

# FinDKG: Dynamic Knowledge Graphs with Large Language Models for Detecting Global Trends in Financial Markets

Xiaohui Victor Li\*

Imperial College London  
London, United Kingdom

xiaohui.li21@alumni.imperial.ac.uk

## ABSTRACT

Dynamic knowledge graphs (DKGs) are popular structures to express different types of connections between objects over time. They can also serve as an efficient mathematical tool to represent information extracted from complex unstructured data sources, such as text or images. Within financial applications, DKGs could be used to detect trends for strategic thematic investing, based on information obtained from financial news articles. In this work, we explore the properties of large language models (LLMs) as dynamic knowledge graph generators, proposing a novel open-source fine-tuned LLM for this purpose, called the Integrated Contextual Knowledge Graph Generator (ICKG). We use ICKG to produce a novel open-source DKG from a corpus of financial news articles, called FinDKG, and we propose an attention-based GNN architecture for analysing it, called KGTransformer. We test the performance of the proposed model on benchmark datasets and FinDKG, demonstrating superior performance on link prediction tasks. Additionally, we evaluate the performance of the KGTransformer on FinDKG for thematic investing, showing it can outperform existing thematic ETFs.

## KEYWORDS

Dynamic knowledge graphs, graph attention networks, graph neural networks, graph transformers, large language models.

## 1 INTRODUCTION

A knowledge graph (KG) is a data structure that encodes information consisting in entities and different types of relations between them. Formally, a KG can be represented as  $\mathcal{G} = \{\mathcal{E}, \mathcal{R}, \mathcal{F}\}$ , where  $\mathcal{E}$  and  $\mathcal{R}$  denote the sets of entities and relations respectively, and  $\mathcal{F} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$  represents a set of facts, consisting in relations of different types between entities. The triplet  $(s, r, o) \in \mathcal{F}$  is the fundamental building block of a KG, where  $s \in \mathcal{E}$  represents the source entity,  $r \in \mathcal{R}$  the relation, and  $o \in \mathcal{E}$  the object entity. For instance, the triplet *(OpenAI, Invent, ChatGPT)* shows how entities and relations combine to form a fact, with *OpenAI* and *ChatGPT* as entities and *Invent* as the relation.

Temporal or dynamic knowledge graphs (DKGs) extend static KGs by incorporating temporal dynamics. Each fact in a DKG is associated with a timestamp  $t \in \mathbb{R}_+$ , allowing the model to capture the temporal evolution of events. Therefore, events occur in quadruples  $(s_i, r_i, o_i, t_i) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E} \times \mathbb{R}_+$ , where  $t_i$  is the event time, such that  $t_i \leq t_j$  for  $i < j$ ,  $i, j \in \mathbb{N}$ . Then, the DKG  $\mathcal{G}_t = (\mathcal{E}, \mathcal{R}, \mathcal{F}_t)$  at time  $t$  can be expressed via a time-varying set of facts  $\mathcal{F}_t$  defined as

$$\mathcal{F}_t = \{(s_i, r_i, o_i, t_i) : s_i, o_i \in \mathcal{E}, r_i \in \mathcal{R}, t_i < t\}. \quad (1)$$

\*This work was completed as part of XVL's MSc project at Imperial College London.

Francesco Sanna Passino

Imperial College London  
London, United Kingdom

f.sannapassino@imperial.ac.uk

The task of estimating a model for  $\mathcal{G}_t$  from observed data is called *dynamic knowledge graph learning*. This typically involves data-driven training of graph neural networks, designed to model both the structure and the temporal dynamics of the KGs over time.

In real-world applications such as finance, entities and relations can be further grouped into *categories*, often called *meta-entities*. For example, consider the relation between the entity *Jeff Bezos* which is of type *Person*, and the entity *Amazon*, which is of type *Company*. The relation between them is *Founder Of*, which could be considered to have the type *Business action*. In this work, inspired by heterogeneous graph transformers [HGT, 13], we discuss a way to introduce the additional meta-entity information within a dynamic knowledge graph learning procedure based on graph attention networks [GAT, 34] and EvoKG [25]. This results in the *Knowledge Graph Transformer (KGTransformer)*, an attention-based graph neural network (GNN) designed to create dynamic lower-dimensional representations of entities and relations.

In addition to DKGs, Large Language Models (LLMs) have also been gaining popularity recently within the financial sector, demonstrating potential in enhancing various financial tasks through advanced natural language processing (NLP) capabilities [23]. Popular models such as BERT, the GPT series, and financial-specific variants such as FinBERT [2] and FinGPT [41] leverage LLMs to improve the state-of-the-art in tasks such as financial sentiment analysis.

The application of LLMs to dynamic knowledge graphs has been so far limited in the literature. Therefore, one of the main contributions of this work is to also propose a pipeline for generative knowledge graph construction (KGC) via Large Language Models (LLMs), resulting in the *Integrated Contextual Knowledge Graph Generator (ICKG)* large language model. In particular, we develop a fine-tuned LLM to systematically extract entities and relationships from textual data via engineered input queries or “prompts”, subsequently assembling them into event quadruples of the same form as (1). We use the proposed ICKG LLM to generate an open-sourced financial knowledge graph dataset, called FinDKG.

In summary, our contributions in this work are threefold:

- (1) We propose KGTransformer, an attention-based GNN architecture for dynamic knowledge graph learning that includes information about meta-entities (cf. Section 4), combining existing work on GATs [34], HGTs [13] and EvoKG [25]. We demonstrate substantial improvements in link prediction metrics (cf. Section 5.1) on real-world DKGs.
- (2) We develop an open-source LLM for dynamic knowledge graph generation for finance called *Integrated Contextual Knowledge Graph Generator (ICKG)*, cf. Section 3).
- (3) We utilise ICKG to create an open-source dynamic knowledge graph based on financial news articles, called *FinDKG*

(cf. Section 3.1). FinDKG is used for thematic investing upon capitalizing on the AI trend, improving upon other AI-themed portfolios (cf. Section 5.3).

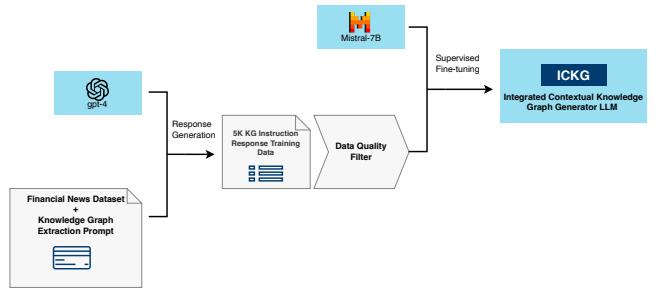
The remainder of this work is organised as follows: Section 2 discusses related literature. Next, Section 3 and 4 discuss the main contributions of our work: ICKG and KGTransformer. Finally, Section 5 discusses applications on real-world DKGs.

## 2 RELATED LITERATURE

*Graph representation learning.* Graph representation learning via graph neural networks (GNNs) is a fast-growing branch of deep learning, focused on extracting lower-dimensional latent space representations of graphs, to improve performance in downstream applications [7]. These methods have demonstrated significant capabilities in tasks such as node classification, edge prediction, and graph classification [18, 19, 40]. When applied to knowledge graphs, representation learning is aimed at deriving low-dimensional vector representations of entities and relations [15], called embeddings. Within the context of KGs, embeddings are then used for tasks such as information retrieval [27], question answering [4], and recommendations [36, 37]. Recent advancements in temporal knowledge graph learning have also integrated temporal information [6].

