Dynamic Representations of Global Crises: A Temporal Knowledge Graph For Conflicts, Trade and Value Networks

Julia Gastinger*
University of Mannheim
and NEC Laboratories Europe

Timo Sztyler
NEC Laboratories Europe

Nils Steinert
Implisense GmbH

Sabine Gruender-Fahrer
InfAI / Leipzig University

Michael Martin
Chemnitz University of Technology
and InfAI / Leipzig University

Anett Schuelke
NEC Laboratories Europe

Heiner Stuckenschmidt
University of Mannheim

Abstract

This paper presents a novel approach to understanding global crises and trade patterns through the creation and analysis of a Temporal Knowledge Graph (TKG), and the application of Temporal Knowledge Graph Forecasting[2]. Combining data from the Armed Conflict Location & Event Data Project (ACLED) and Global Trade Alerts (GTA), the TKG offers a comprehensive view of the intersection between worldwide crises and global trade over time. We detail the process of TKG creation, including the aggregation and merging of information from multiple sources. Furthermore, we report key statistics of the TKG, providing insights into its potential applicability to data-driven Resilience Research. Leveraging the constructed TKG, we predict global trade events, such as trade sanctions across various categories and countries, and conflict events, such as worldwide military actions, to identify potential trade disruptions and anticipate the economic impact of global conflicts. To achieve this, state-of-the-art models for TKG Forecasting are applied and rigorously evaluated, contributing to a deeper understanding of the complex relationship between global crises and trade dynamics.

1 Introduction

Today, the world is facing multiple crises with different social, economic, and ecological consequences. Recent events like the Covid-19 pandemic and the Russia-Ukraine War have highlighted the interdependencies of global supply chains and economic value networks. Challenges such as supply chain disruptions and healthcare availability, define a new era where "managing disruptions defines sustainable growth more than managing continuity" [2]. Economic adversities can occur at any time and in various granularities, such as company crises, market crises, or global economic crises. To effectively address these challenges, businesses must continually adapt their operating models, value chains, and global networks to improve their flexibility and ability to respond quickly and agilely to changing environmental factors. This is encapsulated by the concept of resilience.

Resilience research is a challenging but urgently needed scientific field which will contribute to solving pressing societal issues. In response to this need, researchers in various fields, including information and communications technology, data science, and artificial intelligence (compare e.g., [3], [4], [5], [6]), have made significant contributions to resilience and crisis research.

*`julia.gastinger@uni-mannheim.de`

[2]A preliminary version of this work was presented at the Second International Workshop on Linked Data-driven Resilience Research (D2R2'23). In contrast to the workshop presentation [1], this paper incorporates a refined dataset, which has been updated and modified. Furthermore, it applies state-of-the-art methods for TKG Forecasting using the enhanced dataset, and analyses their predictions.

J. Gastinger et al., Dynamic Representations of Global Crises: A Temporal Knowledge Graph For Conflicts, Trade and Value Networks. *Proceedings of the Third Learning on Graphs Conference (LoG 2024)*, PMLR 269, Virtual Event, November 26–29, 2024.

Our work focuses on specific aspects of resilience research, namely trade-related policy measures, sanctions, and political violence and conflicts. It models these events with Temporal Knowledge Graphs. Temporal Knowledge Graphs (TKG) are Knowledge Graphs (KG) where facts occur, recur, or evolve over time [7]. KG triples are extended with timestamps to indicate that they are valid at a given time, allowing to hold time-evolving multi-relational data [8]. Given their capacity to depict both the interconnectivity of systems and their dynamic evolution, TKG are highly suitable for application in crisis and resilience research. They can be used to understand the evolution of complex economic supply chains over time, with a particular focus on the impact of interlinked crises. An example for such interlinked crises is Russia's invasion of Ukraine in late February 2022 leading to a significant increase in global food prices, and thus fuelling a global food crisis [9]. The research field of TKG forecasting (see e.g., [10], [11], [8]) predicts facts at future timesteps based on a history of a KG [12]. In crisis and resilience research, this capability can be applied not only to analyse the interconnectivity of systems, but also to predict the future evolvement and links in these systems, allowing for timely interventions.

This paper introduces a novel Temporal Knowledge Graph that covers interlinked worldwide crisis and trade sanction events for the year 2023, providing a comprehensive view of the dynamic relationships of these events. By using the presented TKG for downstream learning tasks, we can predict future developments in the global landscape of crisis and trade sanctions. Specifically, we propose the following Use Case: We want to use TKG Forecasting to predict upcoming global trade alert events and their links, as well as upcoming crisis events based on historic events. As an example, we aim at predicting links such as (France, IsAffectedbyEventofType, ?, t^+), (France, IsAffectedByInterventionOnSector, ?, t^+), where t^+ denotes a future timestep.

