

# Foundations and modelling of dynamic networks using Dynamic Graph Neural Networks: A survey

JOAKIM SKARDING, BOGDAN GABRYS AND KATARZYNA MUSIAL

Corresponding author: Joakim Skarding (email: joakim.skarding@uts.edu.au)

动态网络不只是静态网络的拓展，它具备了很不同的结构特性。文章主要工作有：（1）介绍了dynamic network的基本框架和分类；（2）总结了现有的动态网络模型；（3）介绍了DGNN (Dynamic Graph Neural Network)；（4）DGNN如何去做预测

**Abstract**—Dynamic networks are used in a wide range of fields, including social network analysis, recommender systems and epidemiology. Representing complex networks as structures changing over time allow network models to leverage not only structural but also temporal patterns. However, as dynamic network literature stems from diverse fields and makes use of inconsistent terminology, it is challenging to navigate. Meanwhile, graph neural networks (GNNs) have gained a lot of attention in recent years for their ability to perform well on a range of network science tasks, such as link prediction and node classification. Despite the popularity of graph neural networks and the proven benefits of dynamic network models, there has been little focus on graph neural networks for dynamic networks. We aim to provide a review that demystifies dynamic networks, introduces dynamic graph neural networks (DGNNs) and appeals to researchers with a background in either network science or data science. We contribute: (i) a comprehensive dynamic network taxonomy, (ii) a survey of dynamic graph neural networks and (iii) an overview of how dynamic graph neural networks can be used for dynamic link prediction.

**Index terms** – Dynamic network models, graph neural networks, link prediction, temporal networks.

## I. INTRODUCTION

The bulk of network science literature focuses on static networks, yet every network existing in the real world changes over time. In fact, dynamic network structure has been frequently seen as a complication to be suppressed, to ease progress in the study of networks [1]. Since networks have been used as representations of complex systems in fields as diverse as biology and social science, advances in dynamic network analysis can have a large and far reaching impact on any field using network analytics [2].

Dynamic networks add a new dimension to network modelling and prediction – time. This new dimension radically influences network properties which enable a more powerful representation of network data which in turn increases predictive capabilities of methods using such data [3], [4]. In fact, dynamic networks are not mere generalizations of static networks, they exhibit different structural and algorithmic properties [5].

This work surveys how the dynamic network topology can be modelled using dynamic graph neural networks. We consider a dynamic network to be a network where nodes and edges appear and/or disappear over time. A dynamic graph neural network, is considered to be a neural network

architecture which can encode a dynamic network and where the aggregation of neighbouring node features is part of the neural network architecture. Dynamic graph neural networks (DGNN) often make use of a graph neural network (GNN) and a recurrent neural network (RNN). Graphs where the topology is static and only node or edge features change (spatio-temporal graphs [6]) are out of the scope of this survey and thus so are spatio-temporal graph neural networks [6], [7].

Dynamic networks suffer from a known terminology problem [8]. Complex networks which change over time have been referred to, among others, as; dynamic networks [9], [10], temporal networks [2], [11], evolutionary networks [3] or time-varying networks [12]. With models often working only on specific types of networks, a clear and more detailed terminology for dynamic networks is necessary. We provide a dynamic network foundation with an associated taxonomy to contextualize the models in this survey. We are unaware of any work with a comprehensive taxonomy of dynamic networks and therefore it can be considered as the first major contribution of this paper.

Dynamic networks is a vast and interdisciplinary field. Models of dynamic networks are designed by researchers from different disciplines and they usually use modelling methods from their fields. This survey provides a cross-disciplinary overview of dynamic network models. This overview is not intended to be seen as a dynamic models survey, but rather as a context for dynamic graph neural networks and as a reference point for further exploration of the field of dynamic networks modelling.

There are several surveys on graph neural networks [10], [13], [14] as well as surveys on network representation learning [15], [16]. Kazemi *et al.* [17] is the work most similar to this paper as they survey dynamic network representation learning. The distinction is that they survey the broader topic of representation learning on graphs whereas we survey dynamic graph neural networks which is a subset of representation learning on dynamic networks. We thus survey a more narrow scope than Kazmei *et al.* and a different network type from the GNN surveys which focus on static networks [10], [13], [14]. Wu *et al.* [14] and Zhou *et al.* [10] also survey spatio-temporal graph neural networks, which encode spatio-temporal networks (static networks with dynamic node attributes).

This work follows the encoder-decoder framework used by Hamilton *et al.* [15] and is split into three distinct sections each building upon the previous one.

1) Section II is a discussion on dynamic networks. It serves

as a foundation to the following sections. In this section we explore different definitions of links and introduce a novel dynamic network taxonomy. We also give a brief overview of the dynamic network model landscape, which contextualizes the rest of the survey.

- 2) Section III is a survey of the deep learning models for encoding dynamic network topology. This covers dynamic network encoders.
- 3) Section IV is an overview of how the encoders from section III are used for prediction. This includes dynamic network decoders, loss functions and evaluation metrics.

This survey's contributions are: (i) A conceptual framework and a taxonomy for dynamic networks, (ii) an overview of dynamic network models, (iii) a survey of dynamic graph neural networks (iv) an overview of how dynamic graph neural networks are used for prediction of dynamic networks.

## II. DYNAMIC NETWORKS

A complex network is a representation of a complex system. A network that changes over time can be represented as a dynamic network. A dynamic network have both temporal and structural patterns, and these patterns are described by a dynamic network model.

The definition of a link is essential to any network representation. It is even more essential in dynamic networks, as it dictates when a link appears and disappears. Different link definitions affect network properties which in turn affect which models are capable of representing the dynamic network.

Dynamic network are complex networks that change over time. Links and nodes may appear and disappear. With only this insight we can form a general definition for dynamic networks. Our definition is inspired by Rossetti and Cazabet [18].

**Definition 1 (Dynamic Network)** A Dynamic Network is a graph  $G = (V, E)$  where:  $V = \{(v, t_s, t_e)\}$ , with  $v$  a vertex of the graph and  $t_s, t_e$  are respectively the start and end timestamps for the existence of the vertex (with  $t_s \leq t_e$ ).  $E = \{(u, v, t_s, t_e)\}$ , with  $u, v \in V$  and  $t_s, t_e$  are respectively the start and end timestamps for the existence of the edge (with  $t_s \leq t_e$ ).

This definition and any of the later definitions represent unlabeled and undirected networks, but they can however trivially be extended with both direction and labels taken into account.

Whereas dynamic networks are defined as complex networks where links and nodes may appear and disappear, dynamic network models are often designed to work on specific kinds of dynamic networks and specific dynamic network representations. It therefore makes sense to distinguish between different kinds of dynamic networks and how they are represented.

Table VI is as overview of the notation and abbreviations used in this work.

There are several surveys on dynamic network methods [2], [3], [8], [17]–[23]. These surveys focus either on specific kinds of dynamic networks or on a specific discipline and limit the scope of the survey to models in that discipline. To the best of

our knowledge there is no comprehensive survey of dynamic networks, nor does any dynamic network model survey present a complete foundation or framework for dynamic networks. The aim of this section is to set the stage for the dynamic graph neural network survey by creating a conceptual framework for dynamic networks with more precise terminology and to add context by giving an overview of methods used for modelling dynamic network topology.

### A. Dynamic network representations

Dynamic networks can be represented in different ways and there are advantages and disadvantages inherent to the different representation types.

Dynamic network representations can be grouped into four distinct levels ordered by temporal granularity: (i) static, (ii) edge weighted, (iii) discrete, and (iv) continuous networks [24].



Fig. 1: Network representations ordered by temporal granularity. Static networks are the most coarse grained and continuous representations are the most fine grained. With increasing temporal granularity comes increasing model complexity. The figure is inspired by Fig. 5.1 from Rossetti [24]

Fig. 1 shows those four representations with increasing model complexity as the model becomes more temporally fine-grained:

- **Static networks** have no temporal information.
- **Edge weighted networks** have temporal information included as labels on the edges and/or nodes of a static network. The most straightforward example of this is a static network with the edges labeled with the time they were last active.
- **Discrete networks** are represented in discrete time intervals. These can be represented by multiple snapshots of the network at different time intervals.
- **Continuous networks** have no temporal aggregation applied to them. This representation carries the most information, but is also the most complex.

Static and edge weighted networks are used to model stable patterns or the actual state of the network, whereas discrete and continuous methods are used for more dynamic modelling [18]. This work focuses on dynamic networks and will therefore only cover discrete and continuous representations.

Fine grained representations can be trivially aggregated to produce coarser representations. For example links in a continuous representation can be aggregated into snapshots (or time-windows) which is a discrete representation. Any discrete representation can combine the snapshots, yielding an edge-weighted representation and any edge-weighted representation can discard the weights thus yielding a static network.

1) *Discrete Representation:* Discrete representations use an ordered set of graphs (snapshots) to represent a dynamic graph.

离散型表示使用一系列动态网络的快照来表示网络, 即  $DG = \{G^1, G^2, \dots, G^T\}$ , (1)

这种表示的优势在于可以直接在快照上面应用各种静态网络的分析方法获取信息, 然后使用RNN等处理时间序列的工具进行整合。此外还存在其他有效使用快照的方法, 例如利用滑动时间窗口处理多个快照, 从而减少从一个快照到下一个的变化幅度。

where  $T$  is the number of snapshots. Discrete representations, often simply referred to as "snapshots" is common for dynamic networks [2], [3], [11]. Using a discrete representation of the dynamic network allows for the use of static network analysis methods on each of the snapshots. Repeated use of the static methods on each snapshot can then collectively give insight into the network's dynamics.

There are other approaches which effectively use snapshots as well. Overlapping snapshots such as sliding time-windows [25] are also used in a dynamic network analysis to have less radical change from one network snapshot to the next [26]. Discrete dynamic networks need not be represented as an ordered set of graphs, they may also be represented as a multi-layered network [27] or as a tensor [28].

2) *Continuous Representation*: Continuous network representations are the only representations that have exact temporal information. This makes them the most complex but also the representation with the most potential. We cover three continuous representations: (i) the event-based; (ii) the contact sequence; and (iii) the graph streams. The first two representations are taken from the temporal network literature and they are suitable for networks where links do not persist for long [2], [8], [11]. The third representation, i.e. the graph stream, is used in dynamic networks where edges persist for longer [3]. The focus in these representations is on when edges are active, with no mention of change on nodes. All three representations are described in more detail below:

- 1) **The event-based representation** includes the time-interval at which the edge on a graph is active [11]. An event is synonymous with a link in this case. It is a representation for dynamic networks focusing on link duration. The network is given by a time ordered list of events which include the time at which the event appeared and the duration of the event.

$$EB = \{(u_i, v_i, t_i, \Delta_i); i = 1, 2, \dots\}, \quad (2)$$

where  $u_i$  and  $v_i$  is a node pair on which the  $i$ -th event occurs,  $t_i$  is the timestamp for when the event starts and  $\Delta_i$  is the duration of the event. This is very similar to, and serves the same purpose as, the interval graph [2]. The difference is that the interval graph have the time at which the event ends while the event-based representation have the duration of the event.

- 2) **The contact sequence representation** is a simplification of the event-based representation. In a contact sequence the link is instantaneous and thus no link duration is provided.

$$CS = \{(u_i, v_i, t_i); i = 1, 2, \dots\}, \quad (3)$$

It is common to consider event times in real systems instantaneous if the duration of the event is short or not important [2], [11]. Examples of systems where this representation is suitable include message networks such as text message or email networks.

- 3) **The graph stream representation** is used to represent static graphs which are too large to fit in memory but can

also be used as a representation of a dynamic network [20]. It is similar to the event based representation, however, it treats link appearance and link disappearance as separate events.

$$GS = \{e_1, e_2, \dots\}, \quad (4)$$

where  $e_i = (u_i, v_i, t_i, \delta_i)$ , and  $u_i$  and  $v_i$  is the node pair on which the  $i$ -th event occurs,  $t_i$  is the time at which the event occurs, and  $\delta_i \in \{-1, 1\}$  where  $-1$  represents an edge removal and  $1$  represents that an edge is added. The original representation (used for large graphs) does not include time-stamped information of *when* an edge is added/removed [20]. Time stamps will have to be added for retrieving temporal information.

Since graph streams are mostly used to circumvent hardware limitations rather than a limitation of network representations, we will not survey them in detail here. For a more in depth discussion of the graph streams we refer the interested reader to [3], [20], [22].

Which of the above representations is suitable for the network depends on the link duration with the intricacies of link duration covered in the next section.

### B. Link duration spectrum

Dynamic networks go by many names and sometimes these names indicate specific types of dynamic networks. There is substantial literature on 'temporal networks' [2], [8], [11] which focuses on highly dynamic networks where links may represent events such as human interactions or a single email. On the other hand there is also literature which refers to slowly evolving networks, where links represent persistent relations [3]. To the best of our knowledge there are only two works which take note of this distinction, Rossetti and Cazabet [18], and Holme [8].

Rossetti and Cazabet [18] refer to temporal interaction and relational networks (our temporal and evolving networks respectively), but they do not categorize or make formal distinction between the different networks.

Holme [8] suggests that temporal networks can be distinguished by two requirements: (i) The dynamics on the network being at the same or at a similar time scale as the dynamics of the network; and (ii) The dynamic network is non-trivial at any given time (an instantaneous snapshot yield little to no network structure).

The distinction manifests itself in networks even when not considering dynamics on the networks, and this work is limited to dynamics of the network. Therefore we distinguish temporal networks purely based on network topology. We use the second requirement noted by Holme [8].

This work not only provides a way to distinguish between temporal networks and dynamic networks, but it also proposes a framework in which all networks of dynamic topology fit. We do this by introducing the link duration spectrum.

Fig. 2 shows different types of networks on the link duration spectrum. The scale goes from interactions with no link duration to links that have infinite link duration. No link ever disappears in a network with infinite link duration. Temporal

连续型表示能够记录动态网络完整的时序信息，但这也使得连续型表示在四种不同表示方式中最复杂。



根据网络中链路持续时间的长短可以将动态图进行如下的分类，两个特例：Interaction Networks中边的出现是瞬时的，而Strictly Evolving Networks中边一旦出现则不会消失。

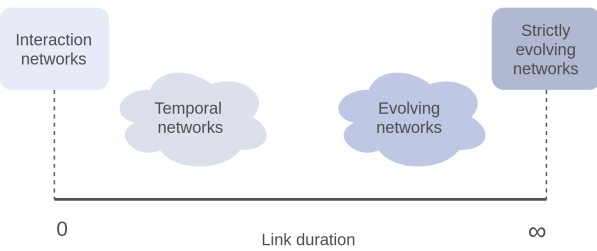


Fig. 2: Temporal and evolving networks on the link duration spectrum. The spectrum goes from 0 (links have no duration) to infinity (links last forever).