*Financial knowledge graphs.* Financial systems are often characterised by intricate and dynamically evolving relationships [1], which can be represented as DKGs for applications such as fraud transaction identification [38], stock return prediction [10], stock linkage discovery [9], and network-based portfolio construction [32]. However, the heterogeneous and dynamic nature of financial networks poses challenges for existing static GNN models, and the study of dynamic extensions of these models within a financial context remains relatively underdeveloped, despite advancements in financial natural language processing [12]. Early industry applications of financial KGs were based on static knowledge graph models [8, 11]. Also, [42] highlighted the potential of KGs in finance by developing a static macroeconomics knowledge graph for selecting variables in economic forecasting. Their KG-based methods improved forecasting accuracy. In this work, we propose an architecture which incorporates *meta-entities* within DKGs, and demonstrate its performance on finance-related tasks.

*LLMs in finance.* LLMs have been applied to a wide array of financial tasks. For example, [2] and [41] demonstrate the effectiveness of LLMs in extracting sentiment from financial news, social media, and corporate disclosures. [21] demonstrates good performance of GPT-4 in predicting stock market returns based on financial news headlines, claimed to be superior to sentiment analysis. Despite these advancements, challenges such as interpretability and computational costs with closed-sourced LLMs remain. [14] emphasises the need for improved interpretability in LLMs to promote transparency for financial applications. Moreover, while existing commercial LLMs such as GPT-4 offer substantial capabilities, their closed-source architecture imposes constraints on their usage. Open-source models such as Meta’s LLaMA [31] and Mistral AI’s LLM [16] offer more efficient alternatives, albeit often less precise.



**Figure 1: Flowchart of the fine-tuned ICKG LLM for knowledge graph construction, outlining the training methodology.**

## 3 THE INTEGRATED CONTEXTUAL KNOWLEDGE GRAPH GENERATOR (ICKG)

One of the objectives of this work is to propose an automated and scalable pipeline to extract temporal knowledge graphs from unstructured data sources, such as text. Large language models represent a natural choice for this task. Generative LLMs, while usually proficient in a wide array of tasks related to language, often require customization in more specialized applications, such as knowledge graph construction. This can be achieved via *supervised fine-tuning*, which involves the further training of a pre-trained LLM on a curated dataset that is tailored to the task at hand [23].

For the purposes of this work, we develop the *Integrated Contextual Knowledge Graph Generator (ICKG)*<sup>1</sup>, an open-sourced fine-tuned LLM, which is optimised for knowledge graph construction tasks and uses the GPT-4 API for data generation. The training workflow of ICKG was divided into the following steps:

- (1) First, a fine-tuning dataset is constructed from a small set of 5,000 open-sourced financial news articles. These are passed to GPT-4 one-by-one with a knowledge graph extraction prompt giving detailed instructions on the required output type, consisting in triplets extracted from the article. Additionally, our prompt asks to classify entities into a pre-defined set of *categories*, or meta-entities.
- (2) Next, an additional data quality filter is applied to the resulting output. Only responses that strictly adhere to the instruction prompt and return more than 5 quadruples per article were retained. This helps reducing the effect of noise and randomness in the GPT-4 output, refining the quality of the quadruples beyond the native capabilities of GPT-4.
- (3) The resulting set of quadruples is used to fine-tune the open-sourced Mistral 7B model [16], obtaining the final Integrated Contextual Knowledge Graph Generator (ICKG). The fine-tuning process was conducted over approximately 10 hours, utilizing 8 A100 GPUs with 40GB memory each.

The full workflow is depicted in Figure 1 diagram. Figure 2 displays an example of this pipeline, where an open-access news article is passed as input to the LLM, describing a set of predefined entity categories and relations and required output type. The output of the procedure is a set of quintuples representing the resulting KG.

<sup>1</sup>The ICKG-v3.2 model is publicly available on the HuggingFace platform for non-commercial research at <https://huggingface.co/victorlxh/ICKG-v3.2>.

### 3.1 The Financial DKG (FinDKG) dataset

Open-source real-world knowledge graphs are relatively scarce, particularly in the financial sector. Therefore, a contribution of this article is to provide an open-sourced financial dynamic knowledge graph dataset, called FinDKG<sup>2</sup>, constructed from scratch utilising our ICKG LLM proposed in the previous section. We collected approximately 400,000 financial news articles from the Wall Street Journal via open-source web archives, spanning from 1999 to 2023. Each article includes metadata such as release time, headlines, categories, in addition to the full textual content. We excluded articles with themes not closely related to economics and finance (such as entertainment, book recommendations, opinion columns).

ICKG is used to extract quintuples consisting in entities, entity categories, and relation type from each news article, with timestamps corresponding to the release date. The possible relations are restricted to 15 types relevant to financial news, summarised with examples in Table 1. The entities are tagged with a *category* selected from the list in Table 2. Additionally, the resulting quintuples undergo entity disambiguation via Sentence-BERT [26, 43]. An example of this procedure is given in Figure 2.

Figure 3 presents a snapshot subgraph of FinDKG as of January 2023, highlighting the most relevant entities at the time, ranked by graph centrality metrics. The graph shows signs of the geopolitical tensions between the United States and China, the rising global economic pressure of high inflation, and the effect of the COVID-19 pandemic. The resulting dataset is used in Section 5 for testing the graph learning procedure for DKGs proposed in Section 4.

## 4 GRAPH LEARNING VIA KGTransformers

Dynamic knowledge graph learning consists in the task of estimating a model which captures the structural and temporal characteristics of the observed data. The focus of this work is the *extrapolation* task, aimed at predicting future facts beyond the known time horizon, particularly *link prediction*: given a DKG  $\mathcal{G}_t$ , source entity  $s$ , a relation  $r$ , and a future time  $t$ , the objective is to predict the most likely object entity  $o^*$  which will complete the connection, forming the quadruple  $(s, r, o^*, t)$ . More formally, for each triplet  $(s, r, t)$ ,  $s \in \mathcal{E}$ ,  $r \in \mathcal{R}$ ,  $t \in \mathbb{R}_+$ , the objective is to estimate ranking functions expressing the likelihoods of quadruples  $(s, r, o, t)$ ,  $o \in \mathcal{E}$  to occur, as a function of  $o \in \mathcal{E}$ . In this work, we learn these functions via the novel *KGTransformer*, described in the next section.

### 4.1 The Knowledge Graph Transformer

In this section, we introduce the KGTransformer, an attention-based graph neural network (GNN) designed to construct lower dimensional representations of the entities, called *graph embeddings*. In addition to standard GNN architectures, KGTransformer incorporates meta-entities via an extended graph attention mechanism based on [13], borrowing strength across entity categories.