1.1 Terminology

A *Temporal Knowledge Graph* G is a set of quadruples (s, r, o, t) with $s, o \in E$, relation $r \in R$, and time stamp $t \in T$ with $T = \{1 \ldots n\}, n \in \mathbb{N}^+$. More precisely, E is the set of entities, R is the set of possible relations, and T is the set of timesteps. An example for such a quadruple is (France, IsAffectedbyEventofType, PeacefulProtest, 2023-07-01).

Temporal Knowledge Graph Forecasting is the task of predicting quadruples for future timesteps t^+ given a history of quadruples G, with $t^+ > n$ and $t^+ \in \mathbb{N}^+$. In this work we focus on entity forecasting, also termed future link prediction, that is, predicting object or subject entities for queries $(s, r, ?, t^+)$ (tail prediction) or $(?, r, o, t^+)$ (head prediction).

1.2 Contribution

With this paper, we showcase how TKG Forecasting can be applied to a real-world use case, to understand and predict the complex interactions between global crises and trade events. Further, we provide and analyse a newly curated dataset for the TKG research community. The main contributions of our work are:

1. We introduce a new Temporal Knowledge Graph that models global conflicts and trade over time (Section 3) and conduct an analysis of the TKG (Section 3.4).
2. We show how TKG Forecasting can be applied in resilience research, by applying state-of-the-Art models for TKG Forecasting to predict upcoming global trade alert events and their links, as well as upcoming crisis events (Section 5).
3. We thoroughly evaluate these models and their performance on various aspects to gain insights for future research (Section 5.2).
4. We release the dataset and the code for evaluating the models, to facilitate further research in this field `https://github.com/JuliaGast/GTA_ACLED_TKG/`

2 Related Work

Knowledge Graphs and Vocabularies for Resilience Research: The use of Knowledge Graphs in resilience and crisis research has gained increasing attention in recent years [13]. KG offer a flexible and comprehensive approach to modelling and analysing complex systems [14], making them suitable for a wide range of domains, including macro-economical analysis [15].

Creating a KG requires a structured and standardized way to represent data in a machine-readable format. Ontologies offer a means to provide a shared vocabulary of terms and concepts that enable data to be integrated and analysed in a consistent and interoperable way. Although there exist established vocabularies [16] to model events, including their relevant actors, occurrence, locality, and other significant properties, the reuse of such vocabularies presents several challenges. These challenges arise from the complex, highly domain-specific nature of these vocabularies, divergent levels of granularity, lack of easy extensibility, and the difficulty of creating interoperable mappings between different ontologies. As a result the CoyPu COY ontology [17] was presented to model macro-economically relevant and market-specific data, and information on current global crisis and conflict events. The TKG presented in our work bases on the COY ontology (details in Appendix A.4).

Temporal Knowledge Graph Datasets: In the domain of TKG analysis, completion and forecasting, six datasets have been published and utilized, including different versions of the Integrated Crisis and Early Warning System (ICEWS) [18]: ICEWS05-15 [19], ICEWS14 [19], and ICEWS18 [10] (the numbers indicate the covered years); GDELT [20]; YAGO [21]; and WIKI [22] (preprocessed by [10]). The ICEWS datasets consist of coded interactions between socio-political actors (i.e., cooperative or hostile actions between individuals, groups, sectors and nation states). However no TKG currently exists that describes trade relations and sanctions over time, or that merges the information of trade data and crisis data.

3 Creation of the Temporal Knowledge Graph

In this section, we explain the resources, describe the relationships between the source datasets, and outline the TKG's creation process. We also define the resulting dataset schema. Additionally, we analyze dataset properties and compare it to existing TKG datasets.

3.1 Resources

GTA: Global Trade Alerts (GTA) [23] is a comprehensive database that tracks trade-related policy measures implemented by nation-states around the world. It contains a wide range of measures, including tariff and non-tariff barriers, export taxes and subsidies, import measures, and other trade-related policies. It is updated in real-time and is provided as open data[3]. GTA provides information on the broader context of each measure, including the sectors and industries that are affected by their implementation, as well as the implementing and the affected jurisdictions. Overall, these aspects allow for analysing the impact of trade regulations on specific countries or regions, and to identify patterns and trends in trade policy over time on the global economy.

ACLED: The Armed Conflict Location & Event Data Project (ACLED) [24] is a non-profit organization that collects and analyses data on political violence and protest events across the world. It uses a combination of media monitoring, crowd-sourcing, and other open-source data collection methods to track and record information about incidents of political violence, including battles, bombings, riots, and protests. The organization's database provides information on the actors involved in each conflict, as well as the location, date, type, and intensity of the violence. ACLED is updated weekly and can be accessed via an API or downloaded as a data dump[4].

Relationships between GTA and ACLED: There are several dependencies between the ACLED dataset and the GTA dataset, e.g.:

- **ACLED events can lead to trade sanctions** If a country experiences political violence or conflict, other countries may respond by imposing trade sanctions or embargoes. For example, if a country is involved in a war, other countries may decide to stop trading with it. A recent example are the sanctions adopted from the European Union following Russia's military aggression against Ukraine. In this case, ACLED informs on the political violence that led to the sanctions, while GTA tracks the implemented trade policies.
- **Trade policies can exacerbate conflicts** Trade policies can exacerbate political conflicts or tensions between countries. For example, when one country imposes trade restrictions on another, it can result in economic hardship and political instability, potentially leading to conflicts. An instance of trade policies fueling conflicts is the US semiconductor sanctions against China.