边的持续时间是有所不同的：Email发送往往只是暂时的状态，而citation出现了则会永久保留，雇佣关系往往会持续一段时间等。

networks reside on the lower end of the link duration spectrum, whereas evolving networks reside on the higher end. The distinction is as follows:

**Temporal networks.** Highly dynamic networks which are too dynamic to be represented statically. The network is at any given time non-trivial. These networks are studied in the temporal network literature [2], [11]. Network properties such as degree distribution and clustering coefficient cannot be adopted directly from static networks and are non-trivial to define. It is more natural to think of a link as an event with a duration.

**Evolving networks.** Dynamic networks where events persist for long enough to establish a network structure. An instantaneous snapshot yields a well defined network. Network properties such as degree distribution and clustering coefficient can be adopted from static networks and gradually updated. These are the networks most often referred to when the term *dynamic network* is used. Links persist for so long that it is more natural to think of link appearance as an event and link disappearance as another event.

Furthermore, there is one notable special case for each of the dynamic network types. These are types of networks which reside on the extreme ends of the link duration spectrum:

- **Interaction networks.** A type of temporal network where links are instantaneous events. These networks are studied in the **temporal network** literature and often represented as contact sequences [2], [11].
- **Strictly evolving networks.** A type of evolving network where **events have infinite duration**. This implies that the **links never disappear**. We refer to these networks as strictly **evolving networks**.

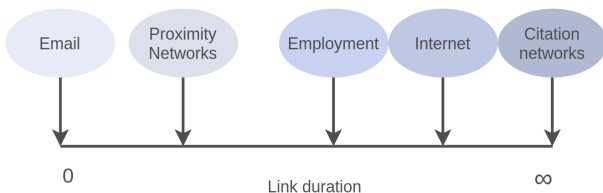


Fig. 3: Examples of networks on the link duration spectrum.

Fig. 3 shows examples of networks on the link duration spectrum.

- An email is a nearly instantaneous event, an email network can therefore be considered an *interaction network*.
- Proximity networks are used as an example of a temporal network in [2]. The link is defined by who is close to whom at what time. Links require maintenance and do not typically last very long.
- Employment networks are social networks where links are formed between employees and employers. The link requires an action after it has been established (termination of contract) to change its state, but also maintenance (continued work from the employee). This network resides in the fuzzy area between temporal and evolving networks and can be treated as either.
- The Internet is an example of the network where we consider nodes linked if data-packets can flow between nodes. A link tends to persist for a long time once established and thus the internet can be thought of as an evolving network.
- Citation networks where links are defined as one paper citing another, have the most persisting links. Once a paper cites another paper, the link lasts forever. This leads to a strictly growing network where no edges disappear. These networks have the additional special characteristic that edges only appear when new nodes appear.

Link definitions influence link duration, which in turn influences a network type. Links can be modified in ways that alter their link duration (also known as time to live, TTL [18]). An email network could define a link as: Actors have once sent an email between each other. This would modify the email link, which is usually nearly instant in duration to a link which will never disappear. This modification moves the network all the way to the right on the spectrum shown in Fig. 2. It transforms an interaction network to a strictly evolving one. Another example of a modification is to use a time-window to force forgetting. A time-window can be applied to a citation network such that only citations which occurred during the time-window appear as links. This will move the network to the left on the link duration spectrum. Depending on the size of the time-window the modified network may be either an evolving or a temporal network.

An additional theoretical special case which is not covered by this concept is a network where links may only disappear. This special case may justify another dimension along which dynamic networks should be distinguished.

### C. Node dynamics 动态网络的另一个区别因素是节点

Another distinguishing factor among dynamic networks is whether nodes may appear or disappear. When modelling networks, it is sometimes simpler to assume that the number of nodes may not change so that the only possible new links are links between already existing nodes.

Many evolving network models assume that edges appear as a new node appears. These models include pseudo-dynamic models such as preferential attachment [29], forest fire [30] and GraphRNN [31]. This is fitting for a citation network where every node is a paper and the edges are cited papers, though in many real world networks, edges can appear and disappear regardless of whether nodes appear.

有时候会假设节点的数量保持不变，用删除所有相关链接表示节点的消失，从而固定邻接矩阵的维度。

TABLE I: Dynamic network types by node dynamics and link duration, excluding special cases.

Node dynamics	Static Dynamic	Link duration	
		Temporal	Evolving
		Node static temporal Node dynamic temporal	Node static evolving Node dynamic evolving

TABLE II: Types of dynamic networks along three dimensions. Static networks and edge-weighted networks are not dynamic networks, but they are included for completeness. If we exclude special cases, we are left with two elements in each dimension.

Dimension	Network types
Temporal granularity	Static, edge-weighted, discrete, continuous
Link duration	Interaction, temporal, evolving, strictly evolving
Node dynamics	Node static, node dynamic, node appearing, node disappearing

With respect to node change we can distinguish between two kinds of networks.

- **Static** where the number of nodes stay static over time; and
- **Dynamic** where the nodes may appear and disappear.

A notable special case of node dynamic networks are the networks where nodes may only appear:

- **Growing** networks are those where nodes may only appear. We consider this a special case of node dynamic networks.

We are unaware of any real world networks where nodes may only disappear. But it should be noted as at least a theoretical special case. Node growing networks on the other hand are rather common.

Any kind of node dynamics can be combined with any kind of link duration network. We can thus have, a *growing evolving network* or a *node static temporal network*. Similarly to the edge duration spectrum a node duration spectrum could theoretically be established, but it has no direct impact on dynamic network structure and we therefore chose to keep node dynamics a discrete distinction.

The node dynamics is an important consideration when modeling the network. Some models support node dynamics whereas others do not.

#### D. The dynamic network cube

Many models assume that nodes disappear when there are no longer any links connected to such nodes. This scheme can work for evolving networks, but in temporal networks it is common that nodes have no links for the majority of the time. Thus for a temporal network it makes sense to model node dynamics separately from link dynamics.

Different aspects of dynamic network representation have been covered in the previous sections. Section II-A defined different dynamic representations ordered by temporal granularity, section II-B defined network types by link duration and section II-C defined network types by node dynamics. This section will consider these previous sections jointly and discuss how the different network types fit together.

Table II includes a comprehensive list of the different dynamic network types. The types are grouped by node dynamic, temporal granularity and link duration type. Types of networks in each group can generally be combined, thus we can have a continuous node static temporal network. The three groups can be thought of as dimensions of a space where different points in the space would represent different types of dynamic networks.

The 3D network type space resulting from excluding special cases is visualised in Fig. 4. When excluding special cases there are two types of networks along each dimension. The dimensions are: (i) link duration (temporal and evolving), (ii) node dynamics (node static and node dynamic) and (iii) temporal granularity (discrete and continuous).

Additionally, Table III presents the suggested terminology for each of the dynamic network types. This completes the suggested dynamic network typology and terminology.

#### E. Dynamic network models

根据网络模型的生成和使用方式，可以分为 Reference Models 和 Realistic Models。前者常被用于研究网络的拓扑结构特征，其特点是会保持某些重要的特征，例如节点度数分布等，而在其他特征上是随机的，以便比较观察到的网络与具有某些相同特征的随机网络之间的区别。后者的目标是尽可能接近现实中的网络，从而进行链接预测、社区检测等任务。

A network model may model a variety of different network characteristics or dynamics. In this work we focus on models of dynamic network structure. Many models define rules for how links are established [29], [30]. The rules are defined such that a network evolved with those rules express some desired features. These features are often observed in real world networks and then included in models as a rule. The search for a good dynamic network model is thus also a search for accurate rules on link formation.

Network models might aim to replicate characteristics like node degree distribution or average shortest path between nodes [32]. The models define probabilistic rules for how links form such that the emerging network has certain distributions of given characteristics observed in real world networks [32]. Some dynamic network models, particularly temporal network models, focus on temporal aspects. An example of a temporal characteristic is the distribution of inter-event times [11].

There are several use cases for network models. They may be used as reference models [2], [8] or as realistic models [33]–[35], and depending on their purpose there are several tasks the model can be used for. These include:

- Reference models are used in the analysis of static networks to study the importance and role of topological features of static networks. Reference models aim to preserve some characteristic such as node degree distribution and otherwise create maximally random networks. The goal is to determine how the observed network is different from a completely random network with the same characteristics. This approach has been adapted to temporal networks [2].
- Realistic models aim to replicate the change in the network as closely as possible. They can be used for several tasks such as network prediction [17], [35], [36] and community detection [18]. Examples include probabilistic models such as the dynamic stochastic block model [37] and representation learning based models such as E-LSTM-D [35]. Some realistic models aim to generate (simulate) realistic networks [31], [38].

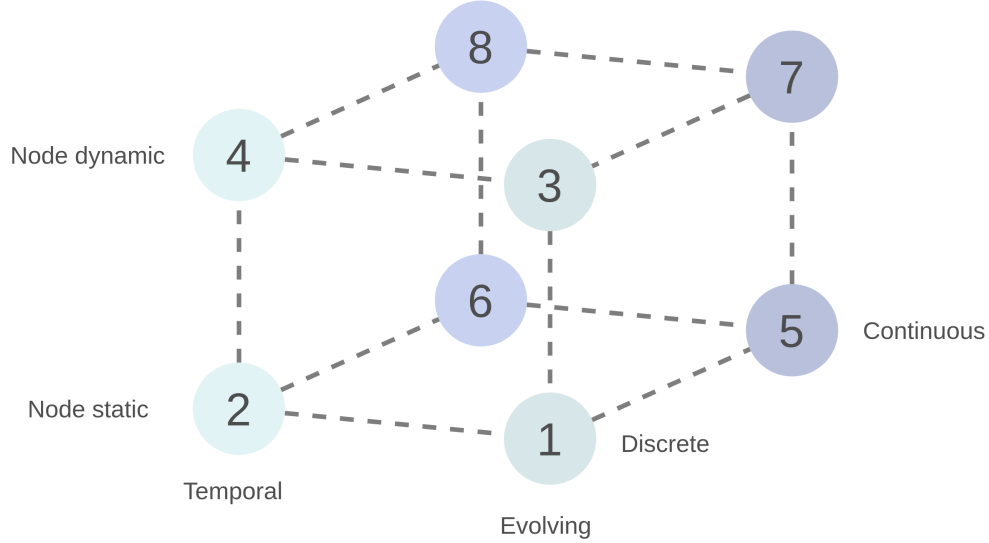


Fig. 4: The dynamic network cube. Each node represents a family of dynamic networks. The nodes are organised along three dimensions: temporal granularity, link duration spectrum and node dynamics.

TABLE III: Terminology of the dynamic network cube.

Node	Temporal granularity	Node dynamics	Link duration	Precise dynamic network term
1	Discrete	Node static	Evolving	Discrete node static evolving network
2			Temporal	Discrete node static temporal network
3		Node dynamic	Evolving	Discrete node dynamic evolving network
4			Temporal	Discrete node dynamic temporal network
5	Continuous	Node static	Evolving	Continuous node static evolving network
6			Temporal	Continuous node static temporal network
7		Node dynamic	Evolving	Continuous node dynamic evolving network
8			Temporal	Continuous node dynamic temporal network

We establish a typology of models for dynamic network topology. The typology is based on the type of method used to model the network (see Fig. 5).

Random graph models (RGM) and Exponential random graph models (ERGM) are random graph models which produce randomly connected graphs while following known common network topology [32]. Stochastic actor oriented models (SAOM) are continuous time models which consider each node an actor and model actor *behaviour*. SAOMs learn to represent the dependencies between a network structure, the position of the actor and the actor behaviour [39]. Relational event models are continuous time models for interaction networks, they define the propensity for a future event to happen between node pairs.

Dynamic network representation learning includes a diverse set of methods which can be used to embed the dynamic graph in a latent space. Representation learning on dynamic networks includes models based on tensor decomposition, random walks and deep learning. Since latent space models and stochastic block models also generate variables in a latent space they are closely related to dynamic network representation learning.

Latent space models and stochastic block models are generative probabilistic models. Latent space models require fitting of parameters with Markov chain Monte Carlo (MCMC) methods and are very flexible but scale to only a few hundred nodes [40]. Stochastic block models on the other hand, scale to an order of magnitude larger networks, at a few

thousand nodes [40]. Tensor decomposition is analogous to matrix factorization where the extra dimension is time [17]. Random walk approaches for dynamic graphs are generally extensions of random walk based embedding methods for static graphs or they apply temporal random walks [11]. Deep learning models include deep learning techniques to generate embeddings of the dynamic network. Deep models can be contrasted with the other networks representation learning models which are shallow models. We distinguish between two types of deep learning models: (i) Temporal restricted Boltzmann machines and (ii) Dynamic graph neural networks. Temporal restricted Boltzmann machines are probabilistic generative models which have been applied to the dynamic link prediction problem [4], [41], [42]. Dynamic graph neural networks combine deep time series encoding with aggregation of neighbouring nodes. Often discrete versions of these models take the form of a combination of a GNN and an RNN. Continuous versions of dynamic graph neural networks cannot make direct use of a GNN since a GNN require a static graph. Continuous DGNNs must therefore modify how node aggregation is done.

A detailed survey of all kinds of dynamic network models is too broad a topic to cover in detail by one survey. Deep learning based models for dynamic networks is a rapidly growing and exciting field, however no existing survey focuses exclusively on dynamic graph neural networks (Kazemi *et al.* [17] being the closest).