Consider a KG  $\mathcal{G} = (\mathcal{E}, \mathcal{R}, \mathcal{F})$ , where  $N = |\mathcal{E}|$ . The KGTransformer layer produces an embedding  $Y^{(\ell)} \in \mathbb{R}^{N \times D_\ell}$  of the entities, where  $D_\ell \in \mathbb{N}$  is the latent dimension of the  $\ell$ -th layer, for  $\ell \in \{1, \dots, L\}$ , initialised from a latent representation  $Y^{(0)} \in \mathbb{R}^{N \times D_0}$ ,  $D_0 \in \mathbb{N}$ . The latent features  $Y^{(\ell)}$  obtained as output of

the  $\ell$ -th layer are passed as input to the  $(\ell + 1)$ -th layer of the full network architecture, until a final output  $Y^{(L)} \in \mathbb{R}^{N \times D_L}$  is obtained.

At the  $\ell$ -th layer, the latent features  $Y^{(\ell)} \in \mathbb{R}^{N \times D_\ell}$  consist in an aggregation operation between  $H \in \mathbb{N}$  sub-vectors of the form  $Y^{(\ell)} \in \mathbb{R}^{N \times D_{\ell,h}}$ ,  $D_{\ell,h} \in \mathbb{N}$ , where  $\sum_{h=1}^H D_{\ell,h} = D_\ell$ , such that:

$$Y^{(\ell)} = \left[ Y_1^{(\ell)}, \dots, Y_H^{(\ell)} \right] \in \mathbb{R}^{N \times D_\ell},$$

by concatenation. Each component refers to a part of the input from the previous layer, creating a so-called *multi-head* system [33].

At the  $\ell$ -th layer, the basic update function for latent features  $Y_h^{(\ell)}[o]$  for an entity  $o \in \mathcal{E}$  in the KGTransformer consists in combination between the so-called *message vectors*, weighted by *attention scores*, according to the following aggregation equation:

$$Y_h^{(\ell)}[o] = \psi \left( \sum_{s \in \mathcal{E}, r \in \mathcal{R}: s \in \mathcal{N}_r(o)} \text{Atn}_h^{(\ell)}(s, r, o) \text{Msg}_h^{(\ell)}(s, r, o) \right), \quad (2)$$

where  $\mathcal{N}_r(o) = \{s \in \mathcal{E} : (s, r, o) \in \mathcal{F}\}$  is the set of type- $r$  neighbours for the entity  $o$ , and  $\text{Atn}_h^{(\ell)}(\cdot) \in \mathbb{R}$  and  $\text{Msg}_h^{(\ell)}(\cdot) \in \mathbb{R}^{D_{\ell,h}}$  are attention and message vectors, calculated from  $Y^{(\ell-1)}$ . Additionally,  $\psi(\cdot)$  is the element-wise Leaky-ReLU activation function.

*KGTransformer attention vectors.* The KGTransformer attention scores  $\text{Atn}_h^{(\ell)}(s, r, o)$  in (2) are calculated by applying the softmax transformation (denoted  $\sigma$ ) on a concatenation of scores  $\alpha_h^{(\ell)}(s, r, o)$  across entities in neighbourhoods  $\mathcal{N}_r(o)$  for each relation  $r \in \mathcal{R}$ :

$$\text{Atn}_h^{(\ell)}(s, r, o) = \sigma \left( \left\| \sum_{s \in \mathcal{E}, r \in \mathcal{R}: s \in \mathcal{N}_r(o)} \alpha_h^{(\ell)}(s, r, o) \right\| \right), \quad (3)$$

where  $\|\cdot\|$  denotes the concatenation operator. The normalisation via the softmax ensures that the weights in the update (2) sum to 1.

Each of the attention scores  $\alpha_h^{(\ell)}(s, r, o)$ ,  $h = 1, \dots, H$  in (3) is obtained after incorporating *meta-entities*. In particular, we assume that a function  $\tau : \mathcal{E} \rightarrow C_E$  exists, mapping each entity to an entity type, where all possible types are described by the set  $C_E$ . For example, consider the relation *Invent* between the source entity *OpenAI*, which is of type *Company*, and the object entity *ChatGPT*, which is of type *Product*. In the context of meta-entities, this could be represented as  $\tau(\text{OpenAI}) = \text{Company}$  and  $\tau(\text{ChatGPT}) = \text{Product}$ . Meta-entities are incorporated in the architecture via tensors  $\mu_h^{(\ell)} \in \mathbb{R}^{|C_E| \times |\mathcal{R}| \times |C_E|}$ ,  $h = 1, \dots, H$ ,  $\ell = 1, \dots, L$ , following the same approach of [13] on heterogeneous graphs. Following [13], the proposed KGTransformer attention score for the  $h$ -th head is:

$$\alpha_h^{(\ell)}(s, r, o) = \frac{K_h^{(\ell)}[s]^T W_{h,r}^{(\ell)} Q_h^{(\ell)}[o] \cdot \mu_h^{(\ell)}[\tau(s), r, \tau(o)]}{\sqrt{D_{\ell,h}}}, \quad (4)$$

where the vectors  $K_h^{(\ell)}[s]$ ,  $Q_h^{(\ell)}[o] \in \mathbb{R}^{D_{\ell,h} \times 1}$  in (4) are called *key* and *query* vectors for entities  $s$  and  $o$ , and  $W_{h,r}^{(\ell)} \in \mathbb{R}^{D_{\ell,h} \times D_{\ell,h}}$  is a trainable weighting matrix. The key and query vectors are derived from the latent features at the previous layer:

$$K_h^{(\ell)}[s] = P_{h,\tau(s)}^{(\ell)} Y^{(\ell-1)}[s], \quad Q_h^{(\ell)}[o] = R_{h,\tau(o)}^{(\ell)} Y^{(\ell-1)}[o],$$

where  $P_{h,c}^{(\ell)}, R_{h,c}^{(\ell)} \in \mathbb{R}^{D_{\ell,h} \times D_{\ell-1}}$ ,  $c \in C_E$ , are trainable matrices.

<sup>2</sup>FinDKG is available to download at <https://xiaohui-victor-li.github.io/FinDKG/#data>.

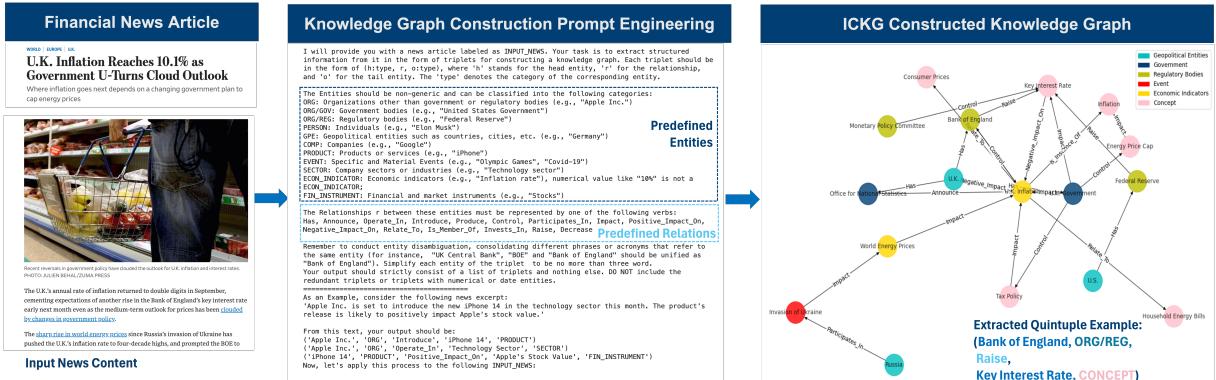


Figure 2: Illustration of the ICKG-enabled knowledge graph generation pipeline for FinDKG, representing the conversion of textual news articles into structured dynamic knowledge graph quintuples.

