# Time to Cite: Modeling Citation Networks using the Dynamic Impact Single-Event Embedding Model

**Nikolaos Nakis**
Technical University of Denmark

**Abdulkadir Çelikkanat**
Technical University of Denmark

**Louis Boucherie**
Technical University of Denmark

**Sune Lehmann**
Technical University of Denmark

**Morten Mørup**
Technical University of Denmark

## Abstract

Understanding the structure and dynamics of scientific research, i.e., the science of science (SciSci), has become an important area of research in order to address imminent questions including how scholars interact to advance science, how disciplines are related and evolve, and how research impact can be quantified and predicted. Central to the study of SciSci has been the analysis of citation networks. Here, two prominent modeling methodologies have been employed: one is to assess the citation impact dynamics of papers using parametric distributions, and the other is to embed the citation networks in a latent space optimal for characterizing the static relations between papers in terms of their citations. Interestingly, citation networks are a prominent example of single-event dynamic networks, i.e., networks for which each dyad only has a single event (i.e., the point in time of citation). We presently propose a novel likelihood function for the characterization of such single-event networks. Using this likelihood, we propose the Dynamic Impact Single-Event Embedding model (DISEE). The DISEE model characterizes the scientific interactions in terms of a latent distance model in which random effects account for citation heterogeneity while the time-varying impact is characterized using existing parametric representations for assessment of dynamic impact. We highlight the proposed approach on several real citation networks finding that the DISEE well reconciles static latent distance network embedding approaches with classical dynamic impact assessments.

Proceedings of the 27[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238. 

## 1 Introduction

The abundance of scientific data has established the science of science (SciSci) as a vital tool for understanding scientific research, as well as predicting future outcomes, research directions, and the overall evolution of science (Fortunato et al., 2018). More specifically, SciSci studies the methods of science itself, searching for answers to important questions such as how scholars interact to advance science, how different disciplinary boundaries are removed, and how research impact can be quantified and predicted. SciSci is an interdisciplinary field with various prominent research directions including but not limited to, scientific novelty and innovation quantification (Uzzi et al., 2013; Stephan et al., 2017; Wang et al., 2017; Cokol et al., 2005; Foster et al., 2015), analysis of career success dynamics of scholars (Petersen et al., 2012; Freeman et al., 2001; Larivière et al., 2013; Duch et al., 2012; Noorden, 2015; Letchford et al., 2015, 2016), characterization of scientific collaborations (Council et al., 2015; Larivière et al., 2015; Wu et al., 2017; Milojević, 2014; Borjas and Doran, 2015) as well as citation and research impact dynamics (Waltman, 2016a; Hirsch, 2005; Garfield, 1972; Radicchi et al., 2008; Waltman et al., 2012; Golosovsky and Solomon, 2012; Waltman, 2016b).

A major focus has been given to the understanding of SciSci through the lens of complex network analysis, studying the structural properties and dynamics, of

the naturally occurring graph data describing SciSci. These include collaboration networks describing how scholars cooperate to advance various scientific fields. In particular, pioneering works (Newman, 2001a,b,c) have analyzed various network statistics such as degree distribution, clustering coefficient, and average shortest paths. Furthermore, citation networks define an additional prominent case where graph structure data describe SciSci. Citation networks, essentially describe the directed relationships of papers (nodes) with an edge occurring between a dyad if paper $A$ cites paper $B$, e.g. A→B. Studies focusing on citation networks have shown power-law and exponential family degree distributions (Redner, 1998), sub-field community structures (Gualdi et al., 2011a), and tree-like backbone topologies (Gualdi et al., 2011b). Lastly, bipartite network structures can emerge by defining networks describing author-paper relationships, including indirect author connections through a collaboration paper or through their citing patterns (Goldstein et al., 2005; Zhang et al., 2008; Barabási et al., 2012).

In this paper, we focus on citation networks that allow for paper impact characterization. Notably, such networks are directed and dynamic ideally having an upper triangular adjacency matrix when nodes are sorted by time due to the time-causal structure of citations (i.e., new papers can only cite past papers, where rows denote cited papers and columns citing papers).

Initial works for paper impact quantification utilized classical machine learning methods on various scholarly features, as well as paper textual information. Methods used to estimate future citations included linear/logistic regression, k-nearest neighbors, support vector machines, random forests, and many more (Davletov et al., 2014; Singh et al., 2015; Bhat et al., 2015; Shibata et al., 2012; Yan et al., 2011; Ibáñez et al., 2009). These studies focused primarily on carefully designing and including proper features (H-index, impact factor (IF), etc.) to be used for the impact prediction task. While these methods attracted lots of attention, they have a major limitation where papers with very similar features define much different citation distributions and attention patterns that are not characterized.

Later works tried to define impact on the paper level by treating the accumulation of citations through time as a time series. In the original work of (Redner, 2005), Redner proposed a log-normal distribution to fit the cumulative citation distribution for papers published during a 110 year period in Physical Review. This was followed by (Eom and Fortunato, 2011) using a shifted power-law distribution on the same networks. Furthermore, another widely used distribution modeling citation dynamics is the Tsallis distribution proposed in (Wallace et al., 2008). The log-normal and Tsallis distribution share a lot of similarities but in literature, the log-normal is preferred due to its simplicity. Later works combined the important characteristic of preferential attachment with the log-normal distribution (Wu et al., 2014; Wang et al., 2013, 2014), as well as, the Poisson process (Shen et al., 2014; Xiao et al., 2016).

Various prominent Graph Representation Learning (GRL) methods have also been applied to citation graphs (Sen et al., 2008; Leskovec et al., 2005) as they have been very popular network choices for assessing downstream task performance, such as link prediction and node classification (Perozzi et al., 2014; Grover and Leskovec, 2016; Qiu et al., 2018). Recently, Graph Neural Networks (GNN)s have also been used including GraphSAGE (Hamilton et al., 2017), the Adaptive Channel Mixing GNN (Luan et al., 2021) and Convolutional Graph Neural Networks (Kipf and Welling, 2017). Despite these works defining strong models, powerful link predictors, and node classifiers they do not explicitly account for impact characterization, nor for the dynamic way that paper citations appear.

Notably, citation networks are dynamic. Whereas dynamic modeling approaches can uncover structures obscured when aggregating networks across time to form static networks, the dynamic modeling approaches are in general based on the assumption that multiple links occur between the dyads in time. Importantly, for continuous-time modeling, this has typically been accounted for using Poisson Process likelihoods (Blundell et al., 2012; Fan et al., 2021; Çelikkanat et al., 2022; Celikkanat et al., 2023) including likelihoods accounting for burstiness and reciprocating behaviors by use of the Hawkes process (Blundell et al., 2012; Arastuie et al., 2020; Delattre et al., 2016; Zuo et al., 2018; Fan et al., 2021). To account for the high degree of complex interactions in time, advanced dynamic latent representations have further been proposed considering both discrete-time (Ishiguro et al., 2010; Herlau et al., 2013; Heaukulani and Ghahramani, 2013; Perozzi et al., 2014; Grover and Leskovec, 2016; Durante and Dunson, 2014, 2016; Kim et al., 2018; Sarkar and Moore, 2005; Sewell and Chen, 2015) and continuous-time dynamics (Blundell et al., 2012; Arastuie et al., 2020; Delattre et al., 2016; Fan et al., 2021; Çelikkanat et al., 2022; Celikkanat et al., 2023), including GNNs with time-evolving latent representations (Trivedi et al., 2019; Rossi et al., 2020). For surveys of such dynamic modeling approaches see also (Xue et al., 2022; Kazemi et al., 2020).

Importantly, citation networks are a class of dynamic networks characterized by a single event occurring between dyads. We presently denote such network a Single-Event Network (SEN). I.e., links occur only once at the time of the paper publication. However, neither

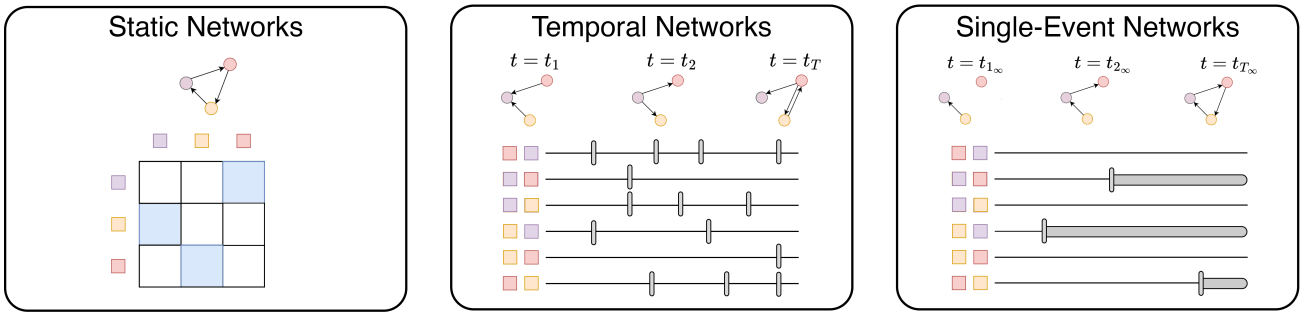



Figure 1: Examples of three different types of networks based on their temporal structure. Round points represent network nodes, square points make up the corresponding colored node dyads, arrows represent directed relationships between two nodes, vertical lines represent events, and black lines are the timelines while grey bold lines show that a link (event) appeared once and cannot be observed again. *Left panel:* Static networks where links occur once and there is no temporal information available. *Middle panel:* Temporal networks where links are events in time and can be observed multiple times along the timeline. *Right panel:* Single-event networks (SENs) where links appear in a temporal manner but can occur only once for each dyad, defining edges as single events.

of the existing dynamic network modeling approaches explicitly account for SENs. Whereas continuous-time modeling approaches are designed for multiple events, thereby easily over-parameterizing such highly sparse networks, static networks can easily be applied to such networks by disregarding the temporal structure but thereby potentially miss important structural information given by the event time. Despite these limitations, to the best of our knowledge, existing dynamic network modeling approaches do not explicitly account for single-event occurrences. In Figure 1, we provide an example of three cases of networks that define static, traditional event-based dynamic networks, as well as SENs. We here observe how static networks are completely blind to the temporal information that single-event networks capture while it is also evident that they differ from traditional event-based temporal networks where each dyad can have multiple events across time.