张量分解的方法类似于矩阵分解，但是增加了一个时间维度。随机游走模型通常是对静态图上的类似嵌入方法的扩展，有的方法会在游走中考虑时间因素。

两种不同类型的深度学习模型。时间限制玻尔兹曼机是应用于动态链接预测问题的概率生成模型。动态图神经网络通常是GNN和RNN相结合的形式，用GNN处理网络，用RNN处理时序信息。

动态网络表示学习有多种方法，包括张量分解、随机游走、深度学习、潜在空间模型和随机块模型等。

其中潜在空间模型和随机块模型都是生成概率模型。

潜在空间模型需要用马尔科夫链和蒙特卡洛方法拟合参数，该模型非常灵活，但规模只有几百个节点。

随机块模型的规模更大，一般是几千个节点。

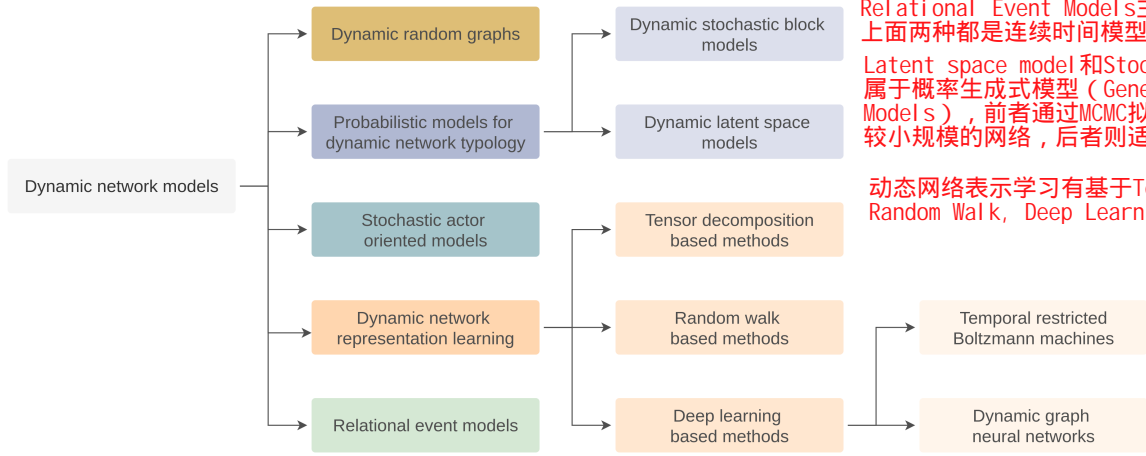


Fig. 5: An overview of dynamic network models. This work explores dynamic graph neural networks in detail.

For the models not discussed in section III there are several works describing and discussing them in detail. Random reference models for temporal networks are surveyed in [2] and [8]. For an introduction to SOAM see Snijders *et al.* [39]. For an introduction to the Relational Event Model (REM) see Butts [43]. See Hanneke *et al.* [44] for Temporal ERGMs (TERGM) on discrete dynamic networks. Block *et al.* [45] provides a comparison of TERGM and SAOM. Fritz *et al.* [21] provide a comparison of a discrete time models, based on the TERGM, and continuous time models represented by the Relational Event Model (REM). Goldenberg *et al.* [46] survey dynamic network models and their survey include dynamic random graph models and probabilistic models. Kim *et al.* [19] surveys latent space models and stochastic block models for dynamic networks. For a comprehensive survey of representation learning on dynamic networks see Kazemi *et al.* [17], and for a survey of dynamic link prediction, including Temporal restricted Boltzmann machines, see Divakaran *et al.* [41].

#### F. Discussion and summary

We have given a comprehensive overview of dynamic networks. This establishes a foundation on which dynamic network models can be defined and thus sets the stage for the survey on dynamic graph neural networks. Establishing this foundation included the introduction of a new taxonomy for dynamic networks and an overview of dynamic network models.

Section II-A presents representations of dynamic networks and distinguishes between discrete and continuous dynamic networks. In section II-B we introduce the link duration spectrum and distinguish between temporal and evolving networks, and in section II-C node dynamics is discussed, we distinguish between node static and node dynamic networks. Section II-D brings together the previous sections to arrive at a comprehensive dynamic network taxonomy.

Discrete representations have seen great success in use on evolving networks with slow dynamics. Graph streams are used on evolving networks which update too frequently to be represented well by snapshots [3]. Both discrete and

continuous representations are used to represent temporal networks [2], [11].

Table IV combines information from section II-A and section II-B and summarizes the existing representations in terms of temporal granularity and link duration.

TABLE IV: Suitable dynamic network representations for temporal and evolving networks.

Representation	Temporal network	Evolving network
Continuous	Event-based representation or Contact sequence	Graph stream
Discrete	Time-windows	Snapshots

Discrete representations have several advantages. A model which works on the static network case can be extended to dynamic networks by applying it on each snapshot and then aggregating the results of the model [17], [19]. This makes it relatively easy, compared to the continuous representation to design dynamic network models. Furthermore, the distinction between an evolving and a temporal network is less important. If modelling a temporal network, one only needs to make sure that a time-window size is large enough that the network structure emerges in each snapshot. However, the discrete representations have their disadvantages too. Chief among them is coarse grained temporal granularity. When modelling a temporal network the use of a time-window is a must. By using a time-window the appearance order of the links and temporal clustering (links appearing frequently together) is lost.

Reducing the size of the time-window or the interval between snapshots is a way to increase temporal granularity. There are however some fundamental problems with this. In the case of a temporal network, a small time-window will eventually yield a snapshot with no network structure. In the case of an evolving network, we will have a sensible network no matter how small the time-window, however there is a trade off with run-time complexity. Discrete models tend to process the entire graph in each snapshot. In which case the run-time will increase linearly with the number of snapshots. The run-time problem is exacerbated by the fact that a lot of real world graphs are huge which make the run-time on each snapshot significant.

random graph models根据已知的结构产生随机连接图；  
Stochastic Actor Oriented Models (SAOM) 将每个节点当做actor考虑并对actor的行动建模；  
Relational Event Models主要用于交互网络，上面两种都是连续时间模型。  
Latent space model和Stochastic block model都属于概率生成式模型 (Generative Probability Models)，前者通过MCMC拟合参数，但是只适用于较小规模的网络，后者则适用大规模网络。

动态网络表示学习有基于Tensor decomposition, Random Walk, Deep Learning的不同方法。

Discrete的优势在于可以将静态网络上使用的模型通过Snapshots的方式拓展到动态网络，同时也不需要严格区分evolving和temporal network；劣势在于对temporal network建模时必须使用时间窗（会导致出现顺序和出现时间聚集相关的信息丢失）。Continuous提供了更细的时间粒度，然而模型更为复杂（目前使用也比较少）。使用连续型建模时必须明确动态网络的类型。



Continuous representations offer superior temporal granularity and thus theoretically a higher potential to model dynamic networks. However, continuous time models tend to be more complex and require either completely new models or significant changes to existing ones to work on the continuous representation. Continuous models are less common than discrete time models [3], [17], [18]. This is likely due to continuous methods being significantly more difficult to develop than discrete methods [3].

When modelling dynamic networks in continuous time it is essential to specify *which* kind of network is being modelled. As models for temporal and evolving networks may not be mutually exclusive and many models work on only specific types of networks. In these cases it might be possible to modify the link duration of a network to run a model on the network. This modification may come at the loss of information, for example when modifying an interaction network to a strictly evolving network, any reappearing link will be removed.

This entire background section establishes a foundation and a conceptual framework in which dynamic networks can be understood. By providing an overview of dynamic network models, it maps out the landscape around deep learning on dynamic graphs thus providing the necessary context. The following sections will explore dynamic graph neural networks in detail.

### III. DYNAMIC GRAPH NEURAL NETWORKS

Network representation learning and Graph Neural Networks (GNN) have seen rapid progress recently and they are becoming increasingly important in complex network analysis. Most of the progress has been done in the context of static networks, with some advances being extended to dynamic networks. Particularly GNNs have been used in a wide variety of disciplines such as chemistry [47], [48], recommender systems [49], [50] and social networks [51], [52].

GNNs are deep neural network architectures that encode graph structures. They do this by aggregating features of neighbouring nodes together. One might think of this node aggregation as similar to the convolution of pixels in convolutional neural networks (CNN). By aggregating features of neighbouring nodes together GNNs can learn to encode both local and global structure.

Several surveys exist of works on static graph representation learning [16], [53] and static graph neural networks [10], [13], [14]. Time-series analysis is relevant for work on dynamic graphs, thus recent advances in this domain is of relevance. For and up to date survey of deep learning on time series we refer to Fawaz *et al.* [54]. Kazemi *et al.* [17] survey representation learning on dynamic graphs and is the closest work to this survey. Their survey has a broader scope than our work as we provide more focus on dynamic graph neural networks.

If dealing with an evolving graph, a static graph algorithm can be used to maintain a model of the graph. Minor changes to the graph would most likely not change the predictions of a static model too much, and the model can then be updated at regular intervals to avoid getting too outdated. We suspect that a spatial GNN is likely to stay accurate for longer than

a spectral GNN, since the spectral graph convolution is based on the graph laplacian which will go through more changes than the local changes in a spatial GNN.

It is important to define what we mean by a dynamic graph neural network (DGNN). Informally we can say that a DGNN is a neural network which encodes a dynamic graph. However, there are some representation learning models for dynamic graphs using deep methods, which we do not consider dynamic graph neural networks. A key characteristic of a graph neural network is an aggregation of neighbouring node features (also known as message passing) [10]. Thus, if a deep representation learning model aggregates neighbouring nodes as part of its neural architecture we call it a dynamic graph neural network. In the discrete case, a DGNN is a combination of a GNN and a time series model. Whereas in the continuous case we have more variety since the node aggregation can no longer be done using traditional GNNs.

Given this definition of representation learning, network models where RNNs are used but network structure is learned using other methods than node aggregation (temporal random walks for example), are not considered DGNNs.

An overview of the types of DGNN encoders is seen in Fig. 6. The encoders aim to capture both structural and temporal patterns and store these patterns in embeddings. This section survey DGNNs, identifies different types of DGNNs and covers how embeddings are encoded. The next section (Section IV) covers decoding of the embeddings.

#### A. Pseudo-dynamic models

Goldenberg *et al.* [46] refer to network models as "pseudo-dynamic" when they contain dynamic processes, but the dynamic properties of the model are not fit to the dynamic data. A well known example of this kind of model is the Barabasi-Albert model [33].

G-GCN [55] can be seen as an extension of the Variational Graph Autoencoder (VGAE) [56] which is able to predict links for nodes with no prior connections, the so called cold start problem. It uses the same encoder and decoder as VGAE, namely a GCN [57] for encoding and the inner product between node embeddings as a decoder. The resulting model learns to predict links of nodes which have only just appeared.

#### B. Discrete Dynamic Graph Neural Networks

Modelling using discrete graphs has the advantage that static graph models can be used on each snapshot of the graph. Discrete DGNNs use a GNN to encode each graph snapshot. We identify two kinds of discrete DGNNs: Stacked DGNNs and Integrated DGNNs.

Autoencoders use either static graph encoders or DGNN encoders, however since they are trained a little differently from DGNNs and generally make use of (and thus extend) a DGNN encoder they are here distinguished from other models.

A discrete DGNN combines some form of deep time-series modelling with a GNN. The time-series model often comes in the form of an RNN, but self-attention has also been used.

对离散表示的动态网络进行建模有一个很大的优点，即可以使用静态的图模型处理每一个快照，然后在使用RNN等模型处理时序信息。离散DGNN可以进一步细分成堆叠型DGNN、整合型DGNN和动态图自动编码器。



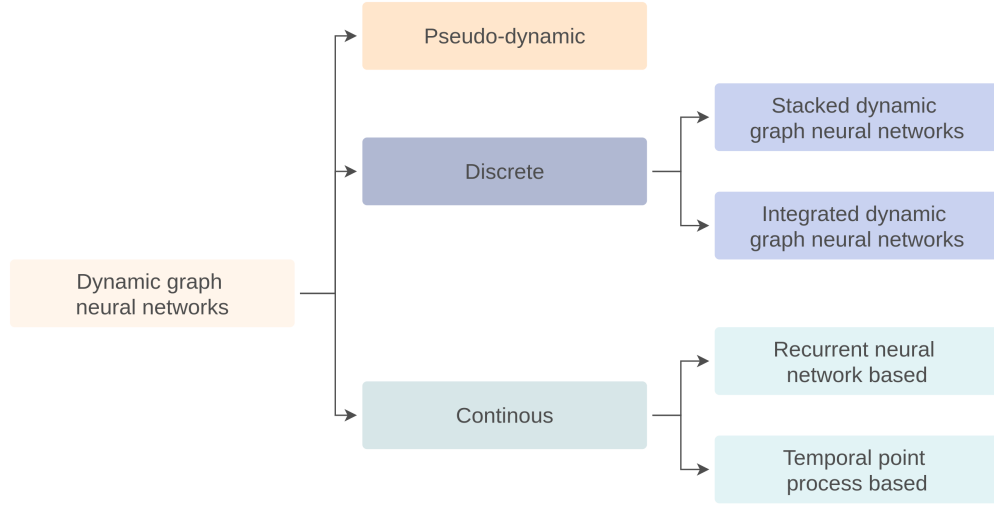


Fig. 6: An overview of dynamic graph neural networks. The main distinction is between discrete and continuous models. This is an extension of Fig 5.

Given a discrete graph  $DG = \{G^1, G^2, \dots, G^T\}$  a discrete DGNN using a function  $f$  for temporal modelling can be expressed as:

$$\begin{aligned} z_1^t, \dots, z_n^t &= \text{GNN}(G^t) \\ h_j^t &= f(h_j^{t-1}, z_j^t) \text{ for } j \in [1, n] \end{aligned} \quad (5)$$

where  $f$  is a neural architecture for temporal modelling (in the methods surveyed  $f$  is almost always an RNN but can also be self-attention),  $z_i^t \in \mathbb{R}^l$  is the vector representation of node  $i$  at time  $t$  produced by the GNN, where  $l$  is the output dimension of the GNN. Similarity  $h_i^t \in \mathbb{R}^k$  is the vector representation produced by  $f$ , where  $k$  is the output dimension of  $f$ .

This can also be written as:

$$\begin{aligned} Z^t &= \text{GNN}(G^t) \\ H^t &= f(H^{t-1}, Z^t) \end{aligned} \quad (6)$$

Informally we can say that the GNN is used to encode each network snapshot and  $f$  (the RNN or self-attention) encodes across the snapshots.

Seo *et al.* [58] introduce two deep learning models which encode a static graph with dynamically changing attributes. Whereas the modelling of this kind of graph is outside the scope of the survey, the two models they introduced are, to the best of our knowledge, the first DGNNs. They introduce both a stacked DGNN and an integrated DGNN: (i) Graph Convolutional Recurrent Network Model 1 (GCRN-M1) and (ii) GCRN model 2 (GCRN-M2) respectively. Very similar encoders have been used in later publications for dynamic graphs.

1) *Stacked Dynamic Graph Neural Networks*: The most straightforward way to model a discrete dynamic graph is to have a separate GNN handle each snapshot of the graph and feed the output of each GNN to a time series component, such as an RNN. We refer to a structure like this as a stacked DGNN.