Relation	Definition	Example
Has	Indicates ownership or possession, often of assets or subsidiaries in a financial context.	Google Has Android
Announce	Refers to the formal public declaration of a financial event, product launch, or strategic move.	Apple Announces iPhone 13
Operate In	Describes the geographical market in which a business entity conducts its operations.	Tesla Operates In China
Introduce	Denotes the first-time introduction of a financial instrument, product, or policy to the market.	Samsung Introduces Foldable Screen
Produce	Specifies the entity responsible for creating a particular product, often in a manufacturing or financial product context.	Pfizer Produces Covid-19 Vaccine
Control	Implies authority or regulatory power over monetary policy, financial instruments, or market conditions.	Federal Reserve Controls Interest Rates
Participates In	Indicates active involvement in an event that has financial or economic implications.	United States Participates In G20 Summit
Impact	Signifies a notable effect, either positive or negative, on market trends, financial conditions, or economic indicators.	Brexit Impacts European Union
Positive Impact On	Highlights a beneficial effect on financial markets, economic indicators, or business performance.	Solar Energy Positive Impact On ESG Ratings
Negative Impact On	Underlines a detrimental effect on financial markets, economic indicators, or business performance.	Covid-19 Negative Impact On Tourism Sector
Relate To	Points out a connection or correlation with a financial concept, sector, or market trend.	AI Relates To FinTech Sector
Is Member Of	Denotes membership in a trade group, economic union, or financial consortium.	Germany Is Member Of EU
Invests In	Specifies an allocation of capital into a financial instrument, sector, or business entity.	Warren Buffett Invests In Apple
Raise	Indicates an increase, often referring to capital, interest rates, or production levels in a financial context.	OPEC Raises Oil Production
Decrease	Indicates a reduction, often referring to capital, interest rates, or production levels in a financial context.	Federal Reserve Decreases Interest Rates

Table 1: Relation types in the FinDKG dataset.

Category	Definition	Example
ORG	Non-governmental and non-regulatory organisations.	Imperial College London
ORG/GOV	Governmental bodies.	UK Government
ORG/REG	Regulatory bodies.	Bank of England
GPE	Geopolitical entities like countries or cities.	United Kingdom
PERSON	Individuals in influential or decision-making roles.	Jerome Powell
COMP	Companies across sectors.	Apple Inc.
PRODUCT	Tangible or intangible products or services.	iPhone
EVENT	Material events with financial or economic implications.	Brexit
SECTOR	Sectors or industries in which companies operate.	Technology Sector
ECON_IND	Non-numerical indicators of economic trends or states.	Inflation Rate
FIN_INST	Financial and market instruments.	S&P 500 Index
CONCEPT	Abstract ideas, themes, or financial theories.	Artificial Intelligence

Table 2: Entity categories in the FinDKG dataset.

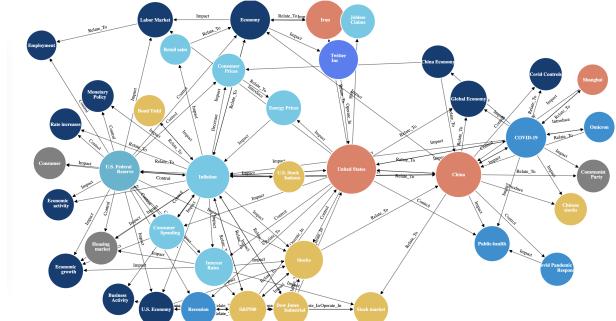


Figure 3: Subgraph of FinDKG’s most influential entities as of January 1, 2023. Entities are coloured by category.

*KGTransformer message vectors.* Similarly to the attention scores in (3), message vectors are obtained via different linear projections applied to the embedding  $Y^{(\ell-1)}$  from the previous layer [13]:

$$\text{Msg}_h^{(\ell)}(s, r, o) = Z_{h,r}^{(\ell)} M_{h,\tau(s)}^{(\ell)} Y^{(\ell-1)}[s],$$

where  $M_{h,\tau(s)}^{(\ell)} \in \mathbb{R}^{D_{\ell,h} \times D_{\ell-1}}$ ,  $Z_{h,r}^{(\ell)} \in \mathbb{R}^{D_{\ell,h} \times D_{\ell,h}}$  are matrices specific to the  $h$ -th head, meta-entity  $\tau(s)$ , and relation  $r$ .

## 4.2 Time-evolving updates for DKGs

So far, Section 4.1 only considered the case of a static knowledge graph. In this section, we discuss how to incorporate two different

types of time-varying representations, called *temporal* and *structural* embeddings, following the EvoKG framework in [25].

Let  $\mathcal{G}_t = (\mathcal{E}, \mathcal{R}, \mathcal{F}_t)$  be a DKG observed at discrete time points  $t = 1, \dots, T$ , such as  $\mathcal{F}_t \subseteq \mathcal{F}_{t'}$  for  $t < t'$ . We write  $\tilde{\mathcal{F}}_t = \mathcal{F}_t \setminus \mathcal{F}_{t-1}$  to denote the set of facts occurring in the time interval between  $[t-1, t]$ . This representation can be used to construct a set of KGs  $\tilde{\mathcal{G}}_t = (\mathcal{E}, \mathcal{R}, \tilde{\mathcal{F}}_t)$  where  $\tilde{\mathcal{F}}_t \cap \tilde{\mathcal{F}}_{t'} = \emptyset$  for  $t \neq t'$ .

First, we apply KGTransformer independently on each graph  $\tilde{\mathcal{G}}_t$ , obtaining an embedding representation  $Y_t^{(\ell)} \in \mathbb{R}^{N \times D_\ell}$  via (2), starting from an input embedding  $Y_t^{(\ell-1)} \in \mathbb{R}^{N \times D_{\ell-1}}$ :

$$Y_t^{(\ell)} = \text{KGTransformer}\left(Y_t^{(\ell-1)}, \tilde{\mathcal{G}}_t\right).$$

The evolution of the embeddings  $Y_t^{(\ell)}$ ,  $t = 1, \dots, T$  over time is modelled via a recurrent neural network (RNN), resulting in:

$$V_t^{(\ell)} = \text{RNN}\left(Y_t^{(\ell)}, V_{t-1}^{(\ell)}\right).$$

The values  $V_t^{(\ell)} \in \mathbb{R}^{N \times D_\ell}$ ,  $t = 1, \dots, T$ , are called *temporal embeddings*. Following [25], the temporal embeddings for the unique entities appearing in  $\mathcal{F}_{r,t} = \{(s, r', o, t) \in \tilde{\mathcal{F}}_t : r' = r\}$  are averaged to obtain a latent representation for the relations  $\tilde{Y}_t^{(\ell)} \in \mathbb{R}^{|\mathcal{R}| \times D_\ell}$ , which is analogously modelled via an RNN, giving a sequence of *temporal relation embeddings*  $\tilde{V}_t^{(\ell)} \in \mathbb{R}^{|\mathcal{R}| \times D_\ell}$ ,  $t = 1, \dots, T$ , where:

$$\tilde{V}_t^{(\ell)} = \text{RNN}\left(\tilde{Y}_t^{(\ell)}, \tilde{V}_{t-1}^{(\ell)}\right).$$

We denote the rows of  $V_t^{(\ell)}$  and  $\tilde{V}_t^{(\ell)}$  as  $v_{i,t}^{(\ell)}$  and  $\tilde{v}_{r,t}^{(\ell)}$ , for entity  $i$  and relation  $r$  respectively. These embedding representations will be used to model the conditional probability of the arrival time of the triplets  $(s, r, o) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}$ , as in the EvoKG framework [25].

