{glob}}(TG)$  is isomorphic to  $(a', t_2)$  in  $\mathcal{K}_{\text{glob}}(TG')$  so, by Theorem 2,  $(a, t_2)$  and  $(a', t_2)$  cannot be distinguished by global MP-TGNNs. However,  $(a, t_2)$  has one outgoing path of length 2 in  $\mathcal{K}_{\text{loc}}(TG)$ , but not in  $\mathcal{K}_{\text{loc}}(TG')$ . Hence, two iterations of  $\text{rwl}$  distinguish these nodes. Thus, by Theorem 6,  $(a, t_2)$  and  $(a', t_2)$  can be distinguished by local MP-TGNNs. Note that  $TG$  and  $TG'$  are colour-persistent.  $\square$

Theorem 11 and Theorem 12 show us that neither global or local MP-TGNNs are strictly more expressive, when compared over all temporal graphs. Does the same result hold over colour-persistent graphs? Interestingly, it is not

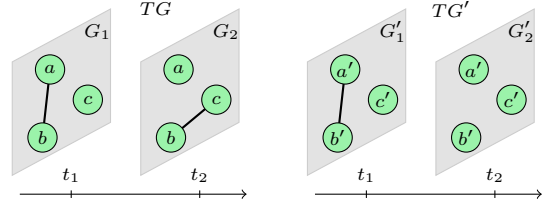


Figure 8:  $(a, t_2)$  and  $(a', t_2)$  which cannot be distinguished by global, but can be distinguished by local MP-TGNNs

the case: in colour-persistent graphs local MP-TGNNs are strictly more expressive than global MP-TGNNs. Hence Theorem 11 cannot hold for colour-persistent graphs.

**Theorem 13.** *In colour-persistent graphs local MP-TGNNs are strictly more expressive than global MP-TGNNs.*

*Proof sketch.* Due to the result established in Theorem 12, it remains to show that over colour-persistent temporal graphs, if  $(v, t)$  and  $(u, t')$  can be distinguished by global MP-TGNNs, then they can be distinguished also by local MP-TGNNs. Hence, by Theorem 2 and Theorem 6, we need to show the  $\text{rwl}_{\text{glob}}^{(\ell)}(v, t) \neq \text{rwl}_{\text{glob}}^{(\ell)}(u, t')$  implies  $\text{rwl}_{\text{loc}}^{(\ell)}(v, t) \neq \text{rwl}_{\text{loc}}^{(\ell)}(u, t')$  for all  $\ell \in \mathbb{N}$ . We show this implication inductively on  $\ell$ , where the inductive step requires proving several non-trivial statements, for example, showing (by another induction) that  $\text{rwl}^{(\ell)}(v, t) \neq \text{rwl}^{(\ell)}(u, t')$  implies  $\text{rwl}^{(\ell)}(v, t+k) \neq \text{rwl}^{(\ell)}(u, t'+k)$ , for any  $k$ .  $\square$

To finish the expressive power landscape announced in Figure 2, it remains to make two more observations. On the one hand, temporal graphs which are not colour-persistent allow global MP-TGNNs to distinguish more elements than colour-persistent graphs. Indeed, this is the case since global MP-TGNNs allow us to pass information about colours between nodes stamped with different time points. On the other hand, this is not allowed in local MP-TGNNs, and so colour-persistence does not impact their expressiveness.

## Experiments

We implement and train basic variants of global and local models on standard temporal link-prediction benchmarks. We emphasise that the goal of our experiments is not to achieve models with high-level performance, but to examine how our expressive power results impact practical performance of MP-TGNNs.

**Benchmarks.** We use the Temporal Graph Benchmark (TGB) 2.0 suite (Gastinger et al. 2024) with small-to-medium temporal datasets `tgbl-wiki`, `tgbl-review`, and `tgbl-coin`, whose statistics are in Table 1. They do not have node features and we discard the edge features. We consider a link-prediction task, where the goal is to predict whether there is a link between two given nodes at the next time point, given information about all previous links. We follow normative training and evaluation procedures supplied by TGB.