There are several works using this architecture with different kinds of GNNs and different kinds of RNNs. We'll use GCRN-M1 [58] as an example of a stacked DGNN. This model stacks the spectral GCN from [59] and a standard peephole LSTM [60]:

$$\begin{aligned} z_t &= \text{GNN}(X_t) \\ i &= \sigma(W_i z_t + U_i h_{t-1} + w_i \odot c_{t-1} + b_i) \\ f &= \sigma(W_f z_t + U_f h_{t-1} + w_f \odot c_{t-1} + b_f) \\ c_t &= f_t \odot c_{t-1} \\ &\quad + i_t \odot \tanh(W_c z_t + U_c h_{t-1} + b_c) \\ o &= \sigma(W_o z_t + U_o h_{t-1} + w_o \odot c_t + b_o) \\ h_t &= o \odot \tanh(c_t) \end{aligned} \quad (7)$$

Let  $X_t \in \mathbb{R}^{n \times d}$ ,  $W \in \mathbb{R}^{k \times nl}$ ,  $U \in \mathbb{R}^{k \times k}$  and  $h, w, c, b, i, f, o \in \mathbb{R}^k$ . The gates which are normally vectors in the LSTM are now matrices. Also,  $z_t \in \mathbb{R}^{nl \times 1}$  is a vector and not a matrix. Even though the GNN used in [58] can output features with the same structure as the input, they reshaped the matrix into a vector. This allows them to use a one dimensional LSTM to encode the entire dynamic network.

Whereas [58] use a spectral GCN and a peephole LSTM this is not a limitation of the architecture as any GNN and RNN can be used. Other examples of stacked DGNNs are: RgCNN [61] which use the Spatial GCN, PATCHY-SAN [62] stacked with a standard LSTM and DyGGNN [63] which uses a gated graph neural network (GGNN) [64] combined with a standard LSTM.

Manessi *et al.* [65] present two stacked DGNN encoders: Waterfall Dynamic-GCN (WD-GCN) and Concatenated Dynamic-GCN (CD-GCN). These architectures are distinct in that they use a separate LSTM per node (although the weights across the LSTMs are shared). The GNN in this case is a GCN [57] stacked with an LSTM per node. The WD-GCN encoder with a vertex level decoder is shown in Fig. 7. WD-GCN and CD-GCN differ only in that CD-GCN adds

这是处理离散表示的动态网络的最直接的模型，其基本思路是使用一个GNN模型处理每一个快照，然后利用RNN等时序模型处理GNN的输出，完成图信息在时间轴上的传递和编码。最简单的堆叠型DGNN可以表示为：

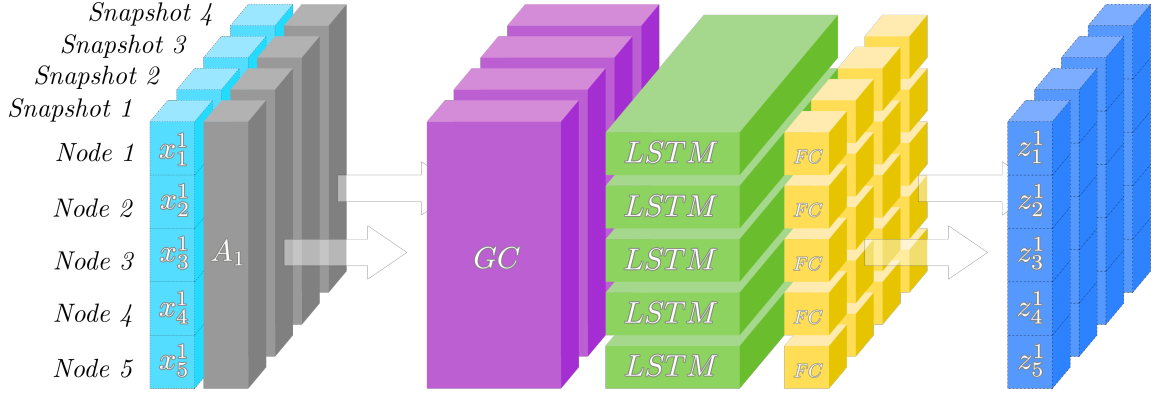


Fig. 7: Stacked DGNN structure from Manessi *et al.* [65]. The graph convolution layer (GC) encode the graph structure in each snapshot while the LSTMs encode temporal patterns.

将每个snapshot输入一个GNN，再将这些GNN的输出拼起来输入时间序列模型（如RNN）。图卷积层（GC）主要用于encode每个snapshot中图的结构，而LSTM用于encode时间特征。可以看到两个不同的组件以堆叠的方式结合成DGNN模型。

skip-connections past the GCN. The equations below are for the WD-GCN encoder.

$$\begin{aligned} Z_1, \dots, Z_t &= \text{GNN}(A_1, X_1), \dots, \text{GNN}(A_t, X_t) \\ H &= \text{v-LSTM}_k(Z_1, \dots, Z_t) \end{aligned} \quad (8)$$

Let  $A \in \mathbb{R}^{n \times n}$  be the adjacency matrix,  $n$  be the number of nodes,  $d$  be the number of features per node and  $X_t \in \mathbb{R}^{n \times d}$  be the matrix describing the features of each node at time  $t$ .  $Z_t \in \mathbb{R}^{n \times l}$  where  $l$  is the output size of the GNN and  $H \in \mathbb{R}^{k \times n \times t}$  where  $k$  is the output size of the LSTMs.

$$\text{v-LSTM}_k(Z_1, \dots, Z_t) = \begin{pmatrix} \text{LSTM}_k(V'_1 Z_1, \dots, V'_1 Z_t) \\ \vdots \\ \text{LSTM}_k(V'_n Z_1, \dots, V'_n Z_t) \end{pmatrix} \quad (9)$$

where LSTM is a normal LSTM [66] and  $V_p \in \mathbb{R}^n$  is defined as  $V_p = \delta_{pi}$  where  $\delta$  is the Kronecker delta. Due to the v-LSTM layer the encoder can store a hidden representation per node.

Sankar *et al.* [67] present a stacked architecture which consists completely of self-attention blocks. They use attention along the spatial and the temporal dimensions. For the spatial dimension they use the Graph Attention Network (GAT) [68] and for the temporal dimension they use a transformer [69].

2) *Integrated Dynamic Graph Neural Networks*: Integrated DGNNs are encoders which combine GNNs and RNNs in one layer and thus combine modeling of the spatial and the temporal domain in that one layer.

Inspired by convLSTM [70] Seo *et al.* [58] introduced GCRN-M2. GCRN-M2 amounts to convLSTM where the convolutions are replaced by graph convolutions. ConvLSTM uses a 3D tensor as input whereas here we are using a two dimensional signal since we have a feature vector for each node.

$$\begin{aligned} f_t &= \sigma(W_f *_{\mathcal{G}} X_t + U_f *_{\mathcal{G}} h_{t-1} + w_f \odot c_{t-1} + b_f) \\ i_t &= \sigma(W_i *_{\mathcal{G}} X_t + U_i *_{\mathcal{G}} h_{t-1} + w_i \odot c_{t-1} + b_i) \\ c_t &= f_t \odot c_{t-1} \\ &\quad + i_t \odot \tanh(W_c *_{\mathcal{G}} X_t + U_c *_{\mathcal{G}} h_{t-1} + b_c) \\ o_t &= \sigma(W_o *_{\mathcal{G}} X_t + U_o *_{\mathcal{G}} h_{t-1} + w_o \odot c_t + b_o) \\ h_t &= o_t \odot \tanh(c_t) \end{aligned} \quad (10)$$

where  $x_t \in \mathbb{R}^{n \times d}$ ,  $n$  is the number of nodes and  $x_i$  is a signal for the  $i$ -th node at time  $t$ .  $W \in \mathbb{R}^{K \times k \times l}$  and  $U \in \mathbb{R}^{K \times k \times k}$  where  $k$  is the size of the hidden layer and  $K$  is the number of Chebyshev coefficients.  $W_f *_{\mathcal{G}} x_t$  denotes the graph convolution on  $x_t$ .

EvolveGCN [71] integrates a RNN into a GCN. The RNN is used to update the weights  $W$  of the GCN. [71] name their layer the Evolving Graph Convolution Unit (EGCU) and present two versions of it: (i) EGCU-H where the weights  $W$  are treated as the hidden layer of the RNN and (ii) EGCU-O where the weights  $W$  are treated as the input and output of the RNN. In both EGCU-H and EGCU-O the RNN operate on matrices rather than vectors as in the standard LSTM. The EGCU-H layer is given by the following equations, where  $(l)$  indicates the neural network layer:

$$\begin{aligned} W_t^{(l)} &= \text{GRU}(H_t^{(l)}, W_{t-1}^{(l)}) \\ H_t^{(l+1)} &= \text{GNN}(A_t, H_t^{(l)}, W_t^{(l)}) \end{aligned} \quad (11)$$

And the EGCU-O layer is given by the equations:

$$\begin{aligned} W_t^{(l)} &= \text{LSTM}(W_{t-1}^{(l)}) \\ H_t^{(l+1)} &= \text{GNN}(A_t, H_t^{(l)}, W_t^{(l)}) \end{aligned} \quad (12)$$

The RNN in both layers can be replaced with any other RNN, and the GCN [57] can be replaced with any GNN given minor modifications.

Other integrated DGNN approaches are similar to GCRN-M2. They may differ in which GNN and/or which RNN they use, the target use case or even the kind of graph they are built for, but the structures of the neural architecture are similar.

这类DGNN模型的基本思路是将GNN和RNN相互整合到一个统一的架构中，比如用图卷积代替RNN中的线性变换部分，或者用RNN控制GNN的参数等

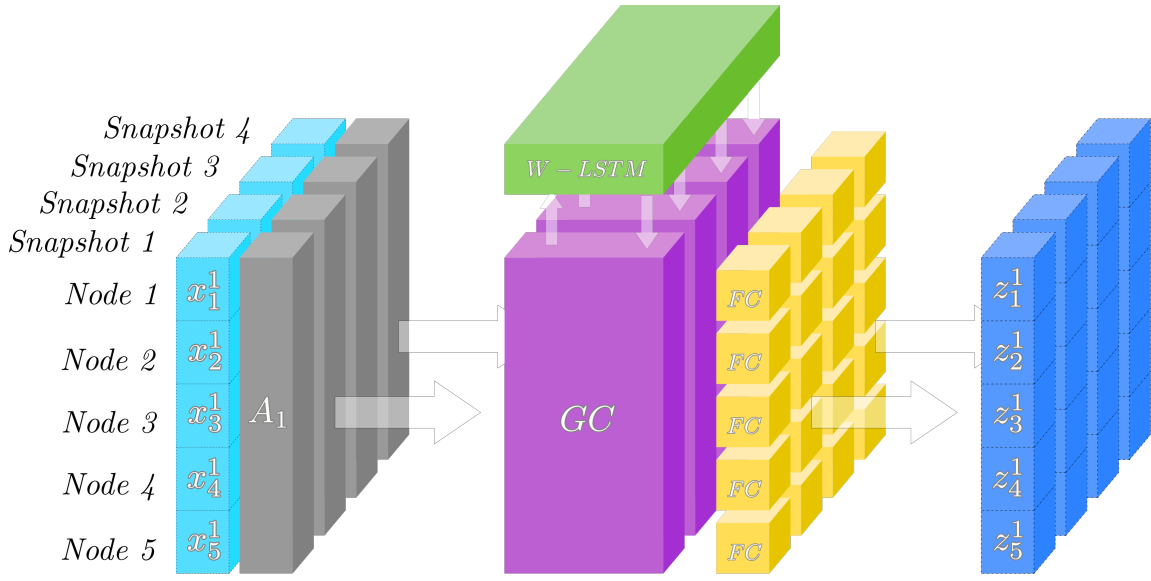


Fig. 8: Integrated DGNN structure of EvolveGCN with a EGCU-O layer [71]. The EGCU-O layer constitute the GC (graph convolution) and the W-LSTM (LSTM for GC weights). W-LSTM is used to initialize the weights of the GC.

将GNN和时间序列模型（RNN等）拼接到一个Layer中构成encoder。GC和LSTM在同一层当中，这里LSTM用于初始化GC的权重。这两个不同的组件整合在同一层当中构成模型的主要部分。

Examples of these include GC-LSTM [72], LRGCN [73] and RE-Net [74].

Chen *et al.* [72] present GC-LSTM, an encoder very similar to GCRN-M2. GC-LSTM takes the adjacency matrix  $A_t$  at a given time as an input to the LSTM and performs a spectral graph convolution [59] on the hidden layer. In contrast GCRN-M2 runs a convolution on both the input and the hidden layer.

LRGCN [73] integrates an R-GCN [75] into an LSTM as a step towards predicting path failure in dynamic graphs.

RE-Net [74] encodes a dynamic knowledge graph by integrating a R-GCN [75] in several RNNs. Other modelling changes enable them to encode dynamic knowledge graphs, thus extending the use of discrete DGNNs to knowledge graphs.

### 3) Dynamic graph autoencoders and generative models:

The Dynamic Graph Embedding model (DynGEM) [76] uses a deep autoencoder to encode snapshots of discrete node dynamic dynamic graphs. Inspired by an autoencoder for static graphs [77] DynGEM makes some modifications to improve computation on dynamic graphs. The main idea is to have the autoencoder initialized with the weights from the previous snapshot. This speeds up computation significantly and makes the embeddings stable (i.e. no major changes from snapshot to snapshot). To handle new nodes the *Net2WiderNet* and *Net2DeeperNet* approaches from [78] are used to add width and depth to the encoder and decoder while the embedding layer stays fixed in size. This allows the autoencoder to expand while approximately preserving the function the neural network is computing.

Dyngraph2vec [79] is a continuation of the work done on DynGEM. dyngraph2vec considers the last  $l$  snapshots in the encoding and can thus be thought of as a sliding time-window. The adjacency matrices  $A^t, \dots, A^{t+l}$  are used to predict  $A^{t+l+1}$ , it is assumed that no new nodes are added. The architecture comes in three variations: (1) dyngraph2vecAE, an autoencoder similar to DynGEM except that it leverages

information from the past to make the future prediction; (2) dyngraph2vecRNN, where the encoder and decoder consist of stacked LSTMs; (3) dyngraph2vecAERN, where the encoder has first a few dense feed-forward layers followed by LSTM layers and the decoder is similar to dyngraph2vecAE, namely a deep feed-forward network.

E-LSTM-D [35] like DynGEM, encode and decode with dense layers, however they run an LSTM on the encoded hidden vector to predict the new embeddings. Although trained as an autoencoder, the model aims to perform a dynamic link prediction.

Hajiramezanali *et al.* [80] introduce two variational autoencoder versions for dynamic graphs: the Variational Graph Recurrent Neural Network (VGRNN) and Semi-implicit VGRNN (SI-VGRNN). They can operate on node dynamic dynamic graphs. Both models use a GCN integrated into an RNN as an encoder (similar to GCRN-M2 [58]) to keep track of the temporal evolution of the graph. VGRNN uses a VGAE [56] on each snapshot that is fed the hidden state of the RGNN  $h_{t-1}$ . This is to help the VGAE take into account how the dynamic graph changed in the past. Each node is represented in the latent space and the decoding is done by taking the inner product decoder of the embeddings [56]. By integrating semi-implicit variational inference [81] with VGRNN they create SI-VGRNN. Both models aim to improve dynamic link prediction.