When modeling SENs, the single event occurrence makes the networks highly sparse. To account for the high degree of sparsity of SENs we use as a starting point the static Latent Distance Modeling (LDM) approaches proposed in (Hoff et al., 2002b) in which static networks are embedded in a low dimensional space and the relative distance between the nodes used to parameterize the probability of observing links between the nodes. Importantly, these modeling approaches have been found to provide easily interpretable low-dimensional ($D = 2$ and $D = 3$) network representations with favorable representation learning performance in tasks including link prediction and node classification (Nakis et al., 2023b,a). The LDM has been generalized to distances beyond Euclidean, including squared Euclidean distances and hyperbolic embeddings (Nickel and Kiela, 2017, 2018) as well as to account for degree heterogeneity through the use of node-specific biases (denoted random effects) (Hoff, 2005; Krivitsky et al., 2009; Nakis et al., 2023b) which we presently refer to as the mass of a paper. Notably, we define paper masses based on their citation dynamics through time, regulated by their distance in a latent space used to embed the structure of the citation network. Specifically, to account for single-event network dynamics, we endow the cited papers (receiving nodes) with a temporal profile in which a parametric function as used for traditional paper impact assessment (Redner, 2005) is employed to regulate the nodes' citation activity in time forming the *Dynamic Impact Single-Event Embedding Model* (DISEE). In particular, our contributions are:

**1) We derive the single-event Poisson Process (SE-PP).** As paper citation networks only include a single event we augment the Poisson Process likelihood to have support only for single events forming the single event Poisson Process.

**2) We propose the DISEE model based on the SE-PP for SENs.** We characterize the rate of interaction within a latent distance model augmented such that citations are generated relative to the degree to which a paper cites and a paper is being cited at a given time point interpreted as masses of the citing and cited papers in which the mass of the cited paper is dynamically evolving.

**3) We demonstrate how DISEE reconciles conventional impact modeling with latent distance embedding procedures.** We demonstrate how DISEE enables accurate dynamic characterization of citation impact similar to conventional paper impact modeling procedures while at the same time provid-

ing low-dimensional embeddings accounting for the structure of citation networks. We highlight this reconciliation on three real networks covering three distinct fields of science.

The paper is organized as follows. In Section 2, we present the single-event Poisson Process (SE-PP) for the modeling of single-event networks (SENs). In Section 3 we discuss procedures for characterizing dynamic impact, and in section 4, we demonstrate how existing embedding procedures can be reconciled with dynamic impact modeling using the SE-PP by the proposed Dynamic Impact Single-Event Embedding Model (DISEE). In Section 5, we present our results on the three distinct citation networks contrasting the performance to the corresponding conventional impact dynamic modeling, as well as, the powerful static LDM (Nakis et al., 2023b). Section 6 concludes our results.

## 2 The Single-Event Poisson Process

Before presenting our modeling strategy for the links of networks, we will first establish the notations used throughout the paper. We utilize the conventional symbol, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, to denote a directed Single-Event-Network over the timeline $[0, T]$ where $\mathcal{V} = \{1, \dots, N\}$ is the vertex and $\mathcal{E} \subseteq \mathcal{V}^2 \times [0, T]$ is the edge set such that each node pair has at most one link. Hence, a tuple, $(i, j, t_{ij}) \in \mathcal{E}$, shows a directed event (i.e., instantaneous link) from source node $j$ to target $i$ at time $t_{ij} \in [T]$, and there can be at most one $(i, j, t_{ij})$ element for each $(i, j) \in \mathcal{V}^2$ and some $t_{ij} \in [0, T]$.

We always assume that the timeline starts at 0 and the last time point is $T$, and we represent the interval by symbol, $[T]$. We employ $t_1 \leq t_2 \leq \cdots \leq t_N$ to indicate the appearance times of the corresponding nodes $1, 2, \dots, N \in \mathcal{V}$, and we suppose that node labels are sorted with respect to the minimum of their incoming edge times. In other words, if $i < j$, then we know that there exists $k \in \mathcal{V}$ such that $t_{ik} \leq t_{jl}\ \forall (j, l, t_{jl}) \in \mathcal{E}$.

The Inhomogeneous Poisson Point (IPP) process is a widely employed approach for modeling the number of events exhibiting varying characteristics depending on the time they occur (González et al., 2016). They are parametrized by an *intensity* or *rate function* representing the average event density, such that the probability of sampling $m$ event points on the interval $[T]$ is

$$p_M(M(T) = m) := \frac{[\Lambda(T)]^m}{m!} \exp(-\Lambda(T)), \quad (1)$$

where $M(T)$ is the random variable showing the number of events occurring over the interval $[T]$, and $\Lambda(T)$ is defined as $\int_0^T \lambda(t')\mathrm{d}t'$ for the intensity function $\lambda : [T] \to \mathbb{R}^+$. We refer unfamiliar readers to the work (Streit, 2010) for more details concerning the process.

We employ a Poisson point process for characterizing the occurrence time of a link (i.e., a single event point indicating the publication or citation time), unlike their conventional practice in modeling the occurrence of an arbitrary number of events between a pair of nodes. Hence, we suppose that a pair can have at most one interaction (i.e., link), and we discretize the probability of sampling $m$ events given in Eq. (1) as having an event and no event cases. More formally, by applying Bayes' rule, we can write it as a conditional distribution of $M(t)$ being equal to $m \in \{0, 1\}$ as follows:

$$\begin{aligned} p_{M|M\leq 1}(M(T) = m) &= \frac{p_M\left(M(T) = m\right)}{\sum_{m'=0}^{1} p_M\left(M(T) = m'\right)} \\ &= \frac{\exp\left(-\Lambda(T)\right)\left[\Lambda(T)\right]^m}{\exp(-\Lambda(T)) + \exp(-\Lambda(T))\Lambda(T)}. \end{aligned} \quad (2)$$

Therefore, the conditional probability of an event for the proposed *Single-Event Poisson Process* is equal to:

$$p_{M|M\leq 1}\left(M(T) = 1\right) = \frac{\Lambda(T)}{1 + \Lambda(T)}. \quad (3)$$

It is also not difficult to derive the likelihood function of the process based on Eq (3). Let $(Y, \Theta)$ be random variables where $Y$ shows whether a link exists and $\Theta$ indicates the time of the corresponding link (if it exists). Then, we can write the likelihood of $(Y, \Theta)$ evaluated at $(1, t^*)$ as follows:

$$\begin{aligned} p_{Y,\Theta}\left(1, t^*\right) &= p_Y\left\{Y = 1\right\} p_{\Theta|Y}\left\{\Theta = t^* | Y = 1\right\} \\ &= \left(\frac{\Lambda(T)}{1 + \Lambda(T)}\right)\left(\frac{\lambda(t^*)}{\Lambda(T)}\right) = \frac{\lambda(t^*)}{1 + \Lambda(T)}. \end{aligned} \quad (4)$$

As a result, we can write the log-likelihood of the whole network by assuming that each dyad follows the Single-Event Poisson Process as follows:

$$\begin{aligned} &\mathcal{L}_{SE-PP}(\Omega) := \log p(\mathcal{G}|\Omega) \\ &= \sum_{1\leq i,j\leq N} \Big( y_{ij} \log \lambda(t_{ij}) - \log\left(1 + \Lambda_{ij}(t_i, T)\right)\Big), \end{aligned} \quad (5)$$

where $\Omega$ is the model hyper-parameters and $\Lambda_{ij}(t_i, T) := \int_{t_i}^{T} \lambda_{ij}(t')\mathrm{d}t'$. Note that for a homogeneous Poisson process with constant intensity $\lambda_{ij}$ for each $(i, j)$ pair, the probability of having an event throughout the timeline is equal to $\Lambda_{ij}(T)/(1 + \Lambda_{ij}(T)) = T\lambda_{ij}/(1 + T\lambda_{ij})$ by Eq. (3). In this regard, the objective function stated in Eq. (5) is equivalent to a static Bernoulli model (Hoff et al., 2002a):

$$\begin{aligned} &\mathcal{L}_{Bern}(\Omega) := \log p(\mathcal{G}|\Omega) \\ &\qquad = \sum_{i,j\in\mathcal{V}} \Big( y_{ij} \log\left(\tilde{\lambda}_{ij}\right) - \log\Big(1 + \tilde{\lambda}_{ij}\Big)\Big), \end{aligned} \quad (6)$$

using the re-parameterization $\tilde{\lambda}_{ij} := T\lambda_{ij}$.