**Models.** We implement global and local MP-TGNNs with combination and aggregation functions being concatenation ( $\parallel$ ) and summation ( $\Sigma$ ), respectively, which are among standard choices (Rossi and Ahmed 2015; Xu et al. 2019). Hence, embedding  $\mathbf{h}_v^{(\ell)}(t)$  is computed as

$$W_2^{(\ell)}[\mathbf{h}_v^{(\ell-1)}(t) \parallel \sigma(W_1^{(\ell)}(\sum_{(u,t') \in \mathcal{N}(v,t)} \star \|g(t-t')))],$$

where  $W_1$  and  $W_2$  are learnable,  $\sigma$  is the rectified linear unit,  $\star$  is either  $\mathbf{h}_u^{(\ell-1)}(t')$  (giving rise to a global model) or  $\mathbf{h}_u^{(\ell-1)}(t)$  (local model), and  $g(t-t') = t-t'$ . After  $\ell = 4$  layers, a multi-layer perceptron with 1024 hidden units predicts a link between  $u$  and  $v$ , given  $\mathbf{h}_u^{(\ell)}(t)$  and  $\mathbf{h}_v^{(\ell)}(t)$ . Aggregating the entire temporal neighbourhood incurs over time a linear computational penalty, precluding larger benchmarks; real-world models approximate this calculation.

**Implementation.** Our implementation<sup>2</sup> is based on PyTorch (Paszke et al. 2019), and in particular its hardware-accelerated scatter operations for temporal aggregation. The use of scatter operations means that results may differ between hardware-accelerated runs.

As stated, the global model would require enormous compute and memory in order to use node embeddings from all previous time points. In order to make this tractable, we apply a train-time approximation: during an epoch, embeddings from previous timepoints are ‘‘frozen’’: detached from the computation graph and not updated as model weights change. This interferes with training as the model must use a mixture of stale and fresh embeddings, but at test time the result is exact. A further observation is that only those embeddings whose nodes are connected at some time need be computed and retained due to the definition of  $\mathcal{N}$ .

Minibatching can be achieved in the temporal context by predicting the next  $k$  links given all previous links. Unlike traditional minibatching, this can have a detrimental impact on model accuracy, because earlier links in the batch may help to predict links later in the batch. However, it is computationally very demanding to set  $k = 1$  for even ‘‘small’’ datasets like `tgbl-review` containing millions of links, so a compromise must be found. We set  $k = 32$  for `tgbl-wiki` and  $k = 1024$  for all others.

**Training** We used Adam (Kingma and Ba 2015) for optimisation with PyTorch defaults  $\gamma = 0.001$ ,  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and no L2 penalty. We were able to significantly stabilise and accelerate training by normalising  $g(t-t')$  with respect to elapsed time and by applying batch normalisation (Ioffe and Szegedy 2015) immediately after summation of temporal neighbours. With the exception of the above stability measures, we have not tuned further as we are not aiming for state-of-the-art performance. Training continued until validation loss failed to improve for 10 epochs. Experiments involving `tgbl-wiki` and `tgbl-review` can be run on desktop hardware (NVIDIA GT730), or even without acceleration, whereas `tgbl-coin` requires a large GPU.

<sup>2</sup><https://github.com/MichaelRawson/temporal-graph-nn>

	tgbl-wiki	tgbl-review	tgbl-coin
nodes	9,227	352,637	638,486
edges	157,474	4,873,540	22,809,486
global	0.223	0.321	0.628
local	<b>0.264</b>	<b>0.359</b>	<b>0.635</b>

Table 1: Statistics (nodes and edges) and MRR scores

**Results.** Table 1 shows the mean reciprocal rank (MRR) score (higher is better) used in TGB. We observe that the scores are relatively high given the simplicity of models and lack of tuning. We have written in bold higher among MRRs obtained by global and local MP-TGNNs. In all three datasets, local MP-TGNN obtains higher scores, but the difference between the scores of local and global models is relatively small. We observe that higher performance of the local model aligns with our theoretical result from Theorem 13, which states that over colour-persistent temporal graphs (as here), local MP-TGNNs are strictly more expressive than global MP-TGNNs.

**Layers.** We also investigated the effect of the increasing number of layers  $\ell$  on MRR. We performed experiments on `tgbl-wiki` with the number of layers increasing from 1 to 8. As presented in Figure 9, the highest MRR for the global model is obtained when  $\ell = 5$  and for the local model when  $\ell = 7$ . Interestingly, for  $\ell = 5$  (which is optimal for the global model), MRR for both models is almost the same. This, again, aligns with our theoretical results, showing that local MP-TGNNs are more expressive than global.