Generative adversarial networks (GAN) [82] have proven to be very successful in the computer vision field [83]. They have subsequently been adapted for dynamic network generation as well. GCN-GAN [84] and DynGraphGAN [85] are two such models. Both models are aimed towards the dynamic link prediction task. The generator is used to generate an adjacency matrix and the discriminator tries to distinguish between the generated and the real adjacency matrix. The aim is to have the generator that generates realistic adjacency matrices which

GAN在CV领域有很多成功的模型，GCN-GAN和DynGraphGAN将其应用至动态网络中，用于动态链接预测任务。前者使用一个堆叠型DGNN作为生成器，一个神经网络模型作为判别器。或者，生成器更加简单，而判别器由GCN和CNN结合而成。

DynGEM使用深度自编码器对动态图的快照进行编码，并在静态图的自动编码器上进行了一些提高效率的修改。其主要思想是使用前一个快照中的权重初始化自动编码器，从而加快计算速度并使嵌入结果变得稳定。

Dyngraph2vec是在DynGEM的基础上发展得到的，扩展了算法处理快照的窗口，使用快照  $A(1), \dots, A(t+l)$  预测快照  $A(t+l+1)$ 。



can be used as a prediction for the next time step.

GCN-GAN use a stacked DGNN as a generator and a dense feed forward networks as a discriminator [84] and DynGraphGAN use a shallow generator and a GCN [57] stacked with a CNN as a discriminator [85].

目前有两种DGNN方法用于连续建模，其中一种是基于RNN的方法，使用RNN架构维护节点的嵌入模型；另一种是基于时序点（Temporal Point Process, TPP）的方法，通过神经网络处理TPP实现对数据的拟合。目前基于连续表示的DGNN一般只能处理某一类特定的网络而缺乏通用的模型。

### C. Continuous Dynamic Graph Neural Networks

Currently there are two DGNN approaches to continuous modelling. RNN based approaches where node embeddings are maintained by an RNN based architecture and temporal point based approaches where temporal point processes are parameterized by a neural network.

1) **RNN based models**: These models use RNNs to maintain node embeddings in a continuous fashion. A common characteristic for these models is that as soon as an event occurs or there is a change to the network, the embeddings of the interacting nodes are updated. This enables the embeddings to stay up to date continuously. There are two models in this category, Streaming graph neural networks [86] which encode directed strictly evolving networks and JODIE [87] which encodes interaction networks. 这两类模型一旦网络变化就会更新嵌入

这些模型使用RNN维护节点嵌入结果，其共同特征是，一旦事件发生或者网络发生变化，就会对更新嵌入结果进行更新。这类模型可以分为两种：处理流图的SGNN和对交互网络进行编码的JODIE。

The Streaming graph neural network [86] maintains a hidden representation in each node. The architecture consists of two components: (i) an update; and (ii) a propagation component. The update component is responsible for updating the state of the nodes involved in an interaction and the propagation component propagates the update to the involved nodes neighbors.

SGNN会维护每个节点的一个隐藏表示。其架构包括两个组件：一个更新器和一个传播组件，每个组件又可以进一步拆分成3个部分，其中以中间单元最为重要。更新器会在事件发生时更新对应节点的状态，传播组件会在网络中传播信息。模型会为每个节点维护两个向量，分别对应源节点和目标节点两种情况，此外还包含有一个基于这两种状态的隐状态向量。

The update and propagation component consist of 3 units each: (i) the interact unit; (ii) the update / propagate unit; and (iii) the merge unit. The difference between the update component and the propagation component is thus the second unit where the update component makes use of the update unit and the propagate component makes use of the propagate unit.

The model maintains several vectors for each node. Among them are: (i) a hidden state for the source role of the node; and (ii) a hidden state of the target role of the node. This is required to treat source and target nodes differently. The model also contains a hidden state which is based on both the source and target state of the node. The interact unit and merge units can be thought of as wrappers which handle many node states. The interact unit generates an encoding based on the interacting nodes and this can be thought of as an encoding of the interaction. The merge unit updates the combined hidden state of the nodes based on the change done to the source and target hidden states by the middle unit.

更新器的中间单元是一种改进的LSTM模型，能够在间隔不规则的时间序列上面工作。传播组件的中间单元包括注意力函数 $f$ 、衰减函数 $g$ 和过滤器 $h$ 。其中 $f$ 用于计算节点之间的重要性， $g$ 根据时间间隔控制更新的大小，而 $h$ 用于过滤掉已经更新的节点和过于古老的信息。

The middle units and core of the update and propagate components are the update and the propagate units. The update unit generates a new hidden state for the interacting nodes. It is based on a Time-aware LSTM [88], which is a modified LSTM that works on time-series with irregular time intervals. The propagate unit updates the hidden states of the neighboring nodes. It consists of an attention function  $f$ , a time decay function  $g$  and a time based filter  $h$ .  $f$  estimates the importance between nodes,  $g$  gauges the magnitude of the update based on how long ago it was and  $h$  is a binary function which filters out updates when the receiving node has too old information.

$h$  has the effect of removing noise as well as making the computation more efficient.

By first running the update component and afterwards propagating, information of the edge update is added to the hidden states of the local neighborhood.

The second method is JODIE [87]. JODIE embeds nodes in an interaction network. It is however targeted towards recommender systems and built for user-item interaction networks. The intuition is that with minor modifications this model can work on general interaction networks.

JODIE uses an RNN architecture to maintain the embeddings of each node. With one RNN for users ( $RNN_u$ ) and one RNN for items ( $RNN_i$ ), the formula for each RNN is identical except that they use different weights. When an interaction happens between a user and an item, each of the embeddings is updated according to equation 13.

JODIE是建立在顾客 $u$ 和商品 $i$ 的交互图之上的。这一模型使用两个RNN分别维护顾客节点和商品节点的嵌入。当用户和商品发生一次交互时，按照下公式进行更新：

$$\begin{aligned} u(t) &= \sigma(W_1^u u(\bar{t}) + W_2^u i(\bar{t}) + W_3^u f + W_4^u \Delta_u) \\ i(t) &= \sigma(W_1^i i(\bar{t}) + W_2^i u(\bar{t}) + W_3^i f + W_4^i \Delta_i) \end{aligned} \quad (13)$$

where  $u(t)$  is the embedding of the interacting user,  $i(t)$  the embedding of the interacting item,  $u(\bar{t})$  the embedding of the user just before the interaction and similarly  $i(\bar{t})$  is the embedding of the item just before the interaction. The superscript on the weights indicates which RNN they are parameters of, so  $W_1^u$  is a parameter of  $RNN_u$ .  $f$  is the feature vector of the interaction and  $\Delta_u$  is the time since the user interacted with an item and similarly for  $\Delta_i$ .

An additional functionality of JODIE is the projection component of their architecture. It is used to predict the trajectory of the dynamic embeddings. The model predicts the future position of the user or item embedding and is trained to improve this prediction.

时间点过程

2) **Temporal point process based models**: Know-Evolve [89] is the precursor to the rest of the dynamic graph temporal point process models discussed in this section. It models knowledge graphs in the form of interaction networks by parameterizing a temporal point process (TPP) by a modified RNN. With some minor modifications the model should be applicable to any interaction network, but since the original model is specifically for knowledge graphs we will rather focus on its successor, DyREP [36].

这类模型来自于Know-Evolve模型。该模型通过RNN参数化一个TPP，用于为知识图谱谱进行建模。其后继者是DyREP模型。在DyREP中，TPP被用于表示网络的拓扑演化和节点通信两种动态过程。

DyREP uses a temporal point process model which is parameterised by a recurrent architecture [36]. The temporal point process can express both dynamics "of the network" (topological evolution) and "on the network" (node communication). By modelling this co-evolution of both dynamics they achieve a richer representation than most embeddings.

The temporal point process (TPP) is modelled by events  $(u, v, t, k)$  where  $u$  and  $v$  are the interacting nodes,  $t$  is the time of the event and  $k \in \{0, 1\}$  indicates whether the event is a topological evolution,  $k = 0$  (edge added) or a communication  $k = 1$ .

TPP既能表达topological evolution，也能表示node communication，因此可获得维度更丰富的representation

The conditional intensity function  $\lambda$  describes the probability of an event happening.  $\lambda$  is parameterised by two functions  $f$  and  $g$ .

通信意味着节点内容属性变化

$$\lambda_k^{u,v} f_k(g_k^{u,v}(t)) \quad (14)$$

where  $\bar{t}$  is the time just before the current event,  $g$  is a weighted concatenation of node embeddings  $z$ ,  $g_k^{u,v}(\bar{t}) = \omega_k^T \cdot [z^u(\bar{t}); z^v(\bar{t})]$ .  $f$  is a modified softplus,  $f_k(x) = \psi_k \log(1 + \exp(x/\psi_k))$ ,  $\omega_k$  and  $\psi_k$  are four parameters which enable the temporal point process to be modelled on two different time scales.

The TPP is parameterised by an RNN. The RNN incorporates aggregation of local node embeddings, the previous embedding of the given node and an exogenous drive.

$$z^v(t_p) = \sigma(W^{struct} h_{struct}^u(\bar{t}_p) + W^{rec} z^v(\bar{t}_p) W^t(t_p - \bar{t}_p)) \quad (15)$$

where  $h_{struct}^u$  is given by an attention mechanism which aggregates embeddings of neighbors of  $u$ . The attention mechanism uses an attention matrix  $S$  which is calculated and maintained by the adjacency matrix  $A$  and the intensity function  $\lambda$ . In short, the  $\lambda$  parameterises the attention mechanism used by the RNN which in turn is used to parameterise  $\lambda$ . Thus  $\lambda$  influences the parameterisation of itself.

With  $\lambda$  well parameterised it serves as a model for the dynamic network and its conditional intensity function can be used to predict link appearance and time of link appearance.

Latent dynamic graph (LDG) [90] uses Kipf *et al.*'s Neural Relational Inference (NRI) model [91] to extend DyREP. The idea is to re-purpose NRI to encode the interactions on the graph, generate a temporal attention matrix which is then used to improve upon self-attention originally used in DyREP.

Graph Hawkes network (GHN) [92] is another method that parameterize a TPP through a deep neural architecture. Similarly to Know-Evolve [89], it targets temporal knowledge networks. A part of the architecture, the Graph Hawkes Process, is an adapted continuous-time LSTM for Hawkes processes [93].

#### D. Discussion and summary

Deep learning on dynamic graphs is still a new field, however, there are already promising methods which show the capacity to encode dynamic topology. This section has provided a comprehensive and detailed survey of deep learning models for dynamic graph topology encoding.

The encoders are summarised and compared in Table V. Models are listed based on their encoders and the encoders capacity to model link and node dynamics. Any model which cannot model link deletion can only model strictly evolving networks or interaction networks (see section II-B).

Most methods focus on discrete graphs which enable them to leverage recent advances in graph neural networks. This allows for modelling of diverse graphs, including node dynamic graphs, dynamic labels on nodes and due to the use of snapshots, temporal networks can also be handled. Continuous models currently exist for strictly growing networks and specific types of interaction networks. This leaves large classes of dynamic graphs unexplored. Since continuous models have some inherent advantages over discrete graphs (see section II-F), expanding the repertoire of dynamic network classes, which can be modelled by continuous networks, is a promising future direction.

All discrete DGNNs use a GNN to model graph topology and a deep time-series model, typically an RNN, to model the time dependency. Two types of architectures can be distinguished: (i) the stacked DGNN and (ii) the integrated DGNN. Given the same graph, a stacked DGNN would generally have fewer parameters than a typical integrated DGNN (such as GCRN-M2 [58]). Both approaches offer great flexibility in terms of which GNN and RNN can be used. They also are rather flexible in that they can model networks with both appearing and disappearing edges as well as dynamic labels.

Discrete models tend to treat every snapshot as a static graph, thus the complexity of the model is proportional to the size of the graph in each snapshot and the number of snapshots. Whereas a continuous model complexity is generally proportional to the number of changes in the graph. This means that continuous models, where the complexity depends on the number of events, are more computationally efficient. If a discrete approach creates snapshots using time-windows, then it can trade off temporal granularity (and thus theoretically modelling accuracy) for faster computation by using larger time-windows for each snapshot.

**So which one should you chose?** A discrete approach is more straightforward and simpler to test than continuous approaches. If temporal granularity and performance is not a concern then one of the advanced discrete approaches such as DySAT or EvolveGCN will likely be a great fit for most dynamic network problems. Since they naturally support link deletion, node addition and node deletion, they provide good general purpose functionality.

Evolving networks are well served by any discrete approach, however with the recent dominance of attention architectures [69] we would expect DySAT to do well in a comparative test. EvolveGCN is expected to train fast on an evolving network with little change between snapshots. The discrete methods are also suited for temporal networks given that the length of the time-windows covered by snapshots is well selected.

If node dynamics is an important feature of the network you wish to model, then you should choose a model that can encode node dynamics such as DySAT [67] or EvolveGCN [71].

As of now, continuous approaches have a lot of potential, but need further development to become a good alternative to the discrete approaches. The continuous approaches are currently limited to network types without link deletion. If run-time complexity and time granularity is essential to the dynamic complex network at hand (for example in case of a temporal network), then non-deep learning methods which are not covered by this survey are recommended. Those methods can be explored in the literature referred to in section II-E.

## IV. DEEP LEARNING FOR PREDICTION OF NETWORK TOPOLOGY

Any embedding method can be thought of as a concatenation of an encoder and a decoder [15]. Until now, we have discussed encoders, but the quality of embeddings depend on the decoder and the loss function as well. While the encoders in Section III can be paired with a variety of decoders and

这一节的重点是如何去decode，主要讨论link prediction的问题，一般有两种情况：time-prediction和time-conditioned，前者预测某件事什么时候发生，后者预测某件事是否发生。

Encoders及其对动态边和节点建模的能力 TABLE V: Deep encoders for dynamic network typology.