In contrast, the conditional probabilities of the triplets given the graph  $\mathcal{G}_t$  will be modelled via the so-called *structural embeddings* [25]. These are obtained via a similar mechanism as above: the output of the KGTransformer is used within an RNN. Denoting the initial input embedding as  $X_t^{(\ell-1)} \in \mathbb{R}^{N \times D_{\ell-1}}$ , we write:

$$X_t^{(\ell)} = \text{KGTransformer}\left(X_t^{(\ell-1)}, \tilde{\mathcal{G}}_t\right), U_t^{(\ell)} = \text{RNN}\left(X_t^{(\ell)}, U_{t-1}^{(\ell)}\right).$$

The values  $U_t^{(\ell)} \in \mathbb{R}^{N \times D_\ell}$ ,  $t = 1, \dots, T$ , are called *structural embeddings*. As before, averaging over the entities appearing in the sub-graph of type  $r \in \mathcal{R}$  at time  $t$  gives embeddings  $\tilde{U}_t^{(\ell)} \in \mathbb{R}^{|\mathcal{R}| \times D_\ell}$ . As before, these are modelled via a recurrent neural network:

$$\tilde{U}_t^{(\ell)} = \text{RNN}\left(\tilde{X}_t^{(\ell)}, \tilde{U}_{t-1}^{(\ell)}\right).$$

As before,  $u_{i,t}^{(\ell)}$  and  $\tilde{u}_{r,t}^{(\ell)}$  are used to denote the rows of  $U_t^{(\ell)}$  and  $\tilde{U}_t^{(\ell)}$  respectively, corresponding to the structural embeddings at time  $t$  for entity  $i$  and relation  $r$ .

### 4.3 Dynamic knowledge graph learning

In this section, a probabilistic framework for learning DKGs is discussed, based on the work of [17, 25], integrated with the KGTransformer time-varying embeddings discussed in Section 4.2. The objective of the graph learning procedure is to estimate the model parameters that best describe the observed graph  $\mathcal{G}_T$  under the proposed model. Using  $\tilde{\mathcal{G}}_1, \dots, \tilde{\mathcal{G}}_T$ , we can decompose the probabilities associated with events occurred in the graph  $\mathcal{G}_T$  as follows:

$$\begin{aligned} p(\mathcal{G}_T) &= p(\tilde{\mathcal{G}}_1, \dots, \tilde{\mathcal{G}}_T) = \prod_{t=1}^T p(\tilde{\mathcal{G}}_t | \mathcal{G}_{t-1}) \\ &= \prod_{t=1}^T \prod_{(s,r,o,t) \in \tilde{\mathcal{F}}_t} p(t | s, r, o, \mathcal{G}_{t-1}) p(s, r, o | \mathcal{G}_{t-1}). \quad (5) \end{aligned}$$

The decomposition in (5) partitions the conditional probability into two components:  $p(s, r, o | \mathcal{G}_{t-1})$  captures the evolving graph structure, whereas  $p(t | s, r, o, \mathcal{G}_{t-1})$  controls the temporal dynamics. Therefore, a model should be postulated on both these probabilities to capture both temporal and structural characteristics of DKGs.

*Modelling the graph structure.* To approximate  $p(s, r, o | \mathcal{G}_t)$ , we use embeddings that represent the time-varying structural components of both entities and relationships. Let  $u_{i,t}, \tilde{u}_{r,t} \in \mathbb{R}^D$ ,  $D \in \mathbb{N}$ , be the structural embeddings for entity  $i$  and relation  $r$ , updated until time  $t$ , obtained from the final layer of the KGTransformer. Additionally, we combine those into a global embedding  $g_t = (g_{t,1}, \dots, g_{t,D}) \in \mathbb{R}^D$  that aggregates the embeddings of all entities up to time  $t$  [17]. Each entry of  $g_t$  is computed as follows:

$$g_{t,j} = \max_{i \in \mathcal{E}_t} \{u_{i,t,j}\}, \quad j = 1, \dots, D,$$

where  $\mathcal{E}_t = \{s \in \mathcal{E} : (s, r, o) \in \tilde{\mathcal{F}}_t \vee (o, r, s) \in \tilde{\mathcal{F}}_t, r \in \mathcal{R}, o \in \mathcal{E}\}$  is the set of entities involved in events in  $\tilde{\mathcal{F}}_t$ . The vector  $g_t$  is used as a global conditioning variable for computing  $p(s, r, o | \mathcal{G}_t)$  [17].

Following [25], we decompose  $p(s, r, o | \mathcal{G}_t)$  into entity and relationship level components as follows:

$$p(s, r, o | \mathcal{G}_t) = p(o | \mathcal{G}_t) \times p(r | o, \mathcal{G}_t) \times p(s | r, o, \mathcal{G}_t). \quad (6)$$

Each term is parametrised via a multilayer perceptron (MLP) [25]:

$$\begin{aligned} p(s | r, o, \mathcal{G}_t) &= \sigma \{\text{MLP}([\tilde{u}_{r,t}, u_{o,t}, g_t])\}, \\ p(r | o, \mathcal{G}_t) &= \sigma \{\text{MLP}([u_{o,t}, g_t])\}, \\ p(o | \mathcal{G}_t) &= \sigma \{\text{MLP}(g_t)\}. \end{aligned} \quad (7)$$

Similarly to (6), the equivalent decomposition

$$p(s, r, o | \mathcal{G}_t) = p(s | \mathcal{G}_t) \times p(r | s, \mathcal{G}_t) \times p(o | r, s, \mathcal{G}_t)$$

could also be used, and parametrised via three MLPs as in (7).

*Modelling the temporal dynamics.* Following [25], we model  $p(t | s, r, o, \mathcal{G}_t)$  via a mixture of  $M \in \mathbb{N}$  log-normal distributions:

$$p(t | s, r, o, \mathcal{G}_t) = \sum_{m=1}^M w_m \phi_{\text{LN}}(t; \mu_m, \sigma_m),$$

where  $\phi_{\text{LN}}(t; \mu_m, \sigma_m)$  is the log-normal density function, where  $w_m, \mu_m, \sigma_m$  are the weight, mean, and standard deviation of the  $m$ -th component, such that  $w_m, \sigma_m \geq 0$  for all  $m = 1, \dots, M$ , and  $\sum_{m=1}^M w_m = 1$ . Model parameters are learned through an MLP that receives inputs composed of concatenated temporal embeddings for each entity and relation derived from the KGTransformer.