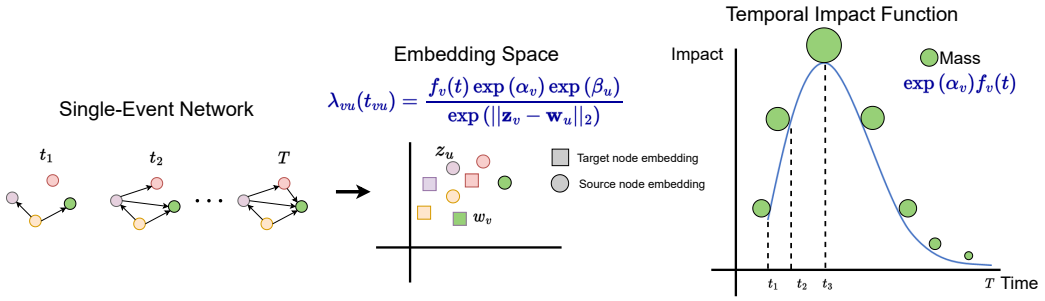



Figure 2: DISEE procedure overview. The model defines for the SE-PP an intensity function introducing two sets of static embeddings distinguishing between source $\mathbf{w}_u$ and target $\mathbf{z}_v$ node embeddings. Furthermore, each node is assigned its own random effect, distinguishing again the source $\beta_u$ and target $\alpha_v$ roles. The random effects can be parameterized to represent source and target masses through the exponential function. Finally, for each target node of the network, the model defines an impact function $f_v(t)$ yielding a temporal impact characterization of the nodes' link dynamics, which controls the nodes' time-varying mass as $\exp(\alpha_v) f_v(t)$.

**Comparison with a Bernoulli process DISEE:** The reason for not working with the Bernoulli Process, is its discrete nature, modeling independent trials indexed by a set and not a time variable (Grimmett and Stirzaker, 2001). The process is usually defined at discrete time points (e.g., coin tosses every minute). Contrary, the Poisson Process is a counting process that models the number of events occurring over continuous intervals of time.

## 3 Dynamic Impact Characterization

In the realm of impact analysis and risk assessment, characterizing dynamic events is pivotal in understanding and managing potential consequences. We know that papers generally undergo the process of aging over time as new concepts replace older concepts. We model the distribution of the impact of paper $i \in \mathcal{V}$ by the probability density function (pdf) of the Log Normal distribution:

$$f_i(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln t - \mu)^2}{2\sigma^2}\right) \tag{7}$$

where $\mu$ and $\sigma$ are the parameters of the distribution.

In addition, as an alternative impact function, we consider the Truncated normal distribution:

$$f_i(t) = \frac{1}{\sigma} \frac{\phi(\frac{t-\mu}{\sigma})}{\Phi(\frac{\kappa-\mu}{\sigma}) - \Phi(\frac{\rho-\mu}{\sigma})}, \tag{8}$$

where $\mu$ and $\sigma$ are the parameters of the distribution which lie in $(\rho, \kappa) \in \mathbb{R}$, $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$, and $\Phi(\cdot)$ is the cumulative distribution function (cdf) $\Phi(x) = \frac{1}{2}\left(1 + \text{erf}(\frac{x}{\sqrt{2}})\right)$. Such distributions are particularly valuable for capturing the inherent variability and asymmetry in the life-cycle of a paper.

As argued in the work of Wang et al. (2013) employing an impact function based solely on the pdf yields a successful characterization of the aging of a paper but does not account for the preferential attachment connecting to papers regarding their impact at a specific time moment. For that, we also consider the approach in Wang et al. (2013) as an impact function, which is the product of the pdf and the cumulative distribution function (cdf), as:

$$f_i(t) = pdf_i(t) cdf_i(t), \tag{9}$$

We here note that Eq. (5) requires the calculation of the integral $\Lambda_{ij}(t_i, T) := \int_{t_i}^{T} \lambda_{ij}(t')\mathrm{d}t'$ which can be trivially calculated for an impact function based on Eq. (9), by applying the substitution rule for $u := cdf_i(t)$. It is also important that the cdf is of analytical form which is the case for both the Truncated normal and Log-Normal distributions we consider.

## 4 Single-Event Network Embedding by the Latent Distance Model

Our main purpose is to represent every node of a given single-event network in a low $D$-dimensional latent space ($D \ll N$) in which the pairwise distances in the embedding space should reflect various structural properties of the network, like homophily and transitivity (Nakis et al., 2023b,a). For instance, in the *Latent Distance Model* (Hoff et al., 2002a), one of the pioneering works, the probability of a link between a pair of nodes depended on the log-odds expression, $\gamma_{ij}$, as

$\alpha - \|\mathbf{z}_i - \mathbf{z}_j\|_2$ where $\{\mathbf{z}_i\}_{i\in\mathcal{V}}$ are the node embeddings, and $\alpha \in \mathbb{R}$ is the global bias term responsible for capturing the global information in the network. It has been proposed for undirected graphs but can be extended for directed networks as well by simply introducing another node representation vector $\{\mathbf{w}_i\}_{i\in\mathcal{V}}$ in order to differentiate the roles of the node as source (i.e., sender) and target (i.e., receiver). By the further inclusion of two sets of random effects $(\alpha_i, \beta_i)_{i\in\mathcal{V}}$ describing the in and out degree heterogeneity, respectively, we can define the log-odds (Bernoulli) and log-rate (Poisson) (Nakis et al., 2023b) expression as:

$$\gamma_{ij} = \alpha_i + \beta_j - \|\mathbf{z}_i - \mathbf{w}_j\|_2. \tag{10}$$

We can now combine a dynamic impact characterization function with the *Latent Distance Model*, to obtain an expression for the intensity function of the proposed *Single-Event Poisson Process*, as:

$$\lambda_{ij}(t_{ij}) = \frac{f_i(t_{ij}) \exp(\alpha_i) \exp(\beta_j)}{\exp(\|\mathbf{z}_i - \mathbf{w}_j\|_2)}. \tag{11}$$

Combining the intensity function of Eq. (11) with the log-likelihood expression of Eq. (5) yields the *Dynamic Impact Single-Event Embedding Model* (DISEE). Under such a formulation, we exploit the time information data indicating when links occur through time, so we can grasp a more detailed understanding of the evolution of networks, generate enriched node representations, and quantify a node's temporal impact on the network.

**Generative Model**: To generate Single-Event Networks defining accurate and powerful impact functions, we consider the following generative process:

1. Input: $\alpha_\lambda, \theta_\lambda, \alpha_\beta, \theta_\beta, \mu_m, \sigma_m, \alpha_s, \theta_s, \sigma_{\mathbf{z}}, \sigma_{\mathbf{w}}$
2. For each node $i \in \mathcal{V}$
   (a) $t_i \sim Uniform(0, T)$, Paper appearance times
   (b) $\lambda_i \sim Gamma(\alpha_\lambda, \theta_\lambda)$, Impact cited paper
   (c) $\beta_i \sim Gamma(\alpha_\beta, \theta_\beta)$, Degree citing paper
   (d) $\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{z}}\mathbf{I})$, Cited paper embedding
   (e) $\mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2_{\mathbf{w}}\mathbf{I})$, Citing paper embedding
   (f) $\mu_i \sim \mathcal{N}(\mu_m, \sigma^2_m)$, Param. of log-normal
   (g) $\sigma_i \sim Gamma(\alpha_s, \theta_s)$, Param. of log-normal
   (h) $f_i \leftarrow$ pdf of $LogNormal(\mu_i, \sigma_i)$
3. Relabel nodes $(1, \ldots, N)$ such that $t_1 \leq \cdots \leq t_N$
4. For each node $i \in \mathcal{V}$
   (a) $\kappa_i \sim Poisson(\lambda_i)$, Number of citations
   (b) For each node $j > i$: $p_{ij} \leftarrow \frac{(\kappa_i f_i(t_j - t_i))\beta_j}{\exp(\|\mathbf{z}_i - \mathbf{w}_j\|)}$
   (c) $\mathbf{p} \leftarrow (p_{i(i+1)}, \ldots, p_{ij}, \ldots, p_{iN}) / \sum_{k=i+1}^{N} p_{ik}$
   (d) $K \leftarrow \min(\kappa_i, N - i)$
   $(j_1, \ldots, j_K) \sim Mult(\mathbf{p}, \text{sample } K \text{ w.o. repl.})$
   (e) Add links $(i, i + j_1), \ldots, (i, i + j_K)$

We could, in Step 4, have generated the observed networks from the Single-Event Poisson Process (SE-PP); however, this would technically assign separate time points to be sampled for each citation. Instead, we draw the citations from fixed time points of publications using multinomial sampling without replacement.

**Model ablations**: We define an Impact Function Model (IFM), where only the impact function is fitted to the target nodes (cited papers) describing their link (citation) dynamics. Comparing with IFM will allow us to validate the quality of the impact characterization of DISEE. We further contrast our model to a Preferential Attachment Model (PAM) setting where the embedding dimension is set as $D = 0$, providing a quantification of the importance of including an impact function and an embedding space in DISEE. In addition, we consider a combination of an Impact Function Model with a Preferential Attachment Model, defining a Temporal Preferential Attachment Model (TPAM). Compared with the TPAM we aim to verify the importance of introducing an embedding space characterization in citation networks. Finally, we systematically contrast the performance of DISEE to conventional static latent distance modeling (LDM) (Hoff et al., 2002a) corresponding to setting the impact function to be constant $f_i(t) \propto 1$ in DISEE. The LDM is a very powerful link predictor (Nakis et al., 2022, 2023a) and contrasting its performance against DISEE will help us showcase the successful reconciliation of static latent space network embedding approaches with classical dynamic impact assessments of citation networks. In Table 1, we provide the rate formulation of each considered model ablations and the corresponding model characteristics in terms of impact characterization and definition of an embedding space.