Model type	Model name	Encoder	Link addition	Link deletion	Node addition	Node deletion	Network type
Discrete networks							
Stacked DGNN	GCRN-M1 [58]	Spectral GCN [59] & LSTM	Yes	Yes	No	No	Any
	WD-GCN [65]	Spectral GCN [57] & LSTM	Yes	Yes	No	No	Any
	CD-GCN [65]	Spectral GCN [57] & LSTM	Yes	Yes	No	No	Any
	RgCNN [61]	Spatial GCN [62] & LSTM	Yes	Yes	No	No	Any
	DyGGNN [63]	GGNN [64] & LSTM	Yes	Yes	No	No	Any
	DySAT [67]	GAT [68] & temporal attention from [69]	Yes	Yes	Yes	Yes	Any
	GCRN-M2 [58]	GCN [59] integrated in an LSTM	Yes	Yes	No	No	Any
	GC-LSTM [72]	GCN [59] integrated in an LSTM	Yes	Yes	No	No	Any
	EvolveGCN [71]	LSTM integrated in a GCN [57]	Yes	Yes	Yes	Yes	Any
	LRGCN [73]	R-GCN [75] integrated in an LSTM	Yes	Yes	No	No	Any
Continuous networks	RE-Net [74]	R-GCN [75] integrated in several RNNs	Yes	Yes	No	No	Knowledge network
	RNN based						
	Streaming GNN [86]	Node embeddings maintained by architecture consisting of T-LSTM [88]	Yes	No	Yes	No	Directed strictly evolving
	JODIE [87]	Node embeddings maintained by an RNN based architecture	Yes	No	No	No	Bipartite and interaction
	TTP based						
	Know-Evolve [89]	TPP parameterised by an RNN	Yes	No	No	No	Interaction, knowledge network
	DyREP [36]	TPP parameterised by an RNN aided by structural attention	Yes	No	Yes	No	Strictly evolving
	LDG [90]	TPP, RNN and self-attention	Yes	No	Yes	No	Strictly evolving
	GHN [92]	TPP parameterised by a continuous time LSTM [93]	Yes	No	No	No	Interaction, knowledge network

离散的更加直观简洁，不过前面的章节提到过，相比离散型，连续型的模型有内在的优势。然而连续型只能应对没有Link deletion的情况，后续还需要继续改进。

loss functions depending on the intended task, we focus in this section on one of the most commonly tackled problems - link prediction.

Prediction problems can be defined for many different contexts and settings. In this survey we refer to the prediction of the future change to the network topology. Much work has been done on the prediction of missing links in networks, which can be thought of as an interpolation task. This section explores how dynamic graph neural networks can be used for link prediction and deal exclusively with the extrapolation (future link prediction) tasks.

Predictions can be done in a time-conditioned or a time-predicting manner. Time-predicting means that a method predicts *when* an event will occur and time-conditioned means that a method predicts *whether* an event will occur at a given time  $t$ . For example, if the method predicts the existence of a link in the next snapshot, it is a time-conditioned prediction. If it predicts when a new link between nodes will appear, it is a time-predicting prediction.

Prediction of links often focuses only on the prediction of the appearance of a link. However, link disappearance is a less explored but also important for the prediction of network topology. We refer to link prediction based on a dynamic network as dynamic link prediction.

For embedding methods, what is predicted and how, is decided by the decoder. You can have both time-predicting and time-conditioned decoders. The prediction capabilities will depend on the information captured by the embeddings. Thus, an embedding which captures continuous time information has a higher potential to model temporal patterns. Well modelled temporal and structural embeddings offer a better foundation for a decoder and thus potentially better predictions.

If dealing with discrete data and few timestamps, a time-conditioned decoder can be used for time prediction. This can be done by applying the time-conditioned decoder to every candidate timestamp  $t$  and then consider the  $t$  where the link has the highest probability of appearing.

The rest of this section is a description of how the surveyed models from the previous section can be used to perform

predictions. This includes mainly a discussion on decoders and loss functions. Since the surveyed models aim to predict the time-conditioned existence of links, the focus will be on the dynamic link prediction task.

Autoencoders can use the same decoders and loss functions as other methods. Their aim is typically a little different. The decoder is targeted at the already observed network and tries to recreate the snapshot. A prediction for a snapshot at time  $t+1$  is marginally different from the decoder of an autoencoder which is targeted at already observed snapshots.

预测“什么时候发生”的decoder基本都需要一个预测“是否发生”的decoder，decode的过程就是从embedding layer输出邻接矩阵。因此，预测问题可以转化为求“时间 $t+1$ 时边出现的概率”（通过输出的邻接矩阵求得）。常见的方式有内积，GraphRNN等。

#### A. Decoders

Of the surveyed approaches which apply a predicting decoder, almost all apply a time-conditioned decoder. A prediction is then often an adjacency matrix  $\hat{A}^\tau$  which indicates the probabilities of an edge at time  $\tau$ . Often  $\tau = t + 1$ .

We consider decoders to be the part of the architecture which produces  $\hat{A}^\tau$  from  $H$  the dynamic graph embeddings.

Dimensions of  $H$  depend on the approach. For discrete graphs we may have  $H \in \mathbb{R}^{n \times s \times d_h}$ , where  $n$  and  $s$  is the number of nodes and snapshots respectively, and  $d_h$  determines the size of the embedding layer  $H$  [63], [72]. In this case  $H$  will need to be reshaped to produce  $\hat{A}^\tau$ .  $H$  may take many different forms, however, it is rather common to apply a feed-forward network to obtain a node embedding layer  $Z \in \mathbb{R}^{n \times d_z}$  [63], [67], [71], [72]. This node embedding layer can be thought of as the embeddings for the nodes in the prediction.

Some approaches decode only from the last snapshot in which case  $H \in \mathbb{R}^{n \times d_h}$  and we can simply consider  $H = Z$ . Thus no feed-forward network is required to reshape the embeddings.

From  $Z$  there are different approaches of obtaining  $\hat{A}^\tau$ . A straightforward way is to further use a feed-forward network [35], [71], [76], [79]. Another approach introduced by Kipf and Welling [56] for a static graph autoencoder is to use the inner product of the node embeddings in  $Z$ :

$$p(A_{ij}^t = 1 | z_i^t, z_j^t) = \sigma((z_i^t)^\top z_j^t) \quad (16)$$



Where  $z_k$  is the node embedding of node  $k$ . An inner product decoder works well if we only want to predict or reproduce the graph topology. If we would like to decode the feature matrix then a neural network should be used [80].

Wu *et al.* [94] uses GraphRNN, a deep sequential generative model as a decoder [31]. What is unique with GraphRNN is that it reframes the graph generation problem as a sequential problem. The GraphRNN authors claim increased performance over feed-forward auto-encoders.

In general there are many options for how decoding can be done. As long as the probability for each edge is produced from the latent variables and the architecture can be efficiently optimized with back-propagation.

The only surveyed method using a time-predicting decoder is DyRep [36]. DyRep uses the conditional intensity function of its temporal point process to model the dynamic network.

While the focus in this section is on decoders that are used directly for the forecasting task, it is important to note that downstream learning is also used. This is where the embeddings which capture temporal and topological patterns are produced by the DGNN, and later used to perform additional learning, for example forecasting. An example of this is seen in [67], where a logistic regression classifier is trained on the node embeddings of snapshot  $t$  to predict links at  $t + 1$ .

### B. Loss functions

The loss function is central to any deep learning method, as it is the equation which is being optimized. Regarding loss functions we can make a distinction between (i) link prediction optimizing methods; and (ii) autoencoder methods. As the prediction methods optimize towards link prediction directly, an autoencoder optimizes towards recreation of the dynamic graph and also can be used and have been shown to perform well in link prediction tasks.

1) *Link prediction*: Prediction of edges is seen as a binary classification task. Traditional link prediction is well known for being extremely unbalanced [40], [95]. For predicting methods the loss function is often simply the binary cross-entropy [63], [67], [71].

Some models use negative sampling [67], [71]. This transforms the problem of link prediction from a multiple output classification (a prediction for each link) to a binary classification problem (is the link a "good" link or a "bad" link). This speeds up computation and deals with the well-known class imbalance problem in link prediction.

$$\mathcal{L}_{CE} = \sum_{i=1}^n \sum_{j=1}^n A_{ij}^t \log(\hat{A}_{ij}^t) \quad (17)$$

Equation 17 is an example of a binary cross entropy loss adapted from [72]. An alternative approach to computing the loss is seen in [72], which uses a regression loss function for link prediction which they regularize by including the squared Forbenius norms of all the weights of the model in the loss.

DySAT [67] sums the loss function only over nodes which are in the same neighbourhood at time  $t$ . The neighbourhoods are extracted by taking nodes which co-occur in a random walks on the graph. The inner product is calculated as a part

of the summation in the loss function. This means that the inner product will be calculated only for the node pairs that the loss is computed on. Together it reduces the amount of nodes which are summed up and should result in a training speed up. Any accuracy trade-off is not discussed by the authors.

2) *Autoencoders*: Autoencoder approaches [35], [76], [79] aim to reconstruct the dynamic network. All surveyed autoencoders operate on discrete networks. Therefore the reconstruction of the network is reduced to reconstruction of each snapshot. This entails creating a loss function which penalizes wrong reconstruction of the input graph. Variational autoencoder approaches [55], [80] also aim to be generative models. To be generative, they need to enable interpolation in latent space. This is achieved by adding a term to the loss function which penalizes the learned latent variable distribution for being different from a normal distribution. It is also common to add regularization to the loss functions to avoid overfitting.

$$\mathcal{L} = \sum_{i=1}^n \sum_{j=1}^n (A_{ij}^t - \hat{A}_{ij}^t) * P_{ij} \quad (18)$$

Equation 18 is the reconstruction penalizing component of E-LSTM-D's loss function [35].  $P$  is a matrix which increases the focus on existing links.  $p_{ij} = 1$  if  $A_{ij}^t = 0$  and  $p_{ij} = \beta > 1$  if  $A_{ij}^t = 1$ .

A common way to regularize is through summing up all the weights of the model, thus keeping the weights small and the model less likely to overfit. E-LSTM-D [35] used the  $L_2$  norm for this, as seen in equation 19.

$$\begin{aligned} \mathcal{L}_{reg} = & \frac{1}{2} \sum_{k=1}^K (\|W_e^{(k)}\|_F^2 + \|\hat{W}_d^{(k)}\|_F^2) \\ & + \frac{1}{2} \sum_{l=1}^L (\|W_f^{(l)}\|_F^2 + \|W_i^{(l)}\|_F^2 \\ & + \|W_C^{(l)}\|_F^2 + \|W_o^{(l)}\|_F^2) \end{aligned} \quad (19)$$

Where  $W$  refers to the weight matrices.  $W_e$ ,  $\hat{W}_d$ ,  $W_f$ ,  $W_i$ ,  $W_C$  and  $W_o$  are the weight matrices of the encoder, decoder, forget gate, input gate, cell memory and output gate respectively.  $K$  and  $L$  is the number of layers in the encoder/decoder and LSTM respectively.

The total loss function (equation 20) is composed of the reconstruction loss and the regularization with a constant  $\alpha$  to balance the terms.

$$\mathcal{L}_{total} = \mathcal{L} + \alpha \mathcal{L}_{reg} \quad (20)$$

The variational autoencoder methods use a different regularizer. They normalize the node embeddings compared to a prior. In traditional variational autoencoders this prior is a Normal distribution with mean 0 and standard deviation 1. In dynamic graph autoencoders [55], [80] the prior is still a Gaussian, but it is parameterised by previous observations. Equation 21 is the regularization term from [80].

$$\begin{aligned} & \text{KL}(q(Z^t | A^{<t}, X^{<t}, Z^{<t}) \\ & \quad \| p(Z^t | A^{<t}, X^{<t}, Z^{<t})) \end{aligned} \quad (21)$$

可以当做二分类问题处理（很容易理解，就是判断某两个节点中是否存在一条边），因此损失函数通常也是binary cross-entropy

变分自编码器使用不同的正则化方式：根据先验规范化node embedding，传统的VAE使用N(0, 1)高斯作为先验，而动态网络的先验高斯分布参数由之前的观测决定。（encoder会受到先前观测和当下状态的影响，和prior只由先前观测决定）

where  $q$  is the encoder distribution and  $p$  is the prior distribution. **KL** is the Kullback-Leibler divergence which measures the difference between two distributions. The  $A^{<t}$  indicate all adjacency matrices up to, but not including  $t$  and similarly for the other matrices. We can see that the prior is influenced by previous snapshots, but not by the current. Whereas the encoder is influenced by the previous and the current snapshot.

3) *Temporal Point Processes*: DyRep [36] models a dynamic network by parameterising a temporal point process. Its loss function influences how the temporal point process is optimized.

$$\mathcal{L} = - \sum_{p=1}^P \log(\lambda_p(t)) + \int_0^T \Lambda(\tau) d\tau \quad (22)$$

where  $P$  is the set of observed events,  $\lambda$  is the intensity function and  $\Lambda(\tau) = \sum_{u=1}^n \sum_{v=1}^n \sum_{k \in \{0,1\}} \lambda_k^{u,v}(\tau)$  is the survival probability for all events that did not happen. Survival probability indicates the probability of an event not happening [96]. The first term thus rewards a high intensity when an event happens, whereas the second term rewards a low intensity (high survival) of events that do not happen.

Trivedi *et al.* [36] further identify that calculating the integral of  $\Lambda$  is intractable. They get around that by sampling non-events and estimating the integral using Monte Carlo estimation, this is done for each mini-batch.

Loss Function的设计需要和实际的任务息息相关, 比如AutoEncoder重构动态网络, 就需要去“惩罚”最后输出与原始网络不一致的地方。而加入各种正则化项就是为了防止过拟合, 正则化项的设计取决于我们期望模型在什么假设或条件限制。

Link prediction is plagued by high class imbalance. It is a binary classification, a link either exists or not and most links will not exist. In fact, actual links tend to constitute less than 1% of all possible links [97]. AUC and precision@k are two commonly used evaluation metrics in static link prediction [95], [98]. If dynamic link prediction requires the prediction of both appearing and disappearing edges, the evaluation metric needs to reflect that. Furthermore traditional link prediction metrics have shortcomings when used in a dynamic setting [40].

For a detailed discussion on evaluation of link prediction we refer to Yang *et al.* [95] for static link prediction and Junuthula *et al.* [40] for dynamic link prediction evaluation.

- 1) **Area under the curve (AUC)**. Area under the curve (AUC) is used to evaluate a binary classification and has the advantage of being independent of the classification threshold. The AUC is the area under the receiver operating characteristic (ROC) curve. The ROC is a plot of the true positive rate and the false positive rate. The AUC evaluates predictions based on how well the classifier ranks the predictions, this provides a measure which is invariant of the classification threshold. In link prediction there has been little research into finding the optimal threshold [99], using the AUC for evaluation avoids this problem. Yang *et al.* [95] note that AUC can show deceptively high performance in link prediction due to the extreme class imbalance. They recommend the use of PRAUC instead.

- 2) **PRAUC**. The PRAUC is similar to the AUC except that it is the area under the precision recall curve. The metric is often used in highly imbalanced information retrieval problems [40].

PRAUC is recommended by Yang *et al.* [95] as a suitable metric for traditional (static) link prediction due to the deceptive nature of the ROC curve and because PRAUC shows a more discriminative view of classification performance.