*Inference on the model parameters.* The model parameters are learned by minimising a composite loss function, which follows again the approach of [25] with a minor adjustment for relational symmetries. In particular, we let the loss function be:

$$\begin{aligned} \mathcal{L} &= - \sum_{t=1}^T \sum_{(s,r,o,t) \in \tilde{\mathcal{F}}_t} \left\{ \lambda_1 \log p(t | s, r, o, \mathcal{G}_{t-1}) \right. \\ &\quad + \lambda_2 [\log p(o | \mathcal{G}_{t-1}) + \log p(r | o, \mathcal{G}_{t-1}) + \log(s | r, o, \mathcal{G}_{t-1}) \\ &\quad \left. + \log p(s | \mathcal{G}_{t-1}) + \log p(r | s, \mathcal{G}_{t-1}) + \log(o | r, s, \mathcal{G}_{t-1})] \right\}, \end{aligned}$$

where  $\lambda_1, \lambda_2 \in \mathbb{R}_+$  are tunable hyperparameters. In order to manage computational and memory requirements, truncated backpropagation through time [TBPTT; see 25, 39] is used to minimise  $\mathcal{L}$ .

*Link prediction.* As described in the introduction, the model performance is evaluated on link prediction, aimed at predicting the most likely object  $o$  for an incomplete quadruple  $(s, r, ?, t)$ . The predicted entity  $\hat{o}$  is obtained as  $\hat{o} = \operatorname{argmax}_{o \in \mathcal{E}} p(o | s, r, \mathcal{G}_t)$ , where the distribution  $p(o | s, r, \mathcal{G}_t)$  is estimated via the MLP in (7).

## 5 EXPERIMENTS AND APPLICATIONS

In this section, we test the performance of KGTransformer for link prediction tasks on popular benchmarks used in the literature and on the newly created FinDKG dataset. Additionally, we evaluate the performance of FinDKG, generated by ICKG LLM, in detecting financial trends from the news articles by analysing graph centrality measures. We also explore its application for thematic investing.

### 5.1 Link prediction on real-world DKGs

We conduct experiments on various real-world knowledge graph datasets to evaluate the efficacy of our proposed KGTransformer model, focusing on its performance for link prediction.

*Performance metrics.* Following existing literature [see, for example, 25], we measure the model’s accuracy for link prediction using Mean Reciprocal Rank (MRR) and Hits@n (specifically Hits@3 and Hits@10). The MRR is defined for a set  $Q$  of test quadruples by summing the inverses of the ranks associated with each quadruple:  $MRR = \sum_{q \in Q} \text{rank}_q^{-1} / |Q|$ , where  $\text{rank}_q$  is the position of the true link in the ranked list of predictions. On the other hand, *Hits@n* measures the proportion of true links ranked within the top- $n$  predictions. A validation set is used to implement an early stopping mechanism to avoid overfitting.

*Baseline models for comparisons.* We compare the performance of the proposed KGTransformer against the following methods:

- Static graph models: R-GCN [29], which treats the graph as time-invariant, providing a baseline.
- Temporal graph models: RE-Net [17] and EvoKG [25].
- A KGTransformer version excluding meta-relations (denoted “KGTransformer w/o node types” in plots).

*Implementation details.* The KGTransformer is implemented with two layers of transformation blocks, with each embedding having a dimensionality of 200. We adhere to the original specifications for baseline KG models. All models are optimized using the AdamW algorithm [22] with a learning rate of  $5 \times 10^{-4}$  and an early stopping mechanism triggered after 10 epochs of no validation improvement.

Both model training and evaluations are consistently conducted on an identical computational environment: a single NVIDIA A100 GPU cloud server with 40GB of memory. To account for the inherent variability in model training, we employ three distinct random seeds, shared across different models. The final results are reported as averages over these training runs. Results across different seeds exhibit minimal variance for the datasets used in this work.

*Datasets for evaluation.* We evaluate the performance of the proposed KGTransformer architecture on publicly accessible real-world DKGs used as benchmarks in the literature [25], alongside

Dataset	$N_{\text{train}}$	$N_{\text{val}}$	$N_{\text{test}}$	$ \mathcal{E} $	$ \mathcal{R} $	$ \mathcal{C}_{\mathcal{E}} $
YAGO	161,540	19,523	20,026	10,623	10	-
WIKI	539,286	67,538	63,110	12,554	24	-
ICEWS14	275,367	48,528	341,409	12,498	260	-
FinDKG	119,549	11,444	13,069	13,645	15	12

Table 3: Summaries of the DKGs used for model evaluation.

Model	YAGO			WIKI			ICEWS14		
	MRR	H@3	H@10	MRR	H@3	H@10	MRR	H@3	H@10
R-GCN	27.43	31.24	44.75	13.96	15.75	22.05	15.03	16.12	31.47
RE-Net	46.35	51.93	61.47	31.45	34.23	41.15	23.81	26.57	42.62
EvoKG	49.86	57.69	65.42	42.56	47.18	52.34	<b>24.24</b>	<b>27.25</b>	41.97
<b>KGTransformer</b>	<b>51.33</b>	<b>59.22</b>	<b>67.15</b>	<b>44.32</b>	<b>49.27</b>	<b>53.81</b>	23.98	26.89	41.22

Table 4: Performance comparison on the benchmark DKGs datasets in terms of MRR, Hits@3,10. Best results are in bold.

the FinDKG introduced as part of this work, described in Section 3.1. Summary statistics about these datasets are described in Table 3. ICEWS consists in a collection of cooperative or hostile actions between socio-political actors corresponding to individuals, groups, sectors and nation states [5]. YAGO dataset is a large, automatically generated knowledge base that combines relational data from Wikipedia and WordNet [30], whereas WIKI is based on triplets extracted from the Wikidata database [20, 35].

It must be remarked that the only dataset containing meta-entities is FinDKG: therefore, we expect the benefits of KGTransformer to be particularly evident for this dataset. For the other benchmarks, the identity mapping function  $\tau(s) = s$  is used, implying that  $\mathcal{E} = \mathcal{C}_{\mathcal{E}}$ .

*Results on benchmarks and FinDKG.* Table 4 displays the temporal link prediction scores across the benchmark DKGs, and Figure 4 depicts the results on FinDKG. From the table, it can be seen that the static method R-GCN under-performs in temporal settings, highlighting the importance of temporal features. KGTransformer outperforms competitors on the YAGO and WIKI datasets, but it does not improve performance on the ICEWS14 dataset. The advantages of the KGTransformer are more evident on the FinDKG, which explicitly contains entity types (cf. Table 2, 3). Integrating these types into the KGTransformer enhances performance significantly, resulting in an approximate 10% improvement in MRR and Hits@3,10 metrics over temporal baselines. This demonstrates the superior performance of KGTransformer when entity categories are also available, providing a way to directly incorporate them into the model architecture. It must be remarked that, when entity categories are not included within the architecture (“KGTransformer w/o node types”), the results align closely with the temporal baselines, demonstrating the benefit of introducing this information.

### 5.2 Trend identification in financial news

Analysing the results of FinDKG gives a way to dynamically track the global financial network and evaluate the performance of the ICKG LLM to extract valuable information from financial news. To visualise this, we form a series of FinDKGs where rolling 1-month snapshot knowledge graphs were assembled every week on Sundays. These graphs stored the event quadruples of the preceding

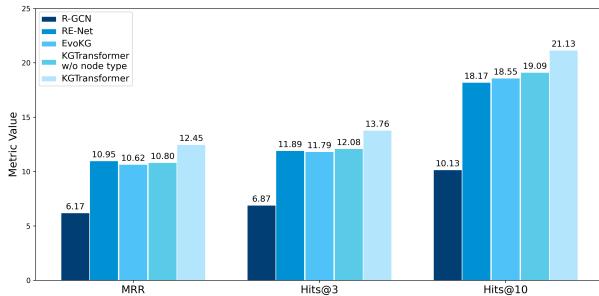


Figure 4: Performance comparison of models on FinDKG.