## 5 Results and Discussion

In this section, we will evaluate how successfully DISEE reconciles traditional impact quantification approaches with latent distance modeling. We denote as DISEE the model with an impact function given as a probability density function ($f_i(t) = pdf_i(t)$), and as DISEE PA the Preferential Attachment (PA) model with an impact function based on Eq. (9). Lastly, we also consider the Fixed Impact (FI) variants of the models, where the impact functions are fixed, during training, to the empirical distribution of the citation data distributions, yielding the FI-DISEE and FI-DISEE PA models. The latter two models essentially combine an IFM model with the proposed DISEE

Table 1: DISEE model and considered model ablations.

| Model Name | IFM | PAM | TPAM | LDM | DISEE |
|---|---|---|---|---|---|
| **Rate formulation** | $\exp(\alpha_i) f_i(t)$ | $\exp(\alpha_i)\exp(\beta_j)$ | $f_i(t)\exp(\alpha_i)\exp(\beta_j)$ | $\frac{\exp(\alpha_i)\exp(\beta_j)}{\exp(\|\|\boldsymbol{z}_i-\boldsymbol{w}_j\|\|_2)}$ | $\frac{f_i(t)\exp(\alpha_i)\exp(\beta_j)}{\exp(\|\|\boldsymbol{z}_i-\boldsymbol{w}_j\|\|_2)}$ |
| **Impact** | ✓ | ✗ | ✓ | ✗ | ✓ |
| **Embedding space** | ✗ | ✗ | ✗ | ✓ | ✓ |

Table 2: Statistics of networks.

| | $\lvert\mathcal{V}_1\rvert$ | $\lvert\mathcal{V}_2\rvert$ | $\lvert\mathcal{E}\rvert$ |
|---|---|---|---|
| *Artificial* (*Art*) | 4,001 | 4,979 | 160,039 |
| *Machine Learning* (*ML*) | 22,540 | 148,703 | 526,226 |
| *Physics* (*Phys*) | 20,012 | 51,996 | 573,378 |
| *Social Science* (*SoSci*) | 12,930 | 100,402 | 288,012 |

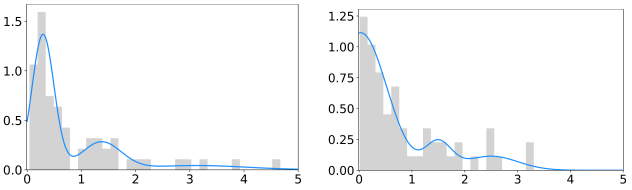
model. All experiments regarding the DISEE models have been conducted on an 8 GB NVIDIA RTX 2070 Super GPU. In addition, we adopted the Adam optimizer (Kingma and Ba, 2017) with a learning rate of lr = 0.1 and ran it for 3000 iterations. Our objective was to minimize the negative log-likelihood of the Single-Event Poisson Process, utilizing a case-control approach (Raftery et al., 2012) to achieve scalable inference. (*For extensive training and implementation details, code release, as well as, a complexity analysis please visit the supplementary.*)

We test the proposed approach's effectiveness in the link prediction task by comparing it to the classical LDM which is not time-aware and does not quantify temporal impact, as well as, against four prominent baselines. We also consider multiple model ablations that are either able to characterize a node's impact or account for GRL, i.e. define node embeddings, but not both. For the task of link prediction, we remove 20% of network links and we sample an equal amount of non-edges as negative samples and construct the test set. The link removal is designed in such a way that the residual network stays connected. For the evaluation, we consider the Precision-Recall Area Under Curve (AUC-PR) score, as it is a metric not sensitive to the class imbalance between links and non-links (*Receiver Operator Characteristic scores are provided in the supplementary*). We then continue by evaluating the quality of impact expression of DISEE by visually presenting the inferred impact functions and comparing them against the FI-DISEE model. Finally, we visualize the model's learned temporal space representing the target papers, accounting for their temporal impact in terms of their mass at a specific time point, and characterizing the different papers' lifespans.

**Datasets**: In our experiments, we employ three real citation networks. Specifically, we use the OpenAlex dataset (Priem et al., 2022), exploring highly impactful scientific domains such as (i) *Machine Learning*, (ii) *Physics*, and (iii) *Social Science*. Lastly, we also consider an artificial network generated based on the generative process as described in Section 4. In order to be able to characterize scientific impact, we consider papers that have been cited at least ten times while we assign zero mass to papers with less than ten incoming links. This yields a directed bipartite structure where target nodes have at least ten citations. Analytically, the network statistics are given in Table 2 where $|\mathcal{V}_1|$ is the number of target nodes, $|\mathcal{V}_2|$ the number of source nodes, and $|\mathcal{E}|$: the total number of links. (*For more details see supplementary.*)

**Baselines**: In addition to the model ablations stated in Table 1, we employed four baselines relying on different encoding strategies. (i) NODE2VEC (Grover and Leskovec, 2016) is a well-known random walk-based approach that learns node embeddings by utilizing the co-occurrence frequencies of nodes within the generated node sequences. (ii) CTDNE (Nguyen et al., 2018) can be considered as its adaptation for temporal networks. It performs random walks based on the temporal characteristics of the network thereby incorporating the time information of the links. (iii) VERSE (Tsitsulin et al., 2018) aims to extract latent representations reflecting Personalized Page Rank (PPR) similarity among nodes. Finally, we have employed the (iv) NETSMF model (Qiu et al., 2019) which learns embeddings by decomposing a designed matrix representing the high and low-order node interactions. (*For more details see supplementary.*)

**Link prediction**: We here compare the results of DISEE variants against the different model ablations and baselines in terms of their AUC-PR scores, as presented in Table 3. Scores for our models are presented as the mean value of three independent runs of the Adam optimizer (error bars were found in the $10^{-4}$ scale and thus omitted). We here provide the scores for the DISEE model and variants under the LOG-NORMAL distribution. Similar results are obtained while also considering the TRUNCATED normal distribution, and they are provided in the supplementary. We here observe, that the best performance is achieved




(a) Artificial paper # 1. (b) Artificial paper # 2.

Figure 3: *Artificial*: Comparison of the inferred impact functions, generated by a DISEE Mixture Model with three TRUNCATED normal distributions, to the true citation histogram of two papers from the *Art* dataset.

Table 3: AUC-PR scores over four networks. (Bold numbers denote the best-performing model.)

| | *Art* | | | | *ML* | | | | *Phys* | | | | *SoSci* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dimension ($D$) | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 |
| IFM | | | .882 | | | | .610 | | | | .639 | | | | .572 | |
| PAM | | | .520 | | | | .810 | | | | .838 | | | | .796 | |
| TPAM | | | .903 | | | | .806 | | | | .836 | | | | .790 | |
| NODE2VEC | .593 | .810 | .767 | .828 | .584 | .799 | .873 | .966 | .592 | .811 | .896 | .974 | .566 | .800 | .897 | .964 |
| VERSE | .606 | .816 | .774 | .831 | .760 | .911 | .946 | .981 | .780 | .882 | .929 | .977 | .800 | .919 | .949 | **.976** |
| CTDNE | .651 | .711 | .768 | .810 | .653 | .768 | .830 | .928 | .636 | .777 | .835 | .937 | .677 | .805 | .859 | .924 |
| NETSMF | .749 | .834 | .834 | .847 | .713 | .750 | .818 | .913 | .750 | .790 | .867 | .936 | .772 | .836 | .836 | .923 |
| LDM | .943 | .953 | .958 | .955 | **.953** | .969 | .976 | **.983** | **.940** | **.963** | **.973** | **.987** | **.943** | .956 | 0.963 | .975 |
| DISEE PA | .947 | .955 | .957 | .956 | .944 | .967 | .973 | .979 | .926 | .958 | .968 | .981 | .930 | .958 | .965 | .971 |
| FI-DISEE PA | .944 | .954 | .956 | .956 | .941 | .966 | .973 | .978 | .923 | .956 | .967 | .980 | .929 | .956 | .963 | .970 |
| DISEE | **.955** | **.962** | **.963** | **.961** | **.953** | **.970** | **.977** | .982 | .937 | **.963** | **.973** | .984 | .939 | **.962** | .968 | **.976** |
| FI-DISEE | **.955** | **.962** | **.963** | **.961** | .952 | **.970** | **.977** | .982 | .934 | .962 | .972 | .983 | .940 | **.962** | **.969** | .975 |

by model specifications that define an embedding space, i.e. the DISEE variants and LDM models. The Preferential Attachment Models in both the static (PAM) and temporal (TPAM) versions are characterized by an approximately 15% decrease in their link prediction score. This highlights the importance and benefits of the predictive performance an embedding space provides. Contrasting the performance of DISEE against the LDM, we witness almost identical scores, verifying that DISEE successfully inherited the link prediction power of the LDM. Comparing now to the baselines, we here observe that all of the DISEE variants and the LDM are characterized by mostly a significantly higher performance. Notably, the DISEE PA and FI-DISEE PA models show a small decrease in the performance. Lastly, fixing the impact functions in the FI-DISEE and FI-DISEE PA models does not have an effect on the model performance. Contrasting the DISEE and LDM models on the *Art* network, we observe a higher performance from the former. However, this advantage diminishes as the dimensionality increases, attributed to the expression and conveyance of temporal information by the latent space.