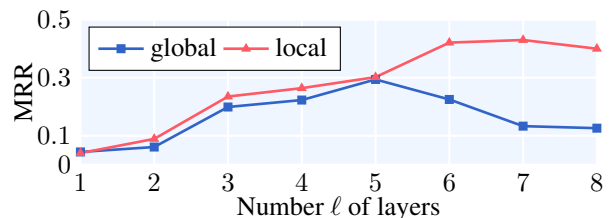


Figure 9: MRR against number of layers on `tgbl-wiki`

**Variations.** The lack of node features in benchmarks led us to try both random node features, which did not significantly alter results, and (transductively) learnable node features, which caused drastic overfitting.

## Conclusions

We have categorised temporal message-passing graph neural networks into global and local, depending on their temporal message passing mechanism. One might expect that global models have higher expressive power than local, but surprisingly we find that the two are incomparable. Further, if node colours (feature vectors) do not change over time, local models are strictly more powerful than global. Our experimental results align with the theoretical findings, showing that local models obtain higher performance on temporal link-prediction tasks.

## Acknowledgements

Michael Rawson was supported by the ERC Consolidator Grant ARTIST 101002685. Przemysław Andrzej Wałęga was partially supported by the EPSRC projects OASIS (EP/S032347/1), ConCuR (EP/V050869/1) and UK FIRES (EP/S019111/1), as well as SIRIUS Centre for Scalable Data Access and Samsung Research UK. For the purpose of Open Access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission.

## References

- Babai, L.; and Kucera, L. 1979. Canonical Labelling of Graphs in Linear Average Time. In *Annual Symposium on Foundations of Computer Science, FOCS*, 39–46.
- Barceló, P.; Galkin, M.; Morris, C.; and Orth, M. A. R. 2022. Weisfeiler and Leman Go Relational. In *Learning on Graphs Conference, LoG*, volume 198, 46.
- Barceló, P.; Kostylev, E. V.; Monet, M.; Pérez, J.; Reutter, J. L.; and Silva, J. P. 2020. The Logical Expressiveness of Graph Neural Networks. In *8th International Conference on Learning Representations, ICLR*.
- Beddar-Wiesing, S.; D’Inverno, G. A.; Graziani, C.; Lachi, V.; Moallemy-Oureh, A.; Scarselli, F.; and Thomas, J. M. 2024. Weisfeiler-Lehman Goes Dynamic: An Analysis of the Expressive Power of Graph Neural Networks For Attributed and Dynamic Graphs. *Neural Networks*, 173: 106213.
- Cai, J.; Fürer, M.; and Immerman, N. 1992. An optimal Lower Bound on The Number of Variables for Graph Identification. *Combinatorica*, 12(4): 389–410.
- Chen, Y.; and Wang, D. 2023. Calibrate and Boost Logical Expressiveness of GNN Over Multi-Relational and Temporal Graphs. In *Annual Conference on Neural Information Processing Systems, NeurIPS*.
- Gao, J.; and Ribeiro, B. 2022. On the Equivalence Between Temporal and Static Equivariant Graph Representations. In *International Conference on Machine Learning, ICML*, volume 162, 7052–7076.
- Gastinger, J.; Huang, S.; Galkin, M.; Loghmani, E.; Parviz, A.; Poursafaei, F.; Danovitch, J.; Rossi, E.; Koutis, I.; Stuckenschmidt, H.; et al. 2024. TGB 2.0: A Benchmark for Learning on Temporal Knowledge Graphs and Heterogeneous Graphs. *arXiv preprint arXiv:2406.09639*.
- Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; and Dahl, G. E. 2017. Neural Message Passing for Quantum Chemistry. In *International Conference on Machine Learning, ICML*, volume 70, 1263–1272.
- Grohe, M. 2023. The Descriptive Complexity of Graph Neural Networks. In *Annual Symposium on Logic in Computer Science, LICS*, 1–14.
- Huang, X.; Romero, M.; Ceylan, İ. İ.; and Barceló, P. 2023. A Theory of Link Prediction via Relational Weisfeiler-Leman on Knowledge Graphs. In *Annual Conference on Neural Information Processing Systems, NeurIPS*.
- Ioffe, S.; and Szegedy, C. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *International Conference on Machine Learning, ICML*, volume 37, 448–456.
- Kapoor, A.; Ben, X.; Liu, L.; Perozzi, B.; Barnes, M.; Blais, M.; and O’Banion, S. 2020. Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks. *CoRR*, abs/2007.03113.
- Kazemi, S. M.; Goel, R.; Jain, K.; Kobayez, I.; Sethi, A.; Forsyth, P.; and Poupart, P. 2020. Representation Learning for Dynamic Graphs: A Survey. *Journal of Machine Learning Research*, 21: 70:1–70:73.
- Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations, ICLR*.
- Longa, A.; Lachi, V.; Santin, G.; Bianchini, M.; Lepri, B.; Lio, P.; Scarselli, F.; and Passerini, A. 2023. Graph Neural Networks for Temporal Graphs: State of the Art, Open Challenges, and Opportunities. *Transactions on Machine Learning Research*.
- Luo, Y.; and Li, P. 2022. Neighborhood-Aware Scalable Temporal Network Representation Learning. In *Learning on Graphs Conference, LoG*.
- Morris, C.; Ritzert, M.; Fey, M.; Hamilton, W. L.; Lenssen, J. E.; Rattan, G.; and Grohe, M. 2019. Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks. In *Conference on Artificial Intelligence, AAAI*, 4602–4609.
- Pareja, A.; Domeniconi, G.; Chen, J.; Ma, T.; Suzumura, T.; Kanezashi, H.; Kaler, T.; Schardl, T. B.; and Leiserson, C. E. 2020. EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs. In *Conference on Artificial Intelligence, AAAI*, 5363–5370.
- Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E. Z.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Annual Conference on Neural Information Processing Systems, NeurIPS*, 8024–8035.
- Qin, M.; and Yeung, D. 2024. Temporal Link Prediction: A Unified Framework, Taxonomy, and Review. *ACM Computing Surveys*, 56(4): 89:1–89:40.
- Qu, L.; Zhu, H.; Duan, Q.; and Shi, Y. 2020. Continuous-Time Link Prediction via Temporal Dependent Graph Neural Network. In *The Web Conference 2020, WWW*, 3026–3032.
- Rossi, E.; Chamberlain, B.; Frasca, F.; Eynard, D.; Monti, F.; and Bronstein, M. M. 2020. Temporal Graph Networks for Deep Learning on Dynamic Graphs. *CoRR*, abs/2006.10637.
- Rossi, R. A.; and Ahmed, N. K. 2015. The Network Data Repository with Interactive Graph Analytics and Visualization. In *Conference on Artificial Intelligence, AAAI*, 4292–4293.