- 3) **Fixed-threshold metrics**. One of the most common fixed threshold metrics in traditional link prediction is Precision@k. It is the ratio of items which are correctly predicted. From the ranking prediction the top  $k$  predictions are selected, then precision is the ratio  $\frac{k_r}{k}$ , where  $k_r$  is the number of correctly predicted links in the top  $k$  predictions.

While a higher precision indicates a higher prediction accuracy, it is dependent on the parameter  $k$ .  $k$  might be given on web scale information retrieval, where we care about the accuracy of the highest  $k$  ranked articles, in link prediction it is difficult to find the right cut-off [99].

Other fixed-threshold metrics include accuracy, recall and F1 among others [95]. These methods suffer from instability in their predictions, where a change of thresholds can lead to contradictory results [95]. This problem is also observed in dynamic link prediction [40]. Fixed-threshold metrics are not recommended unless the targeted problem have a natural threshold [95].

- 4) **Mean Average Precision (MAP)**. Mean average precision has been suggested as a measure for dynamic link prediction [97]. MAP is the mean of the average precision (AP) per node.
- 5) **Sum of absolute differences (SumD)**. Li *et al.* [97] pointed out that models often have similar AUC scores and suggested SumD as a stricter measurement of accuracy. It is simply, the number of mis-predicted links. The metric has different meanings depending on how many values are predicted, since it is not normalized according to the total number of links. Chen *et al.* considers SumD misleading for this reason [35]. The metric strictly punishes false positives, since there are so many links not appearing, a slightly higher rate of false positives will have a large impact on this metric.
- 6) **Error rate**. Since SumD suffers from several drawbacks an extension is suggested by Chen *et al.* [35]. Error rate normalizes SumD by the total number of existing links.

$$\text{Error Rate} = \frac{N_{\text{false}}}{N_{\text{true}}} \quad (23)$$

where  $N_{\text{false}}$  is the number of mis-predicted links and  $N_{\text{true}}$  is the number of existing links.

- 7) **GMAUC**. After a thorough investigation of evaluation metrics for dynamic link prediction, Junuthula *et al.* suggests GMAUC as an improvement over other metrics [40]. The key insight is that dynamic link prediction can be divided into two sub-problems: (i) predicting disappearance of links that already exist or appearance

Link Prediction问题通常是非常不平衡的分类问题, 不平衡分类问题直接使用AUC就不太合适, 因此后面几种指标提出了改进。

PRAUC能比AUC更好应对不平衡问题, Error Rate能有效避免类别不平衡问题, GMAUC包含了重新出现和消失的链接, 并能预测从未出现过的链接等。

当然也可以根据实际需要去结合几个指标对于模型进行评估。

of links that have once existed; and (ii) predicting links that have never been seen before. When the problem is divided in this way, each of the sub-problems takes on different characteristics.

Prediction of links that have never been seen before is equivalent to traditional link prediction, for which PRAUC is a suitable metric [95]. Prediction of already existing links is both the prediction of once seen links appearing and existing links disappearing. This is a more balanced problem than traditional link prediction, thus AUC is a suitable measure. [40] note that both the mean and the harmonic mean will lead to either the AUC or the PRAUC to dominate, thus the geometric mean is used to form a unified metric.

$$\text{GMAUC} = \sqrt{\frac{\text{PRAUC}_{\text{new}} - \frac{P}{P+N}}{1 - \frac{P}{P+N}} \cdot 2(\text{AUC}_{\text{prev}} - 0.5)} \quad (24)$$

$\text{PRAUC}_{\text{new}}$  is the PRAUC score of new links,  $\text{AUC}_{\text{prev}}$  is the AUC score of previously observed links. The authors note the advantages of GMAUC:

- Based on threshold curves, thus avoids pitfall of fixed-threshold metrics
- Accounts for differences between predicting new and already observed edges without having the metric to be dominated by either sub-problem.
- Any predictor that predicts only new edges or only previously observed edges gets a score of 0.

#### D. Discussion and summary

In this section we have provided an overview of how, given a dynamic network encoder, one can perform network topology prediction. The overview includes how methods from section III use their embeddings for prediction. This completes the journey from establishing a dynamic network, to encoding the dynamic topology, to predicting changes in the topology.

Prediction using a deep model requires decoding and the use of a loss function that captures temporal and structural information. Prediction is largely focused on time-conditioned link prediction and the two main modelling approaches are: (1) an architecture directly aimed at prediction; and (2) an architecture aimed at generating node embeddings which are then used for link prediction in a downstream step. Most dynamic network models surveyed fall into the second category, including all autoencoder approaches. All else being equal we would expect an architecture directly aimed at prediction to perform better than a two step architecture. This is because the first case will allow the entire architecture to optimize itself towards the prediction task.

Evaluation of dynamic link prediction is not trivial. Error rate is a good metric which avoids the class imbalance problem. The GMAUC metric incorporates the observation that reappearing and disappearing links are not an imbalanced classification. An evaluation of new methods should include both the Error rate and the GMAUC metric.

Prediction on dynamic networks is in its infancy. Deep models are largely focused on unattributed time-conditioned discrete link appearance prediction. This leaves opportunities for future work in a large range of prediction tasks, with some types of prediction still unexplored. Prediction based on continuous-time encoders is a particularly interesting frontier due to the representations inherent advantages and due to the limited amount of works in that area.

#### V. CHALLENGES AND FUTURE WORK

There are plenty of challenges and multiple avenues for improvement of deep learning both in the context of modelling and prediction of network topology.

**Expanding modelling and prediction repertoire.** In this work we have exclusively focused on dynamic network topology. However, complex networks are diverse and not only topology may vary. Topology dynamics can be represented as a 3-dimensional cube (Section II-D). However, real networks can be much more complex. Complex networks may have dynamic node and edge attributes, they may have directed and/or signed edges, be heterogeneous in terms of nodes and edges and be multilayered or multiplex. Each of these cases can be considered another dimension in the dynamic network hypercube. Designing deep learning models for encoding these network cases expand the repertoire of tasks on which deep learning can be applied. Which types of networks can be encoded can be expanded as well as an expansion of what kind of predictions can be made on those networks. For example, most DGNN models (and most GNN models) encode attributed dynamic networks but predict only graph topology without the node attributes.

**Adoption of advances in closely related fields.** Dynamic graph neural networks are based on GNNs and thus advances made to GNNs trickle down and can improve DGNNs. Challenges for GNNs include increasing modelling depth as GNNs struggle with vanishing gradients [100] and increasing scalability for large graphs [53]. As advancements are made in deep neural networks for time series and in GNNs these advancements can be applied to dynamic network modelling and prediction to improve performance. Similarly improvements in deep time-series modelling can easily be adapted to improve DGNNs.

**Continuous DGNNs.** Modelling temporal patterns is what distinguishes modelling dynamic graphs over modelling static graphs. Capturing these temporal patterns is key to making accurate predictions. However, most models rely on snapshots which are coarse temporal representations. Methods modelling network change in continuous time will offer fine-grained temporal modelling. Future work is needed for modelling and prediction of continuous time dynamic networks.

**Scalability.** Large scale datasets is a challenge for dynamic network modelling. Real world datasets tend to be so large that modelling becomes prohibitively slow. Dynamic networks either use a discrete representation in the form of snapshots, in which case processing of each snapshot is the bottleneck or continuous time modelling which scales with the number of interactions. A snapshot model will need to have frequent

(1) 扩展对图建模和预测的任务范围。比如我们可以不局限于预测图的拓扑结构, 还可以预测node attribute等, 当然这样就会使得任务更为复杂。

(2) 从相近领域的工作中获得灵感。由于DGNN来源于GNN, 因此GNN领域的进展都可以扩展到动态图的任务中加以尝试。

(3) 连续型DGNNs。当前大多数模型依赖于snapshot, 这其实是一种比较粗糙的temporal representations, 连续时间的方法往往能够捕捉更为深入细致的信息。

(4) Scalability。现实中往往存在着非常大规模的数据, 而这对于现有的一些方法提出了挑战。



snapshots in order to achieve high temporal granularity. In addition frequent snapshots might undermine the capacity to model a temporal network. Improvements in continuous-time modelling is likely to improve the performance of deep learning modelling on dynamic networks both in terms of temporal modelling capacity and ability to handle large networks.

Dynamic graph neural networks is a new exciting research direction with a broad area of applications. With these opportunities the field is ripe with potential for future work.

## REFERENCES

- [1] Steven H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268, March 2001.
- [2] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.
- [3] Charu Aggarwal and Karthik Subbian. Evolutionary network analysis: A survey. *ACM Computing Surveys (CSUR)*, 47(1):10, 2014.
- [4] Taisong Li, Bing Wang, Yasong Jiang, Yan Zhang, and Yonghong Yan. Restricted Boltzmann Machine-Based Approaches for Link Prediction in Dynamic Networks. *IEEE Access*, 6:29940–29951, 2018.
- [5] Othon Michail and Paul G Spirakis. Elements of the theory of dynamic networks. *Communications of the ACM*, 61(2):72–81, 2018.
- [6] Ashesh Jain, Amir R Zamir, Silvio Savarese, and Ashutosh Saxena. Structural-rnn: Deep learning on spatio-temporal graphs. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5308–5317, 2016.
- [7] Bing Yu, Haoteng Yin, and Zhanxing Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*, 2017.
- [8] Petter Holme. Modern temporal network theory: A colloquium. *The European Physical Journal B*, 88(9):234, September 2015.
- [9] Kathleen M Carley, Jana Diesner, Jeffrey Reminga, and Maksim Tsvetov. Toward an interoperable dynamic network analysis toolkit. *Decision Support Systems*, 43(4):1324–1347, 2007.
- [10] Jie Zhou, Ganqu Cui, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph Neural Networks: A Review of Methods and Applications. *arXiv:1812.08434 [cs, stat]*, December 2018.
- [11] Naoki Masuda and Renaud Lambiotte. *A Guide to Temporal Networks*. World Scientific, 2016.
- [12] Arnau Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-Varying Graphs and Dynamic Networks. In Hannes Frey, Xu Li, and Stefan Ruehrup, editors, *Ad-Hoc, Mobile, and Wireless Networks*, pages 346–359, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [13] Ziwei Zhang, Peng Cui, and Wenwu Zhu. Deep Learning on Graphs: A Survey. *arXiv:1812.04202 [cs, stat]*, December 2018.
- [14] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2020.
- [15] William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.
- [16] Palash Goyal and Emilio Ferrara. Graph Embedding Techniques, Applications, and Performance: A Survey. *Knowledge-Based Systems*, 151:78–94, July 2018.
- [17] Seyed Mehran Kazemi, Rishab Goel, Kshitij Jain, Ivan Kobyzev, Akshay Sethi, Peter Forsyth, and Pascal Poupard. Relational Representation Learning for Dynamic (Knowledge) Graphs: A Survey. *arXiv:1905.11485 [cs, stat]*, May 2019.
- [18] Giulio Rossetti and Rémy Cazabet. Community Discovery in Dynamic Networks: A Survey. *ACM Computing Surveys*, 51(2):1–37, February 2018.
- [19] Bomim Kim, Kevin H Lee, Lingzhou Xue, and Xiaoyue Niu. A review of dynamic network models with latent variables. *Statistics surveys*, 12:105–135, 2018.
- [20] Jian Zhang. A survey on streaming algorithms for massive graphs. In *Managing and Mining Graph Data*, pages 393–420. Springer, 2010.
- [21] Cornelius Fritz, Michael Lebacher, and Göran Kauermann. Tempus volat, hora fugit: A survey of tie-oriented dynamic network models in discrete and continuous time. *Statistica Neerlandica*, 2019.
- [22] Andrew McGregor. Graph stream algorithms: A survey. *ACM SIGMOD Record*, 43(1):9–20, 2014.
- [23] Saoussen Aouay, Salma Jamoussi, Faiez Gargouri, and Ajith Abraham. Modeling dynamics of social networks: A survey. In *CASoN*, pages 49–54. IEEE, 2014.
- [24] Giulio Rossetti. *Social Network Dynamics*. 2015.
- [25] Mayur Datar, Aristides Gionis, Piotr Indyk, and Rajeev Motwani. Maintaining Stream Statistics over Sliding Windows. *SIAM Journal on Computing*, 31(6):1794–1813, January 2002.
- [26] Matteo Morini, Patrick Flandrin, Eric Fleury, Tommaso Venturini, and Pablo Jensen. Revealing evolutions in dynamical networks. *arXiv preprint arXiv:1707.02114*, July 2017.
- [27] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, November 2014.
- [28] Daniel M Dunlavy, Tamara G Kolda, and Evrim Acar. Temporal link prediction using matrix and tensor factorizations. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(2):10, 2011.
- [29] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [30] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, pages 177–187. ACM, 2005.
- [31] Jiaxuan You, Rex Ying, Xiang Ren, William L. Hamilton, and Jure Leskovec. GraphRNN: A Deep Generative Model for Graphs. *CoRR*, abs/1802.08773, 2018.
- [32] Mark Newman. *Networks*. Oxford university press, 2018.
- [33] Albert-László Barabási, Hawoong Jeong, Zoltan Nédá, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. *Physica A: Statistical mechanics and its applications*, 311(3-4):590–614, 2002.
- [34] Kevin Xu. Stochastic block transition models for dynamic networks. In *Artificial Intelligence and Statistics*, pages 1079–1087, 2015.
- [35] Jinyin Chen, Jian Zhang, Xuanheng Xu, Chengbo Fu, Dan Zhang, Qingpeng Zhang, and Qi Xuan. E-LSTM-D: A Deep Learning Framework for Dynamic Network Link Prediction. *arXiv preprint arXiv:1902.08329*, 2019.
- [36] Rakshit Trivedi, Mehrdad Farajtabar, Prasenjeet Biswal, and Hongyuan Zha. DyRep: Learning representations over dynamic graphs. In *International Conference on Learning Representations*, 2019.
- [37] Kevin S. Xu and Alfred O. Hero III. Dynamic stochastic blockmodels for time-evolving social networks. *IEEE Journal of Selected Topics in Signal Processing*, 8(4):552–562, August 2014.
- [38] Akanda Wahid-Ul Ashraf, Marcin Budka, and Katarzyna Musial. Simulation and Augmentation of Social Networks for Building Deep Learning Models. *arXiv preprint arXiv:1905.09087*, May 2019.
- [39] Tom A. B. Snijders, Gerhard G. van de Bunt, and Christian E. G. Steglich. Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32(1):44–60, January 2010.
- [40] Ruthwik R Junuthula, Kevin S Xu, and Vijay K Devabhaktuni. Evaluating link prediction accuracy in dynamic networks with added and removed edges. In *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)*, pages 377–384. IEEE, 2016.
- [41] Aswathy Divakaran and Anuraj Mohan. Temporal link prediction: A survey. *New Generation Computing*, pages 1–46, 2019.
- [42] Xiaoyi Li, Nan Du, Hui Li, Kang Li, Jing Gao, and Aidong Zhang. A deep learning approach to link prediction in dynamic networks. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 289–297. SIAM, 2014.
- [43] Carter T. Butts and Christopher Steven Marcum. A relational event approach to modeling behavioral dynamics. In Andrew Pilny and Marshall Scott Poole, editors, *Group processes: Data-driven computational approaches*, pages 51–92. Springer International Publishing, Cham, 2017.
- [44] Steve Hanneke, Wenjie Fu, and Eric P. Xing. Discrete temporal models of social networks. *Electronic Journal of Statistics*, 4:585–605, 2010.
- [45] Per Block, Johan Koskinen, James Hollway, Christian Steglich, and Christoph Stadtfeld. Change we can believe in: Comparing longitudinal network models on consistency, interpretability and predictive power. *Social Networks*, 52:180–191, January 2018.
- [46] Anna Goldenberg, Alice X. Zheng, Stephen E. Fienberg, and Edoardo M. Airoldi. A Survey of Statistical Network Models. *Found-*