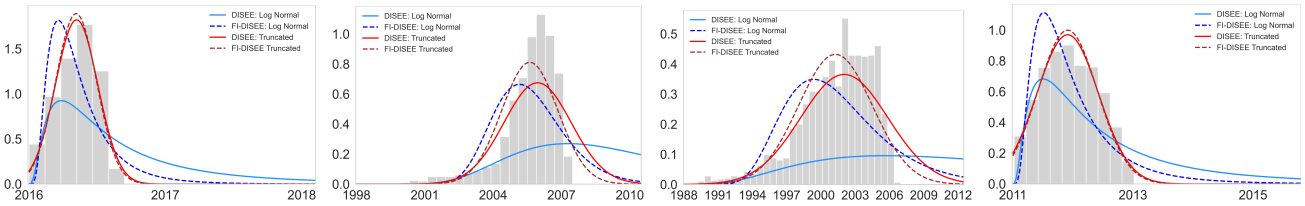
**Impact quantification and space visualization**: We now continue by addressing the quality of paper impact characterization based on a target paper's incoming citation dynamics. In Figure 4, we provide the inferred impact functions of DISEE and FI-DISEE, under the LOG-NORMAL and TRUNCATED normal distributions. We further show the true impact dynamics through the citation histogram for each one of the corresponding papers. For the TRUNCATED normal case we witness a general agreement between the DISEE and FI-DISEE models, describing well the ground truth citation pattern. For the LOG-NORMAL case, we witness a similar impact shape if the paper lifespan is short in terms of its incoming citation pattern. For larger lifespans DISEE defines a much larger standard deviation than FI-DISEE returning heavier tails. In general, when compared to the true citation histogram both models using the LOG-NORMAL distribution provide much heavier tails when the paper lifespan exceeds the 2-year threshold. Nevertheless, the LOG-NORMAL heavier tails may be more appropriate for future impact predictions (as papers stay "alive" longer).

Finally, we here provide embedding space visualizations of the target papers, accounting for their temporal impact in terms of their mass at a specific time point. Analytically, Figures 5 and 6 (a)-(c) show the evolution of the embedding space for the domain of *Machine Learning* and *Social Sciences* from the year 1988 until 2023. We here observe how papers are published in a specific year and after they accumulate a specific amount of impact/mass they perish in the next years/snapshots of the network. Lastly, Figure 6 (d) shows the present day status of *Physics* based on the years passed from the publication for each paper.

**Broader scope:** Our proposed approach has a general aim of providing a principled analysis of single-event dynamic networks, reconciling impact characterization with graph representation learning. We highlight here that the impact function can further be extended to kernel density estimation approaches, non-parametric methods, and mixture models. As an example, in Figure 3, we provide an example of a mixture-model for the *Art* network. We witness how more advanced methods can accurately describe more complex citation patterns. This showcases the generalization power of DISEE which does not solely rely on a specific choice of impact function.

## 6 Conclusion

We have proposed the Dynamic Impact Single-Event Embedding Model (DISEE), a reconciliation between




(a) Deep Residual Learning for Image Recognition (He et al., 2015).
(b) Gradient-based learning applied to document recognition (Lecun et al., 1998).
(c) Structural equation modeling in practice (Anderson and Gerbing, 1988).
(d) LIBSVM: A library for support vector machines (Chang and Lin, 2011).

Figure 4: *Machine Learning*: DISEE and FI-DISEE models TRUNCATED normal, and LOG-NORMAL inferred impact function visualizations compared to the true citation histogram for four highly cited *ML* papers.


(a) 1998 (b) 2008 (c) 2018 (d) 2023

Figure 5: *Machine Learning*: DISEE 2-dimensional embedding space LOG-NORMAL yearly evolution. Node sizes are based on each paper's mass, $f_i(t)\exp(\alpha_i)$. Nodes are color-coded based on their publication year.


(a) *SoSci* 1998 (b) *SoSci* 2008 (c) *SoSci* 2023 (d) *Phys* 2023

Figure 6: DISEE 2-dimensional embedding space LOG-NORMAL: *Social Science* (a)-(c) field evolution throughout the years with nodes color-coded based on their publication year. *Physics* present day status (d) with nodes color-coded based on years passed from their publication. Node sizes are based on each paper's mass, $f_i(t)\exp(\alpha_i)$.

traditional impact quantification approaches with a Latent Distance Model (LDM). We have focused on Single-Event Networks (SENs), and more specifically in citation networks, where we to the best of our knowledge for the first time derived and explored the Single-Event Poisson Process (SE-PP). Such a process defines an appropriate likelihood allowing for a principled analysis of SENs. In order to define powerful ultra-low dimensional network embeddings, we turned to the representation power of the LDM. We introduced impact functions parameterized through an appropriate probability density function, such as the log-normal distribution. Through extensive experiments, we showed that the DISEE inherited the link prediction performance of the powerful LDM. Furthermore, we showed how the temporal impact successfully characterized citation patterns, showcasing that the DISEE model successfully reconciles powerful embedding approaches with citation dynamics impact characterization. Finally, visualizations of the embedding space clearly depict the lifecycle of target papers, highlighting their emergence, impact duration, and eventual decline as science evolves over the years.

An important aspect of SciSci is the future impact prediction for newly published/unseen (during training) papers. Predicting such citation patterns requires an inductive setting that can be obtained by extending our proposed DISEE model defining paper embeddings via Graph Neural Networks based on the citing pattern (and potentially the inclusion of node features). We leave such formulations as a future research direction.

## Acknowledgements

We would like to express sincere appreciation and thank the reviewers for their constructive feedback and their insightful comments. We thank the Independent Research Fund Denmark for supporting this work [grant number: 0136-00315B].

## References

J. Anderson and D. Gerbing. Structural equation modeling in practice: A review and recommended two-step approach. *Psychological bulletin*, 103(3):411–423, May 1988. ISSN 0033-2909. doi: 10.1037/0033-2909.103.3.411.

M. Arastuie, S. Paul, and K. Xu. CHIP: A hawkes process model for continuous-time networks with scalable and consistent estimation. In *NeurIPS*, volume 33, pages 16983–16996, 2020.

A.-L. Barabási, C. Song, and D. Wang. Handful of papers dominates citation. *Nature*, 491(7422):40–41, 2012.

H. S. Bhat, L.-H. Huang, S. Rodriguez, R. Dale, and E. Heit. Citation prediction using diverse features. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 589–596, 2015. doi: 10.1109/ICDMW.2015.131.

C. Blundell, J. Beck, and K. A. Heller. Modelling reciprocating relationships with hawkes processes. In *NeurIPS*, volume 25, 2012.

G. J. Borjas and K. B. Doran. Which peers matter? the relative impacts of collaborators, colleagues, and competitors. *Review of economics and statistics*, 97 (5):1104–1117, 2015.

A. Çelikkanat, N. Nakis, and M. Mørup. Piecewise-velocity model for learning continuous-time dynamic node representations. *arXiv preprint arXiv:2212.12345*, 2022.

A. Celikkanat, N. Nakis, and M. Mørup. Continuous-time graph representation with sequential survival process, 2023.

C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3), may 2011. ISSN 2157-6904. doi: 10.1145/1961189.1961199. URL `https://doi.org/10.1145/1961189.1961199`.

M. Cokol, I. Iossifov, C. Weinreb, and A. Rzhetsky. Emergent behavior of growing knowledge about molecular interactions. *Nature Biotechnology*, 23: 1243–1247, Nov. 2005. doi: 10.1038/nbt1005-1243.

N. R. Council et al. Enhancing the effectiveness of team science. 2015.

F. Davletov, A. S. Aydin, and A. Cakmak. High impact academic paper prediction using temporal and topological features. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, page 491–498, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450325981. doi: 10.1145/2661829.2662066. URL `https://doi.org/10.1145/2661829.2662066`.

S. Delattre, N. Fournier, and M. Hoffmann. Hawkes processes on large networks. *The Annals of Applied Probability*, 26(1):216 – 261, 2016.

J. Duch, X. H. T. Zeng, M. Sales-Pardo, F. Radicchi, S. Otis, T. K. Woodruff, and L. A. Nunes Amaral. The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PloS one*, 7(12):e51332, 2012.

D. Durante and D. Dunson. Bayesian Logistic Gaussian Process Models for Dynamic Networks. In *AISTATS*, volume 33, pages 194–201, 2014.

D. Durante and D. B. Dunson. Locally adaptive dynamic networks. *The Annals of Applied Statistics*, 10(4):2203–2232, 2016.

Y.-H. Eom and S. Fortunato. Characterizing and modeling citation dynamics. *PLOS ONE*, 6(9):1–7, 09 2011. doi: 10.1371/journal.pone.0024926. URL `https://doi.org/10.1371/journal.pone.0024926`.

X. Fan, B. Li, F. Zhou, and S. SIsson. Continuous-time edge modelling using non-parametric point processes. *NeurIPS*, 34:2319–2330, 2021.

S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A.-L. Barabási. Science of science. *Science*, 359(6379):eaao0185, 2018. doi: 10.1126/science.aao0185.

J. G. Foster, A. Rzhetsky, and J. A. Evans. Tradition and Innovation in Scientists' Research Strategies. *American Sociological Review*, 80(5):875–908, 2015. doi: 10.1177/0003122415601618.

R. Freeman, E. Weinstein, E. Marincola, J. Rosenbaum, and F. Solomon. Competition and careers in biosciences, 2001.

E. Garfield. Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178(4060):471–479, 1972.

M. L. Goldstein, S. A. Morris, and G. G. Yen. Group-based yule model for bipartite author-paper networks. *Phys. Rev. E*, 71:026108, Feb 2005. doi: 10.1103/PhysRevE.71.026108.

M. Golosovsky and S. Solomon. Runaway events dominate the heavy tail of citation distributions. *The European Physical Journal Special Topics*, 205(1): 303–311, 2012.

J. A. González, F. J. Rodríguez-Cortés, O. Cronie, and J. Mateu. Spatio-temporal point process statistics: A review. *Spatial Statistics*, 18:505–544, 2016. ISSN 2211-6753. doi: https://doi.org/10.1016/j.spasta.2016.10.002.

G. Grimmett and D. Stirzaker. *Probability and Random Processes*. Oxford University Press, 2001.

A. Grover and J. Leskovec. Node2Vec: Scalable feature learning for networks. In *KDD*, pages 855–864, 2016.

S. Gualdi, M. Medo, and Y.-C. Zhang. Influence, originality and similarity in directed acyclic graphs. *Europhysics Letters*, 96(1):18004, sep 2011a. doi: 10.1209/0295-5075/96/18004.