- Schlichtkrull, M. S.; Kipf, T. N.; Bloem, P.; van den Berg, R.; Titov, I.; and Welling, M. 2018. Modeling Relational Data with Graph Convolutional Networks. In *The Semantic Web - 15th International Conference, ESWC*, volume 10843, 593–607.
- Skarding, J.; Gabrys, B.; and Musial, K. 2021. Foundations and Modeling of Dynamic Networks Using Dynamic Graph Neural Networks: A Survey. *IEEE Access*, 9: 79143–79168.
- Souza, A. H.; Mesquita, D.; Kaski, S.; and Garg, V. K. 2022. Provably Expressive Temporal Graph Networks. In *Annual Conference on Neural Information Processing Systems, NeurIPS*.
- Vashishth, S.; Sanyal, S.; Nitin, V.; and Talukdar, P. P. 2020. Composition-based Multi-Relational Graph Convolutional Networks. In *International Conference on Learning Representations, ICLR*.
- Weisfeiler, B.; and Leman, A. 1968. The Reduction of a Graph to Canonical Form and the Algebra Which Appears Therein. *nti, Series*, 2(9): 12–16.
- Wu, S.; Sun, F.; Zhang, W.; Xie, X.; and Cui, B. 2023. Graph Neural Networks in Recommender Systems: A Survey. *ACM Comput. Surv.*, 55(5): 97:1–97:37.
- Xu, D.; Ruan, C.; Körpeoglu, E.; Kumar, S.; and Achan, K. 2020. Inductive Representation Learning on Temporal Graphs. In *International Conference on Learning Representations, ICLR*.
- Xu, K.; Hu, W.; Leskovec, J.; and Jegelka, S. 2019. How Powerful are Graph Neural Networks? In *International Conference on Learning Representations, ICLR*.
- Yu, B.; Yin, H.; and Zhu, Z. 2018. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *International Joint Conference on Artificial Intelligence, IJCAI*, 3634–3640.
- Zhou, J.; Cui, G.; Hu, S.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; and Sun, M. 2020. Graph Neural Networks: A Review of Methods and Applications. *AI Open*, 1: 57–81.