- dations and Trends® in Machine Learning*, 2(2):129–233, February 2010.
- [47] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: Moving beyond fingerprints. *Journal of computer-aided molecular design*, 30(8):595–608, 2016.
  - [48] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarell, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In *Advances in Neural Information Processing Systems*, pages 2224–2232, 2015.
  - [49] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 974–983. ACM, 2018.
  - [50] Federico Monti, Michael Bronstein, and Xavier Bresson. Geometric matrix completion with recurrent multi-graph neural networks. In *Advances in Neural Information Processing Systems*, pages 3697–3707, 2017.
  - [51] Jiezhong Qiu, Jian Tang, Hao Ma, Yuxiao Dong, Kuansan Wang, and Jie Tang. DeepInf: Social Influence Prediction with Deep Learning. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '18*, pages 2110–2119, 2018.
  - [52] Yozen Liu, Xiaolin Shi, Lucas Pierce, and Xiang Ren. Characterizing and forecasting user engagement with in-app action graph: A case study of snapchat. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2023–2031, 2019.
  - [53] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.
  - [54] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Deep learning for time series classification: A review. *Data Mining and Knowledge Discovery*, 33(4):917–963, July 2019.
  - [55] Da Xu, Chuanwei Ruan, Kamiya Motwani, Evren Korpoglu, Sushant Kumar, and Kannan Achan. Generative Graph Convolutional Network for Growing Graphs. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3167–3171, May 2019.
  - [56] Thomas N. Kipf and Max Welling. Variational graph auto-encoders. *CoRR*, abs/1611.07308, 2016.
  - [57] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.
  - [58] Youngjoo Seo, Michaël Defferrard, Pierre Vandergheynst, and Xavier Bresson. Structured Sequence Modeling with Graph Convolutional Recurrent Networks. In Long Cheng, Andrew Chi Sing Leung, and Seichi Ozawa, editors, *Neural Information Processing, Lecture Notes in Computer Science*, pages 362–373. Springer International Publishing, 2018.
  - [59] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems* 29, pages 3844–3852. Curran Associates, Inc., 2016.
  - [60] Felix A Gers, Nicol N Schraudolph, and Jürgen Schmidhuber. Learning Precise Timing with LSTM Recurrent Networks. *Journal of machine learning research*, 3(Aug):115–143, 2002.
  - [61] Apurva Narayan and Peter H. O’N Roe. Learning Graph Dynamics using Deep Neural Networks. *IFAC-PapersOnLine*, 51(2):433–438, January 2018.
  - [62] Mathias Niepert, Mohamed Ahmed, and Konstantin Kutzkov. Learning convolutional neural networks for graphs. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 2014–2023. JMLR.org, 2016.
  - [63] Aynaz Taheri, Kevin Gimpel, and Tanya Berger-Wolf. Learning to Represent the Evolution of Dynamic Graphs with Recurrent Models. In *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, pages 301–307, New York, NY, USA, 2019. ACM.
  - [64] Yujia Li, Daniel Tarlow, Marc Brockschmidt, and Richard S. Zemel. Gated graph sequence neural networks. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.
  - [65] Franco Manessi, Alessandro Rozza, and Mario Manzo. Dynamic graph convolutional networks. *Pattern Recognition*, 97:107000, January 2020.
  - [66] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Comput.*, 9(8):1735–1780, November 1997.
  - [67] Aravind Sankar, Yanhong Wu, Liang Gou, Wei Zhang, and Hao Yang. Dysat: Deep neural representation learning on dynamic graphs via self-attention networks. In James Caverlee, Xia (Ben) Hu, Mounia Lalmas, and Wei Wang, editors, *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining, Houston, TX, USA, February 3-7, 2020*, pages 519–527. ACM, 2020.
  - [68] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018.
  - [69] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All you Need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 5998–6008. Curran Associates, Inc., 2017.
  - [70] Xingjian Shi, Zhoung Chen, Hao Wang, Dit-Yan Yeung, Wai-Kin Wong, and Wang-chun Woo. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In Corinna Cortes, Neil D. Lawrence, Daniel D. Lee, Masashi Sugiyama, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada*, pages 802–810, 2015.
  - [71] Aldo Pareja, Giacomo Domeniconi, Jie Chen, Tengfei Ma, Toyotaro Suzumura, Hiroki Kanezashi, Tim Kaler, and Charles E. Leiserson. Evolvecn: Evolving graph convolutional networks for dynamic graphs. *CoRR*, abs/1902.10191, 2019.
  - [72] Jinyin Chen, Xuanheng Xu, Yangyang Wu, and Haibin Zheng. GC-LSTM: graph convolution embedded LSTM for dynamic link prediction. *CoRR*, abs/1812.04206, 2018.
  - [73] Jia Li, Zhichao Han, Hong Cheng, Jiao Su, Pengyun Wang, Jianfeng Zhang, and Lujia Pan. Predicting Path Failure In Time-Evolving Graphs. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD '19*, pages 1279–1289, New York, NY, USA, 2019. ACM.
  - [74] Woojeong Jin, He Jiang, Meng Qu, Tong Chen, Changlin Zhang, Pedro Szekely, and Xiang Ren. Recurrent event network: Global structure inference over temporal knowledge graph. *arXiv: 1904.05530*, 2019.
  - [75] Michael Schlichtkrull, Thomas N. Kipf, Peter Bloem, Rianne van den Berg, Ivan Titov, and Max Welling. Modeling Relational Data with Graph Convolutional Networks. In Aldo Gangemi, Roberto Navigli, Maria-Esther Vidal, Pascal Hitzler, Raphaël Troncy, Laura Hollink, Anna Tordai, and Mehdi Alam, editors, *The Semantic Web, Lecture Notes in Computer Science*, pages 593–607. Springer International Publishing, 2018.
  - [76] Palash Goyal, Sujit Rokka Chhetri, Ninareh Mehrabi, Emilio Ferrara, and Arquimedes Canedo. DynamicGEM: A library for dynamic graph embedding methods. *arXiv preprint arXiv:1811.10734*, 2018.
  - [77] Daixin Wang, Peng Cui, and Wenwu Zhu. Structural Deep Network Embedding. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pages 1225–1234, New York, NY, USA, 2016. ACM.
  - [78] Tianqi Chen, Ian J. Goodfellow, and Jonathon Shlens. Net2net: Accelerating learning via knowledge transfer. In Yoshua Bengio and Yann LeCun, editors, *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.
  - [79] Palash Goyal, Sujit Rokka Chhetri, and Arquimedes Canedo. Dyngraph2vec: Capturing Network Dynamics using Dynamic Graph Representation Learning. *Knowledge-Based Systems*, page S0950705119302916, July 2019.
  - [80] Ehsan Hajiramezanali, Arman Hasanazadeh, Krishna R. Narayanan, Nick Duffield, Mingyuan Zhou, and Xiaoning Qian. Variational graph recurrent neural networks. In Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d’Alché-Buc, Emily B. Fox, and Roman Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems*

- 2019, *NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 10700–10710, 2019.
- [81] Mingzhang Yin and Mingyuan Zhou. Semi-Implicit Variational Inference. In *International Conference on Machine Learning*, pages 5660–5669, July 2018.
  - [82] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.
  - [83] Zhengwei Wang, Qi She, and Tomas E Ward. Generative adversarial networks: A survey and taxonomy. *arXiv preprint arXiv:1906.01529*, 2019.
  - [84] Kai Lei, Meng Qin, Bo Bai, Gong Zhang, and Min Yang. Gcn-gan: A non-linear temporal link prediction model for weighted dynamic networks. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 388–396, IEEE, 2019.
  - [85] Yun Xiong, Yao Zhang, Hanjie Fu, Wei Wang, Yangyong Zhu, and S Yu Philip. Dyngraphgan: Dynamic graph embedding via generative adversarial networks. In *International Conference on Database Systems for Advanced Applications*, pages 536–552. Springer, 2019.
  - [86] Yao Ma, Ziyi Guo, Zhaochun Ren, Eric Zhao, Jiliang Tang, and Dawei Yin. Streaming Graph Neural Networks. *arXiv:1810.10627 [cs, stat]*, October 2018.
  - [87] Srijan Kumar, Xikun Zhang, and Jure Leskovec. Predicting dynamic embedding trajectory in temporal interaction networks. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1269–1278, 2019.
  - [88] Inci M. Baytas, Cao Xiao, Xi Zhang, Fei Wang, Anil K. Jain, and Jiayu Zhou. Patient Subtyping via Time-Aware LSTM Networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17*, pages 65–74, Halifax, NS, Canada, 2017. ACM Press.
  - [89] Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 3462–3471. PMLR, 2017.
  - [90] Boris Knyazev, Carolyn Augusta, and Graham W Taylor. Learning temporal attention in dynamic graphs with bilinear interactions. *arXiv preprint arXiv:1909.10367*, 2019.
  - [91] Thomas N. Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard S. Zemel. Neural relational inference for interacting systems. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 2693–2702. PMLR, 2018.
  - [92] Zhen Han, Yuyi Wang, Yunpu Ma, Stephan Günnemann, and Volker Tresp. The graph hawkes network for reasoning on temporal knowledge graphs. *arXiv preprint arXiv:2003.13432*, 2020.
  - [93] Hongyuan Mei and Jason M Eisner. The neural hawkes process: A neurally self-modulating multivariate point process. In *Advances in neural information processing systems*, pages 6754–6764, 2017.
  - [94] Changmin Wu, Giannis Nikolentzos, and Michalis Vazirgiannis. EvoNet: A neural network for predicting the evolution of dynamic graphs. *arXiv preprint arXiv:2003.00842*, 2020.
  - [95] Yang Yang, Ryan N. Lichtenwalter, and Nitesh V. Chawla. Evaluating link prediction methods. *Knowl. Inf. Syst.*, 45(3):751–782, 2015.
  - [96] Odd Aalen, Ormulf Borgan, and Hakon Gjessing. *Survival and Event History Analysis: A Process Point of View*. Springer Science & Business Media, 2008.
  - [97] Taisong Li, Jiawei Zhang, S Yu Philip, Yan Zhang, and Yonghong Yan. Deep dynamic network embedding for link prediction. *IEEE Access*, 6:29219–29230, 2018.
  - [98] Linyuan Lü and Tao Zhou. Link prediction in complex networks: A survey. *Physica A*, 390(6):11501170, 2011.
  - [99] Víctor Martínez, Fernando Berzal, and Juan-Carlos Cubero. A Survey of Link Prediction in Complex Networks. *ACM Comput. Surv.*, 49(4):69:1–69:33, December 2016.
  - [100] Guohao Li, Matthias Müller, Ali K. Thabet, and Bernard Ghanem. Deepgcns: Can gcns go as deep as cnns? In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, pages 9266–9275. IEEE, 2019.



TABLE VI: A summary of notation and abbreviations used in this work. Some model names look like abbreviations but are in fact simply the name of the model (or the authors did not explicitly state what the abbreviation stood for). These include: PATCHY-SAN, DyGGNN, RgGNN, EvolveGCN, JODIE, GC-LSTM, GCN-GAN, DynGraphGAN and DyREP.

Notation/abbreviation	Description
$\odot$	Element wise product
$G$	Static graph
$G^t$	Static graph at time $t$
$DG$	Discrete dynamic graph
$CG$	Continuous dynamic graph
$V$	The set of nodes in a graph
$E$	The set of edges in a graph
$v$	A node $v \in V$
$e$	An edge $e \in E$
$t$	Time step / event time
$\bar{t}$	Time step just before time $t$
$< t$	All time steps up until time $t$
$\Delta$	Duration
$n$	Number of nodes
$d$	Dimensions of a node feature vector
$l$	Dimensions of a GNN produced hidden feature vector
$k$	Dimensions of a RNN/self-attention produced hidden feature vector
$X^t$	Feature matrix at time $t$
$A^t$	Adjacency matrix at time $t$
$\hat{A}$	Predicted adjacency matrix
$z_u^t$	GNN produced hidden feature vector of node $u$ at time $t$
$h_u^t$	RNN/self-attention produced hidden feature vector of node $u$ at time $t$
$Z^t$	GNN produced hidden feature matrix at time $t$
$H^t$	RNN/self-attention produced hidden feature matrix at time $t$
$\lambda$	Conditional intensity function
$\sigma$	The sigmoid function
$W, U, w, b, \omega, \psi,$	Learnable model parameters
GNN	Graph neural network
DGNN	Dynamic graph neural network
RNN	Recurrent neural network
LSTM	Long-term short term memory
GAN	Generative adversarial network
CNN	Convolutional neural network
TPP	Temporal point process
RGM	Random graph model
ERGM	Exponential random graph model
TERGM	Temporal exponential random graph model
SAOM	Stochastic actor oriented model
REM	Relational event model
GCN	Graph Convolutional Network
GGNN	Gated Graph Neural Network
R-GCN	Relational Graph Convolutional Network
convLSTM	Convolutional LSTM
GraphRNN	Graph recurrent neural network
G-GCN	Generative graph convolutional network
VGAE	Variational graph autoencoder
GCRN-M1	Graph convolutional recurrent network model 1
GCRN-M2	Graph convolutional recurrent network model 2
WD-GCN	Waterfall dynamic graph convolutional network
CD-GCN	Concatenated dynamic graph convolutional network
DySAT	Dynamic Self-Attention Network
LRGCN	Long Short-Term Memory R-GCN
RE-Net	Recurrent Event Network
DynGEM	Dynamic graph embedding model
E-LSTM-D	Encode-LSTM-Decode
VGRNN	Variational graph recurrent neural network
SI-VGRNN	Semi-implicit VGRNN
SGNN	Streaming graph neural network
LDG	Latent dynamic graph
GHN	Graph Hawkes network