S. Gualdi, C. H. Yeung, and Y.-C. Zhang. Tracing the evolution of physics on the backbone of citation networks. *Phys. Rev. E*, 84:046104, Oct 2011b. doi: 10.1103/PhysRevE.84.046104.

W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *NIPS*, 2017.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition, 2015.

C. Heaukulani and Z. Ghahramani. Dynamic probabilistic models for latent feature propagation in social networks. In *ICML*, pages 275–283, 2013.

T. Herlau, M. Mørup, and M. Schmidt. Modeling temporal evolution and multiscale structure in networks. In *ICML*, pages 960–968, 2013.

J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, 102(46):16569–16572, 2005.

P. D. Hoff. Bilinear mixed-effects models for dyadic data. *JASA*, 100(469):286–295, 2005.

P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *JASA*, 97(460):1090–1098, 2002a.

P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *JASA*, 97(460):1090–1098, 2002b.

A. Ibáñez, P. Larranaga, and C. Bielza. Predicting citation count of bioinformatics papers within four years of publication. *Bioinformatics (Oxford, England)*, 25:3303–9, 10 2009. doi: 10.1093/bioinformatics/btp585.

K. Ishiguro, T. Iwata, N. Ueda, and J. Tenenbaum. Dynamic infinite relational model for time-varying relational data analysis. *NeurIPS*, 23, 2010.

S. M. Kazemi, R. Goel, K. Jain, I. Kobyzev, A. Sethi, P. Forsyth, and P. Poupart. Representation learning for dynamic graphs: A survey. *JMLR*, 21(70):1–73, 2020.

B. Kim, K. H. Lee, L. Xue, and X. Niu. A review of dynamic network models with latent variables. *Statistics surveys*, 12:105, 2018.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks, 2017.

P. N. Krivitsky, M. S. Handcock, A. E. Raftery, and P. D. Hoff. Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social Networks*, 31 (3):204 – 213, 2009.

V. Larivière, C. Ni, Y. Gingras, B. Cronin, and C. R. Sugimoto. Bibliometrics: Global gender disparities in science. *Nature*, 504(7479):211–213, 2013.

V. Larivière, Y. Gingras, C. R. Sugimoto, and A. Tsou. Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7):1323–1332, 2015.

Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. doi: 10.1109/5.726791.

J. Leskovec, J. Kleinberg, and C. 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*, KDD '05, page 177–187, New York, NY, USA, 2005. Association for Computing Machinery. ISBN 159593135X. doi: 10.1145/1081870.1081893. URL `https://doi.org/10.1145/1081870.1081893`.

A. Letchford, H. Moat, and T. Preis. The advantage of short paper titles. *Royal Society Open Science*, 2: 150266, 08 2015. doi: 10.1098/rsos.150266.

A. Letchford, T. Preis, and H. S. Moat. The advantage of simple paper abstracts. *Journal of Informetrics*, 10(1):1–8, 2016. ISSN 1751-1577. doi: https://doi.org/10.1016/j.joi.2015.11.001.

S. Luan, C. Hua, Q. Lu, J. Zhu, M. Zhao, S. Zhang, X.-W. Chang, and D. Precup. Is heterophily a real nightmare for graph neural networks to do node classification?, 2021.

S. Milojević. Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences*, 111(11):3984–3989, 2014.

N. Nakis, A. Çelikkanat, S. L. Jørgensen, and M. Mørup. A hierarchical block distance model for ultra low-dimensional graph representations. 2022.

N. Nakis, A. Çelikkanat, and M. Mørup. HM-LDM: A hybrid-membership latent distance model. In *CNA XI*, pages 350–363. Springer International Publishing, 2023a.

N. Nakis, A. Çelikkanat, S. Lehmann, and M. Mørup. A hierarchical block distance model for ultra low-dimensional graph representations. *IEEE Transactions on Knowledge and Data Engineering*, pages 1–14, 2023b. doi: 10.1109/TKDE.2023.3304344.

M. E. Newman. The structure of scientific collaboration networks. *Proceedings of the national academy of sciences*, 98(2):404–409, 2001a.

M. E. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical review E*, 64(1):016131, 2001b.

M. E. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):016132, 2001c.

G. H. Nguyen, J. B. Lee, R. A. Rossi, N. K. Ahmed, E. Koh, and S. Kim. Continuous-time dynamic network embeddings. In *TheWebConf*, page 969–976, 2018.

M. Nickel and D. Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.

M. Nickel and D. Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International Conference on Machine Learning*, pages 3779–3788. PMLR, 2018.

R. V. Noorden. Interdisciplinary research by the numbers. *Nature*, 525:306–307, 2015.

B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *KDD*, page 701–710, 2014.

A. M. Petersen, M. Riccaboni, H. E. Stanley, and F. Pammolli. Persistence and uncertainty in the academic career. *Proceedings of the National Academy of Sciences*, 109(14):5213–5218, 2012.

J. Priem, H. Piwowar, and R. Orr. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts, 2022.

J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang. Network embedding as matrix factorization: Unifying DeepWalk, LINE, PTE, and Node2Vec. In *WSDM*, 2018.

J. Qiu, Y. Dong, H. Ma, J. Li, C. Wang, K. Wang, and J. Tang. Netsmf: Large-scale network embedding as sparse matrix factorization. In *The World Wide Web Conference*, pages 1509–1520, 2019.

F. Radicchi, S. Fortunato, and C. Castellano. Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, 105(45):17268–17272, 2008.

A. E. Raftery, X. Niu, P. D. Hoff, and K. Y. Yeung. Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919, 2012.

S. Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B*, 4(2):131–134, aug 1998. doi: 10.1007/s100510050359.

S. Redner. Citation statistics from 110 years of physical review. *Physics Today*, 58, 06 2005. doi: 10.1063/1.1996475.

E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. M. Bronstein. Temporal graph networks for deep learning on dynamic graphs. *ICML 2020 Workshop*, 2020.

P. Sarkar and A. Moore. Dynamic social network analysis using latent space models. In Y. Weiss, B. Schölkopf, and J. Platt, editors, *NeurIPS*, volume 18, 2005.

P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad. Collective classification in network data. *AI magazine*, 2008.

D. K. Sewell and Y. Chen. Latent space models for dynamic networks. *JASA*, 110(512):1646–1657, 2015.

H. Shen, D. Wang, C. Song, and A. Barabási. Modeling and predicting popularity dynamics via reinforced poisson processes. *CoRR*, abs/1401.0778, 2014. URL `http://arxiv.org/abs/1401.0778`.

N. Shibata, Y. Kajikawa, and I. Sakata. Link prediction in citation networks. *Journal of the American Society for Information Science and Technology*, 63(1):78–85, 2012. doi: https://doi.org/10.1002/asi.21664.

M. Singh, V. Patidar, S. Kumar, T. Chakraborty, A. Mukherjee, and P. Goyal. The role of citation context in predicting long-term citation profiles: An experimental study based on a massive bibliographic text dataset. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, page 1271–1280, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450337946. doi: 10.1145/2806416.2806566. URL `https://doi.org/10.1145/2806416.2806566`.

P. E. Stephan, R. Veugelers, and J. Wang. Reviewers are blinkered by bibliometrics. *Nature*, 544:411–412, 2017.

R. L. Streit. *Poisson point processes: imaging, tracking, and sensing*. Springer Science & Business Media, 2010.

R. Trivedi, M. Farajtabar, P. Biswal, and H. Zha. Dyrep: Learning representations over dynamic graphs. In *ICLR*, 2019.

A. Tsitsulin, D. Mottin, P. Karras, and E. Müller. Verse: Versatile graph embeddings from similarity measures. In *Proceedings of the 2018 world wide web conference*, pages 539–548, 2018.

B. Uzzi, S. Mukherjee, M. Stringer, and B. Jones. Atypical combinations and scientific impact. *Science (New York, N.Y.)*, 342:468–72, 10 2013. doi: 10.1126/science.1240474.

M. Wallace, V. Larivière, and Y. Gingras. Modeling a century of citation distributions. *Journal of Informetrics*, 3, 11 2008. doi: 10.1016/j.joi.2009.03.010.

L. Waltman. A review of the literature on citation impact indicators. *Journal of informetrics*, 10(2): 365–391, 2016a.

L. Waltman. A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2):365–391, 2016b. ISSN 1751-1577. doi: https://doi.org/10.1016/j.joi.2016.02.007.

L. Waltman, N. J. van Eck, and A. F. van Raan. Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology*, 63(1):72–77, 2012.

D. Wang, C. Song, and A.-L. Barabási. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013. doi: 10.1126/science.1237825.

J. Wang, Y. Mei, and D. Hicks. Comment on quantifying long-term scientific impact. *Science*, 345(6193): 149–149, 2014. doi: 10.1126/science.1248770.

J. Wang, R. Veugelers, and P. Stephan. Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8):1416–1436, 2017.

L. Wu, D. Wang, and J. A. Evans. Large teams have developed science and technology; small teams have disrupted it. *arXiv preprint arXiv:1709.02445*, 2017.

Y. Wu, T. Z. Fu, and D. M. Chiu. Generalized preferential attachment considering aging. *Journal of Informetrics*, 8(3):650–658, 2014. ISSN 1751-1577. doi: https://doi.org/10.1016/j.joi.2014.06.002.

S. Xiao, J. Yan, C. Li, B. Jin, X. Wang, X. Yang, S. M. Chu, and H. Zhu. On modeling and predicting individual paper citation count over time. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, IJCAI'16, page 2676–2682. AAAI Press, 2016. ISBN 9781577357704.