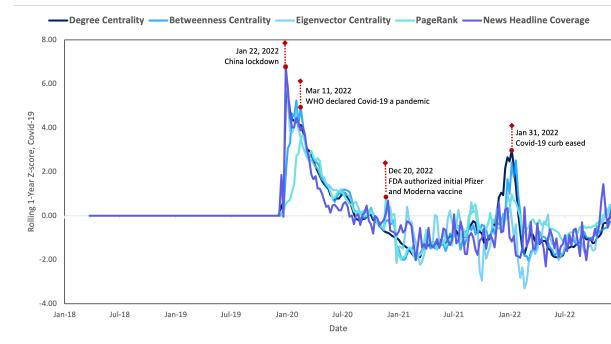


Figure 5: Evolution of the Covid-19 entity centrality measures over time between January 2018 and December 2022.

month. Four graph metrics of centrality were used to quantify the significance of an entity within each temporal knowledge graph: degree centrality, betweenness centrality, eigenvector centrality and PageRank. To standardize these measures over time for comparability, we apply a rolling one-year z-score normalization, making centrality metrics comparable across different times and entities.

We select the global COVID-19 pandemic as a case study. Figure 5 depicts the centrality metrics related to the Covid-19 entity as inferred by FinDKG. We compare the results with a standard measure based on headline coverage of the topic, commonly used in financial NLP applications [3]. These centrality measures appear to effectively capture significant moments in the pandemic timeline.

### 5.3 FinDKG-based thematic investing

Thematic investing is an investment strategy that targets specific themes or trends that are anticipated to influence the future landscape of industries and economies. We demonstrate the utility of FinDKG and link prediction with KGTransformer in estimating corporate exposure to AI, increasingly popular since the launch of OpenAI’s ChatGPT. The objective is to quantitatively measure how closely aligned stock entities are to the prevalence of the AI theme and to generate forward-looking exposure scores.

In an online learning setting, we fit a KGTransformer model within the three-year rolling window FinDKGs at the end of every quarter. At each time  $t$  (corresponding to the end of each month), the fitted KGTransformer is used to predict which stock entities



Figure 6: Cumulative returns of AI-themed long-only portfolios and market indices from June 2022 to December 2023.

are likely to be impacted by AI in the upcoming period  $t + 1$ , corresponding to the quadruple (AI, Impact, ?,  $t + 1$ ). Only stocks with a predicted impact likelihood exceeding the average across all entities are retained. This selection forms the basis of a monthly-rebalanced, AI-focused long-only portfolio within the US S&P 500. The portfolio is constructed by using the normalised predicted likelihood scores as the holding weight, which sum up to 100%. We denote this KGTransformer-powered portfolio as FinDKG-AI. We also fit an EvoKG model as an alternative KG learning methodology, using the same FinDKG data and settings, thereby constructing an EvoKG-based AI portfolio as a baseline strategy.

The out-of-sample backtesting results in Table 5 show the efficacy of the FinDKG-based AI portfolio: FinDKG-AI achieves the highest annualized return and Sharpe ratio across all portfolios. The existing AI ETFs lag behind the market benchmark with less return and comparably larger risk. In contrast, the KGTransformer-based FinDKG AI portfolio outperforms competitors across the evaluation period, with a jump coinciding approximately with the release of OpenAI’s ChatGPT in November 2022, as shown in Figure 6. The FinDKG-AI portfolio also outperforms the EvoKG-based strategy, aligning with the better link prediction capabilities of the KGTransformer model architecture shown in previous section.

## 6 CONCLUSION

In this work, we provided three contributions around the use of dynamic knowledge graphs (DKGs) and large language models (LLMs) within financial applications. First, we investigated the performance of fine-tuned open-source LLMs in generating knowledge graphs, proposing the novel open-source Integrated Contextual Knowledge Graph Generator (ICKG) LLM. Next, the ICKG LLM is used to create an open-source dataset from a corpus of financial news articles, called FinDKG. Additionally, we proposed an attention-based architecture called KGTransformer, which incorporates information from meta-entities within the learning process, combining architectures such as HGT [13] and EvoKG [25].

Our findings show that the proposed KGTransformer architecture improves the state-of-the-art link prediction performance on two benchmark datasets, and it achieves the best performance with over 10% uplift on FinDKG. The generalizability of the ICKG LLM

Portfolio	Name / Model	Annualized Return	Annualized Volatility	Sharpe Ratio	Max DD
SPY	SPDR S&P 500 ETF	18.6%	<b>17.1%</b>	1.084	<b>-16.7%</b>
QQQ	Invesco NASDAQ-100 ETF	<b>29.8%</b>	22.5%	<b>1.323</b>	-21.6%
ARKK	ARK Innovation ETF	20.6%	47.9%	0.431	-43.2%
IRBO	Invesco AI and Next Gen Software ETF	20.4%	25.9%	0.786	-24.9%
IGPT	iShares Robotics and AI ETF	18.7%	22.8%	0.820	-20.0%
EvoKG	FinDKG EvoKG-predicted Portfolio	23.5%	<b>20.2 %</b>	1.163	-19.1%
FinDKG-AI	FinDKG KGTransformer-predicted AI Portfolio	<b>39.6%</b>	21.9%	<b>1.810</b>	<b>-18.2%</b>

**Table 5: Overall performance of market, AI-themed ETF, and FinDKG portfolios. The top two performing portfolios within the metric are highlighted in bold, and the best one is further underlined. The evaluation period is from 30/06/2022 to 29/12/2023.**

extends beyond the financial news and financial domain, as evidenced by applications in the recent literature adopting similar frameworks [24, 28]. Code associated with this work can be found in the GitHub repository [xiaohui-victor-li/FinDKG](https://xiaohui-victor-li/FinDKG), and an online portal to visualise FinDKG is available at <https://xiaohui-victor-li.github.io/FinDKG/>.

## ACKNOWLEDGMENTS

FSP acknowledges funding from the EPSRC, grant no. EP/Y002113/1.