G. Xue, M. Zhong, J. Li, J. Chen, C. Zhai, and R. Kong. Dynamic network embedding survey. *Neurocomputing*, 472:212–223, 2022.

R. Yan, J. Tang, X. Liu, D. Shan, and X. Li. Citation count prediction: Learning to estimate future citations for literature. pages 1247–1252, 10 2011. doi: 10.1145/2063576.2063757.

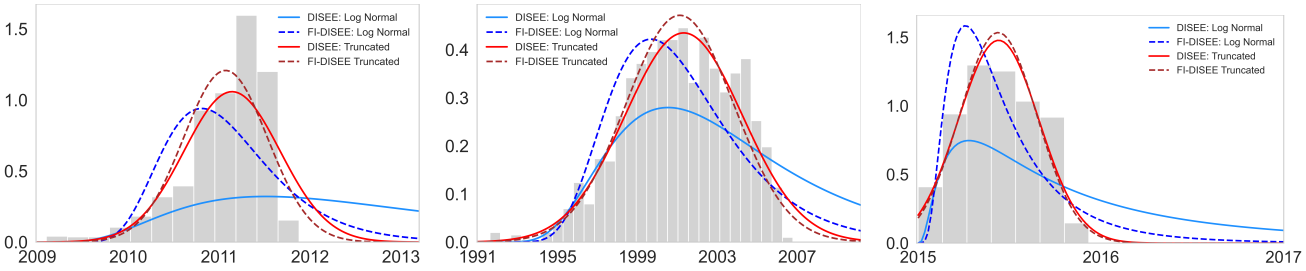
P. Zhang, J. Wang, X. Li, M. Li, Z. Di, and Y. Fan. Clustering coefficient and community structure of bipartite networks. *Physica A: Statistical Mechanics and its Applications*, 387 (27):6869–6875, 2008. ISSN 0378-4371. doi: https://doi.org/10.1016/j.physa.2008.09.006.

Y. Zuo, G. Liu, H. Lin, J. Guo, X. Hu, and J. Wu. Embedding temporal network via neighborhood formation. In *KDD*, page 2857–2866, 2018.

# Supplementary Materials

## 1 Impact Functions

When comparing the TRUNCATED normal to the LOG-NORMAL distribution, it is evident that the latter possess heavier tails, offering a superior description of the papers' potential longevity. However, a significant advantage of the TRUNCATED normal distribution is its invariance to the scale of the time resolution, a feature not shared by the LOG-NORMAL distribution. As a result, by employing an impact based on the TRUNCATED normal, the time resolution of the dataset becomes inconsequential, leading to a straightforward impact inference. In Figure 1, we provide the TRUNCATED normal impact function for DISEE in comparison to the Fixed Impact (FI) mode, FI-DISEE, where the impact function is frozen solely to follow to link structure for each node. We here observe that fixing the impact in the DISEE does not affect paper with relatively small longevity (few years) but leads to heavier tails when the paper longevity increases. In addition, Figure 2 shows the evolution of the embedding space for the cited papers throughout the years. Lastly, Table 1 and Table 2, provide the AUC-PR and AUC-ROC score comparison between different DISEE models using the LOG-NORMAL and the TRUNCATED normal distributions as impact functions.




(a) ImageNet: A large-scale hierarchical image database (Deng et al., 2009).

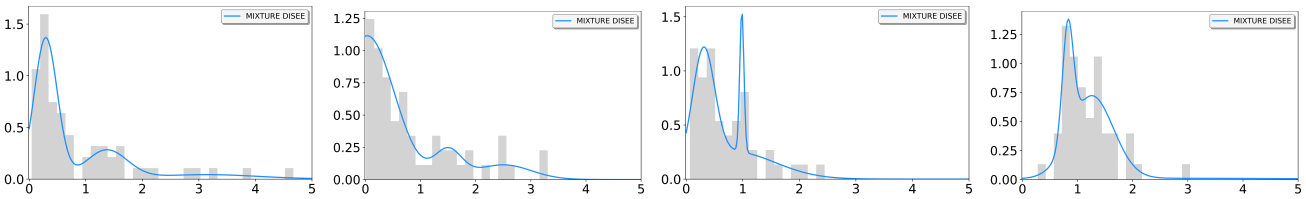
(b) Firm Resources and Sustained Competitive Advantage (Barney, 1991).

(c) Going Deeper with Convolutions (Szegedy et al., 2014).

Figure 1: *Machine Learning*: DISEE and FI-DISEE models TRUNCATED normal inferred impact function visualizations compared to the true citation histogram for four popular papers with different citation dynamics.


(a) 1998 (b) 2008 (c) 2018 (d) 2023

Figure 2: *Machine Learning*: DISEE under the TRUNCATED normal distribution 2-dimensional embedding space evolution throughout the years. Node sizes are based on each paper's mass, $f_i(t)\exp(\alpha_i)$. Nodes are color-coded based on their publication year.




(a) Artificial paper # 1. (b) Artificial paper # 2. (c) Artificial paper # 3. (d) Artificial paper # 4.

Figure 3: *Artificial*: Comparison of the inferred impact functions, generated by a DISEE Mixture Model with three TRUNCATED normal distributions, to the true citation histogram of four papers from the *Art* dataset. Each paper exhibited different citation dynamics.

### 1.1 Impact Characterization via Mixture Models

We can easily extend the impact function to the case of mixture models which can potentially allow for characterizing more complicated impact structures. A small example can be found for the *Artificial* dataset in Figure 3.

## 2 Datasets

In our experiments, we employ three real citation networks. Specifically, we use the OpenAlex API (Priem et al., 2022), exploring highly impactful scientific domains such as (i) *Machine Learning*, (ii) *Physics*, and (iii) *Social Science*. In order to be able to characterize scientific impact, we initially consider papers that have been cited at least ten times, defining three directed networks. Since there is no guarantee that such a filtering approach will define a network made up of nodes with a minimum degree of ten citations, we define a zero mass for the papers (target nodes) that survive the initial thresholding and have less than ten citations. This yields a directed bipartite structure where target nodes have at least ten citations.

## 3 Link Prediction Setting

For the link prediction task, we remove 20% of network links sample an equal amount of non-edges as negative samples, and construct the test set. Notably, these negative samples are sampled in a time-aware manner, meaning that we consider only pairs that are possibly to exist as missing links in the network (i.e., we do not consider node pairs where missing citations refer to papers citing future papers, as the target paper did not exist the time

Table 1: COMPARISON OF AUC-PR SCORES OVER THREE NETWORKS FOR TWO IMPACT FUNCTIONS.

| | | *ML* | | | | *Phys* | | | | *SoSci* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dimension ($D$) | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 |
| LOG-NORMAL | MODEL | .953 | .970 | .977 | .982 | .937 | .963 | .973 | .984 | .939 | .962 | .968 | .976 |
| | FI-MODEL | .952 | .970 | .977 | .982 | .934 | .962 | .972 | .983 | .940 | .962 | .969 | .975 |
| TRUNCATED | MODEL | .952 | .969 | .977 | .982 | .936 | .963 | .972 | .983 | .940 | .962 | .968 | .975 |
| | FI-MODEL | .945 | .970 | .977 | .982 | .935 | .962 | .972 | .983 | .939 | .962 | .968 | .975 |

Table 2: Comparison of AUC-ROC scores over three networks for two impact functions.

| | | *ML* | | | | *Phys* | | | | *SoSci* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Dimension ($D$) | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 |
| Log-Normal | Model | .952 | .969 | .977 | .982 | .936 | .962 | .972 | .984 | .932 | .959 | .964 | .972 |
| | FI-Model | .951 | .968 | .976 | .981 | .933 | .961 | .971 | .983 | .933 | .958 | .966 | .972 |
| Truncated | Model | .952 | .967 | .976 | .981 | .934 | .961 | .971 | .983 | .931 | .957 | .964 | .971 |
| | FI-Model | .948 | .969 | .975 | .981 | .933 | .960 | .970 | .982 | .930 | .957 | .964 | .971 |

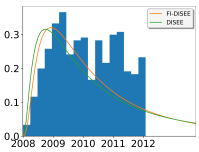
when the source paper was published). The link removal is designed in such a way that the residual network stays connected. Analytically, for each network, we do not consider removing links that make up the minimum spanning tree of the graph. For the evaluation, we consider (in the main paper) the Precision-Recall and (in supplementary) the Receiver Operator Characteristic Area Under Curve scores, as these are metrics not sensitive to the class imbalance between links and non-links.

### 3.1 Receiver Operator Characteristic scores

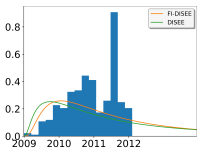
We here provide the Receiver Operator Characteristic Area Under Curve scores for the task of link prediction, as an additional evaluation metric. The results are given in Table 3:

## 4 Additional single-event network experiments:

We here consider a user-item review graph, as such networks can be argued as another prominent example of single-event networks. The major difference when comparing with citation networks now is the absence of the the time-causal structure. Specifically we use a fine food reviews network (McAuley and Leskovec, 2013) where for simplicity we consider a link between a user and a product if a review exists, while defining the temporal impact for each product based on the accumulation of the incoming links (reviews).




(a) Product # 1 Impact




(b) Product # 2 Impact

Table 4: AUC-PR scores over the *Finer-Foods* review network where 20% of the total links have been removed.

| | *LDM* | | | | *FI-DISEE* | | | | *DISEE* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dimension ($D$) | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 |
| Performance | .945 | .956 | .966 | .978 | .942 | .957 | .966 | .978 | .942 | .958 | .966 | .978 |

## 5 Training Details

For inference, we utilize the Adam optimizer (Kingma and Ba, 2017) with a learning rate of $\mathrm{lr} = 0.1$ and iterate this process for 3000 iterations. Our objective is to minimize the negative log-likelihood of the Single-Event Poisson Process, utilizing a case-control approach (Raftery et al., 2012) to achieve scalable inference.