## REFERENCES

- [1] Daron Acemoglu, Ufuk Akcigit, and William Kerr. 2016. Networks and the macroeconomy: An empirical exploration. *NBER Macroeconomics Annual* 30, 1 (2016), 273–335.
- [2] Dogu Araci. 2019. FinBERT: Financial sentiment analysis with pre-trained language models. *arXiv preprint arXiv:1908.10063* (2019).
- [3] Scott R Baker, Nicholas Bloom, and Steven J Davis. 2016. Measuring economic policy uncertainty. *The Quarterly Journal of Economics* 131, 4 (2016), 1593–1636.
- [4] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. 2015. Large-scale simple question answering with memory networks. *arXiv preprint arXiv:1506.02075* (2015).
- [5] Elizabeth Boschee, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz, and Michael Ward. 2015. ICEWS Coded Event Data. <https://doi.org/10.7910/DVN/28075>
- [6] Li Cai, Xin Mao, Yuhao Zhou, Zhaoguang Long, Changxu Wu, and Man Lan. 2024. A survey on temporal knowledge graph: representation learning and applications. *arXiv preprint arXiv:2403.04782* (2024).
- [7] Fenxiao Chen, Yun-Cheng Wang, Bin Wang, and C-C Jay Kuo. 2020. Graph representation learning: a survey. *APSIPA Transactions on Signal and Information Processing* 9 (2020), e15.
- [8] Dawei Cheng, Fangzhou Yang, Xiaoyang Wang, Ying Zhang, and Liqing Zhang. 2020. Knowledge graph-based event embedding framework for financial quantitative investments. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2221–2230.
- [9] Andy Chung and Kumiko Tanaka-Ishii. 2023. Modeling momentum spillover with economic links discovered from financial documents. In *Proceedings of the Fourth ACM International Conference on AI in Finance*. 490–497.
- [10] Fuli Feng, Xiangnan He, Xiang Wang, Cheng Luo, Yiqun Liu, and Tat-Seng Chua. 2019. Temporal relational ranking for stock prediction. *ACM Transactions on Information Systems (TOIS)* 37, 2 (2019), 1–30.
- [11] Xiaoyi Fu, Xinqi Ren, Ole J Mengshoel, and Xindong Wu. 2018. Stochastic optimization for market return prediction using financial knowledge graph. In *2018 IEEE International Conference on Big Knowledge (ICBK)*. IEEE, 25–32.
- [12] Matthew Gentzkow, Bryan Kelly, and Matt Taddy. 2019. Text as data. *Journal of Economic Literature* 57, 3 (2019), 535–574.
- [13] Ziniu Hu, Yuxiao Dong, Kuansan Wang, and Yizhou Sun. 2020. Heterogeneous Graph Transformer. *arXiv preprint arXiv:2003.01332* (2020).
- [14] Pau Rodriguez-Inserte, Mariam Nakhlé, Raheel Qader, Gaëtan Caillaud, and Jingshu Liu. 2024. Large language model adaptation for financial sentiment analysis. *arXiv preprint arXiv:2401.14777* (2024).
- [15] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and S Yu Philip. 2021. A survey on knowledge graphs: representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems* 33, 2 (2021), 494–514.
- [16] Albert Q. Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lampe, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B. *arXiv preprint arXiv:2310.06825* (2023).
- [17] Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. 2019. Recurrent event network: autoregressive structure inference over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530* (2019).
- [18] Shima Khoshrafter and Aijun An. 2024. A survey on graph representation learning methods. *ACM Transactions on Intelligent Systems and Technology* 15, 1 (2024), 1–55.
- [19] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [20] Julien Leblay and Melisachew Wudage Chekol. 2018. Deriving validity time in knowledge graph. In *Companion Proceedings of The Web Conference 2018*. 1771–1776.
- [21] Alejandro Lopez-Lira and Yuehua Tang. 2023. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint arXiv:2304.07619* (2023).
- [22] Ilya Loshchilov and Frank Hutter. 2019. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations (ICLR)*.
- [23] Yuqi Nie, Yaxuan Kong, Xiaowen Dong, John M Mulvey, H Vincent Poor, Qing-song Wen, and Stefan Zohren. 2024. A Survey of Large Language Models for Financial Applications: Progress, Prospects and Challenges. *arXiv preprint arXiv:2406.11903* (2024).
- [24] Kun Ouyang, Yi Liu, Shicheng Li, Ruihan Bao, Keiko Harimoto, and Xu Sun. 2024. Modal-adaptive Knowledge-enhanced Graph-based Financial Prediction from Monetary Policy Conference Calls with LLM. *arXiv preprint arXiv:2403.16055* (2024).
- [25] Namyoung Park, Fuchen Liu, Purvanshi Mehta, Dana Cristofor, Christos Faloutsos, and Yuxiao Dong. 2022. EvoKG: Jointly modeling event time and network structure for reasoning over temporal knowledge graphs. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*. 794–803.
- [26] Nils Reimers and Iryna Gurevych. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084* (2019).
- [27] Ridho Reinanda, Edgar Meij, Maarten de Rijke, et al. 2020. Knowledge graphs: an information retrieval perspective. *Foundations and Trends® in Information Retrieval* 14, 4 (2020), 289–444.
- [28] Bhaskarjit Sarmah, Benika Hall, Rohan Rao, Sunil Patel, Stefano Pasquali, and Dhagash Mehta. 2024. HybridRAG: Integrating Knowledge Graphs and Vector Retrieval Augmented Generation for Efficient Information Extraction. *arXiv preprint arXiv:2408.04948* (2024).
- [29] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *The 15th Semantic Web International Conference*. Springer, 593–607.
- [30] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. Yago: a core of semantic knowledge. In *Proceedings of the 16th International Conference on World Wide Web (Banff, Alberta, Canada) (WWW ’07)*. Association for Computing Machinery, New York, NY, USA, 697–706.
- [31] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lampe. 2023. LLaMA: Open and Efficient Foundation Language Models. *arXiv preprint arXiv:2302.13971* (2023).
- [32] Edward Turner and Mihai Cucuringu. 2023. Graph denoising networks: a deep learning framework for equity portfolio construction. In *Proceedings of the Fourth ACM International Conference on AI in Finance*. 193–201.
- [33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).
- [34] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903* (2017).

- [35] Denny Vrandečić and Markus Krötzsch. 2014. Wikidata: a free collaborative knowledgebase. *Commun. ACM* 57, 10 (Sept. 2014), 78–85. <https://doi.org/10.1145/2629489>
- [36] Hongwei Wang, Fuzheng Zhang, Jialin Wang, Miao Zhao, Wenjie Li, Xing Xie, and Minyi Guo. 2018. RippleNet: Propagating user preferences on the knowledge graph for recommender systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. 417–426.
- [37] Hongwei Wang, Fuzheng Zhang, Mengdi Zhang, Jure Leskovec, Miao Zhao, Wenjie Li, and Zhongyuan Wang. 2019. Knowledge-aware graph neural networks with label smoothness regularization for recommender systems. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 968–977.
- [38] Mark Weber, Giacomo Domeniconi, Jie Chen, Daniel Karl I Weidele, Claudio Bellei, Tom Robinson, and Charles E Leiserson. 2019. Anti-money laundering in Bitcoin: experimenting with graph convolutional networks for financial forensics.
- [39] Ronald J Williams and Jing Peng. 1990. An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation* 2, 4 (1990), 490–501.
- [40] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).
- [41] Hongyang Yang, Xiao-Yang Liu, and Christina Dan Wang. 2023. FinGPT: Open-Source Financial Large Language Models. *arXiv preprint arXiv:2306.06031* (2023).
- [42] Yucheng Yang, Yue Pang, Guanhua Huang, et al. 2020. The knowledge graph for macroeconomic analysis with alternative big data. *arXiv preprint arXiv:2010.05172* (2020).
- [43] Alexandros Zeakis, George Papadakis, Dimitrios Skoutas, and Manolis Koubarakis. 2023. Pre-trained Embeddings for Entity Resolution: An Experimental Analysis. *arXiv preprint arXiv:2304.12329* (2023).