Table 3: AUC-ROC scores over four networks. (Bold numbers denote the best-performing model.)

| | *Art* | | | | *ML* | | | | *Phys* | | | | *SoSci* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dimension ($D$) | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 | 1 | 2 | 3 | 8 |
| IFM | .906 | | | | .645 | | | | .667 | | | | 618 | | | |
| PAM | .532 | | | | .810 | | | | .838 | | | | .796 | | | |
| TPAM | .932 | | | | .814 | | | | .839 | | | | .799 | | | |
| NODE2VEC | .634 | .846 | .806 | .855 | .638 | .839 | .891 | .966 | .638 | .850 | .919 | .976 | .615 | .861 | .911 | .963 |
| VERSE | .632 | .858 | .801 | .865 | .749 | .925 | .956 | **.983** | .779 | .898 | .942 | .981 | .786 | .924 | .953 | **.977** |
| CTDNE | .710 | .742 | .802 | .827 | .650 | .792 | .835 | .925 | .632 | .794 | .837 | .937 | .649 | .792 | .835 | .925 |
| NETSMF | .787 | .868 | .867 | .876 | .753 | .794 | .839 | .916 | .758 | .817 | .885 | .939 | .772 | .835 | .845 | .925 |
| LDM | .948 | .957 | .963 | .960 | .951 | **.969** | .976 | .982 | **.937** | **.962** | **.972** | **.987** | **.935** | **.959** | .964 | .970 |
| DISEE PA | .954 | .961 | .962 | .961 | .946 | .966 | .973 | .978 | .927 | .957 | .968 | .981 | .925 | .954 | .962 | .967 |
| FI-DISEE PA | .951 | .959 | .962 | .960 | .942 | .965 | .972 | .978 | .924 | .955 | .967 | .980 | .924 | .953 | .960 | .967 |
| DISEE | **.958** | **.965** | **.966** | **.964** | **.952** | **.969** | **.977** | .982 | .936 | **.962** | **.972** | .984 | .932 | **.959** | .964 | .972 |
| FI-DISEE | **.958** | .964 | **.966** | **.964** | .951 | .968 | .976 | .981 | .933 | .961 | .971 | .983 | .933 | .958 | **.966** | .972 |

### 5.1 Sequential Learning and Parameter Initialization

Minimizing the negative log-likelihood of the Single-Event Poisson Process presents a challenging task due to the highly non-convex nature of the problem. To navigate this complexity and aid the model in converging to a quality solution, we incorporate a sequential learning approach. Specifically:

- Phase 1: Iteration 0 to 500, the model is trained to learn the parameters of the impact function, focusing exclusively on optimization over the citation patterns (link structure).
- Phase 2: Iteration 501 to 1000 the model shifts its focus to learning the scales of the random effects. During this phase, the full Single-Event Poisson Process negative log-likelihood is optimized. Concurrently, the parameters of the impact function and latent embedding are detached from the optimizer.
- Phase 3: Iteration 1001 to 3000 all parameters are incorporated into the optimizer, marking the commencement of the comprehensive learning process. During this stage, the model is fine-tuned, optimizing all parameters collectively to achieve enhanced performance and accuracy.

We initiate the embedding vectors $\{\boldsymbol{Z}, \boldsymbol{W}\}$ using a Singular-Value Decomposition (SVD) of the adjacency matrix. Concurrently, the random effects $\{\alpha_i, \beta_i\}_{i \in \mathcal{V}}$ are initialized following a Normal distribution with zero mean and unit variance. Additionally, we set the impact function parameter according to the empirical mean and variance derived from the citation pattern (link structure) of each paper.

## 6 Compexity Analysis

With DISEE being a distance model, it scales prohibitively as $\mathcal{O}(N^2)$ since the all-pairs distance matrix needs to be calculated. In order to scale the analysis to large-scale networks, we adopt an unbiased estimation of the log-likelihood similar to a case-control approach (Raftery et al., 2012). In our formulation, we calculate the log-likelihood as:

$$\log p_{ij}(\mathcal{G}|\Omega) = \sum_{j:y_{ij}=1} \left( y_{ij} \log \left( \lambda_{ij}(t^*_{ij}) \right) - \log(1 + \int_{t_i}^{T} \lambda_{ij}(t')dt') \right) + \sum_{j:y_{ij}=0} - \log \left( 1 + \int_{t_i}^{T} \lambda_{ij}(t')dt' \right) = l_1 + l_0 \quad (1)$$

Large networks are usually sparse so the link (case) likelihood contribution term $l_1$ can be calculated analytically, even for massive networks. The non-link (control) likelihood contribution term $l_0$ has a quadratic complexity $\mathcal{O}(N^2)$ in terms of the size of the network, and thus its computation is infeasible. For that, we introduce an unbiased estimator for $l_{i,0}$, which is regarded as a population total statistic (Raftery et al., 2012). We estimate the non-link contribution of a node $\{i\}$ via:

$$l_{i,0} = \frac{N_{i,0}}{n_{i,0}} \sum_{k=1}^{n_{i,0}} - \log \left( 1 + \int_{t_i}^{T} \lambda_{ik}(t')dt' \right), \quad (2)$$

where $N_{i,0}$ is the number of total non-links (controls) for node $\{i\}$, and $n_{i,0}$ is the number of samples to be used for the estimation. We set the number of samples based on the node degrees as $n_{i,0} = 5 * \text{degree}_i$. This makes inference scalable defining an $\mathcal{O}(cE)$ space and time complexity.

## 7 Baselines

In the main paper, we compare our approach against four baselines, each employing a distinct encoding strategy:

(i) NODE2VEC (Grover and Leskovec, 2016): A renowned method based on random walks, learning node embeddings from the co-occurrence frequencies within generated node sequences. The parameters $p$ and $q$ are set to 1, so we performed unbiased random walks. The walk length ($L$), number of walks ($N$), and window size ($w$) are set to 10, 80, and 10, respectively.

(ii) CTDNE (Nguyen et al., 2018): This model extends NODE2VEC to temporal networks, performing random walks that consider the temporal characteristics of the network and incorporate link time information. Its parameters walk length ($L$) is set to 80, number of walks ($N$) to 10, and window size ($w$) is chosen as 2, unlike the conventional practice ($w = 10$) since it was giving error.

(iii) VERSE (Tsitsulin et al., 2018): VERSE targets the extraction of latent representations that capture the Personalized Page Rank (PPR) similarity among nodes. The $\alpha$ value is assigned to 0.85 as suggested in the paper. The number of samples is also considered as 5.

(iv) NETSMF (Qiu et al., 2019): This model learns embeddings through the decomposition of a specially designed matrix that encapsulates both high and low-order node interactions. The context windows size ($T$) and the rank parameters are selected as 10 and 256, respectively, and other parameters are set to their default values.

### 7.1 Link Prediction for the baseline methods

Similar to the proposed models, for the baselines, we utilize the residual network to learn the node embeddings and design a feature vector for each node pair sample by applying a binary operator (Grover and Leskovec, 2016). These feature vectors are then used to train a logistic regression model for the task of link prediction. This is contrary to our DISEE where link prediction can be made directly based on the Single-Event Poisson Process rates, highlighting the advantages of defining statistical models and a valid network likelihood. For the baselines, each operator defines a different link prediction performance for which we always choose the one with the highest score, making the baselines as competitive as they could possibly be. The detailed list of the operators is given in Table 5.

Table 5: Binary operators for constructing feature vectors for node pair samples. Each definition corresponds to d-th coordinate of the embedding vectors $\mathbf{z}_i$ and $\mathbf{z}_j$.

| | Average | Hadamard | Weighted $L_1$ | Weighted $L_2$ |
|---|---|---|---|---|
| Operator | $(z_{i,d} + z_{j,d})/2$ | $(z_{i,d} \times z_{j,d})$ | $\lvert z_{i,d} - z_{j,d} \rvert$ | $\lvert z_{i,d} - z_{j,d} \rvert^2$ |

## References


J. Barney. Firm resources and sustained competitive advantage. *Journal of Management*, 17:99–120, 1991. URL `http://jom.sagepub.com/cgi/content/abstract/17/1/99`.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009. doi: 10.1109/CVPR.2009.5206848.

A. Grover and J. Leskovec. Node2Vec: Scalable feature learning for networks. In *KDD*, pages 855–864, 2016.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization, 2017.

J. McAuley and J. Leskovec. From amateurs to connoisseurs: Modeling the evolution of user expertise through online reviews, 2013.

G. H. Nguyen, J. B. Lee, R. A. Rossi, N. K. Ahmed, E. Koh, and S. Kim. Continuous-time dynamic network embeddings. In *TheWebConf*, page 969–976, 2018.

J. Priem, H. Piwowar, and R. Orr. Openalex: A fully-open index of scholarly works, authors, venues, institutions, and concepts, 2022.

J. Qiu, Y. Dong, H. Ma, J. Li, C. Wang, K. Wang, and J. Tang. Netsmf: Large-scale network embedding as sparse matrix factorization. In *The World Wide Web Conference*, pages 1509–1520, 2019.

A. E. Raftery, X. Niu, P. D. Hoff, and K. Y. Yeung. Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919, 2012.

C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. *CoRR*, abs/1409.4842, 2014. URL `http://arxiv.org/abs/1409.4842`.

A. Tsitsulin, D. Mottin, P. Karras, and E. Müller. Verse: Versatile graph embeddings from similarity measures. In *Proceedings of the 2018 world wide web conference*, pages 539–548, 2018.