[3] https://www.globaltradealert.org/

[4] https://acleddata.com/data-export-tool/

In this case, GTA informs on the trade policies that contributed to the conflict, while ACLED tracks the specific instances of violence or unrest.

- **ACLED events can disrupt trade flows** Political violence or unrest can disrupt trade flows between countries. For example, an attack on a major transportation hub could lead to delays or disruptions in trade. An example is the disruption of the Ukrainian grain trade with North Africa due to the war in Ukraine. In this case, ACLED informs on the incidents of violence that disrupted trade flows, while GTA tracks the affected trade policies or agreements.

Overall, using ACLED and GTA together provides a comprehensive picture of the relationship between political conflict and international trade. By analysing these datasets in tandem, policymakers and researchers can better understand the ways in which political violence and trade policies are interconnected, and develop more effective strategies for promoting peace and economic growth.

3.2 Steps for TKG Dataset Creation

We conduct several steps to create the present TKG to integrate the data into a structured and standardized framework.

Dataset Extraction and Mapping First, we retrieve the source data manually from the corresponding web services in a machine-readable format. Next, we convert the data into RDF format using an ontology schema that defines the relevant concepts, properties, and relationships. This enables the representation of the source data as a set of triples and allows for its integration with other RDF data sources. Both the ALCED and GTA datasets are mapped to RDF based on custom ontology declarations. These declarations specify how to transform the source data into triples and extend the central CoyPu COY ontology [17]. We provide the SPARQL CONSTRUCT queries used for the graph creation process in Appendix A.5, as well as in our repository.

Information Curation For GTA, we focus on including triples that contain information on countries and intervention sectors. We use the relations *IsImplementingInterventionOnSector* and *IsAffectedByInterventionOnSector* to depict dependencies between countries and intervention sectors. We exclude triples containing labels, intervention and state acts IDs, and event types, as they do not add value to our use case (e.g., including nodes representing IDs of certain interventions would not be beneficial and lead to a very sparse graph). For ACLED, we include information on the country location, types, and actors of each event. We utilize the relations *XAction* to link countries with ACLED events, and *XActIn* to connect countries based on the type of involved actors, where *X* represents the type of involved actors, e.g., StateForces or Civilians. Additionally, we employ the relation *IsAffectedByEventOfType* to establish the relationship between countries and ACLED events. Similar to GTA we exclude triples with comments and labels, as well as triples that contain ids of interventions and events.

Aggregation GTA uses the hierarchical industry classification schemes *CPC 2.1* and *HS 2012* to denote the affected sectors and products of an intervention. These schemes may include a very large number of categories, making the analysis more challenging. To address this issue, we use broader sector and product categories, based on higher-level groupings within the respective classification scheme. E.g., instead of considering each individual product category, we group products into broader categories based on their use or production process, such as "primary agricultural products". This is useful for modelling the impact of political violence or other events on trade flows, as it helps to identify the most affected sectors or products.

Merging GTA and ACLED The GTA and ACLED events can be linked via their annotated country information. In GTA, country data is available for the implementing and the affected jurisdiction of each intervention or state act. Meanwhile, in ACLED, country information is available from the locations of involved actors and of the ACLED events themselves.

Temporal Information From the graph in RDF format, we create a TKG with daily granularity, containing quadruples for each day in the year 2023 in TXT format. ACLED provides daily timestamps for each event. We create the quadruples by adding this timestamp data to all triples that are connected to this event. GTA provides an announcement date of each state act, as well as the implementation date, and - if existing - the removal date for each connected intervention. Since our use case aims to predict upcoming global trade events, we focus on the earliest available date for each GTA event. Therefore, we add the announcement date timestamp to all triples.

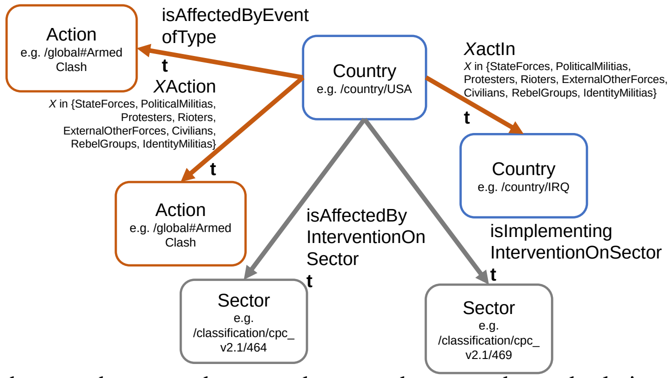

Figure 1: Data Schema, where grey boxes and arrows denote nodes and relations from GTA, orange boxes and arrows denote nodes and relations from ACLED, and blue boxes denote nodes present in both datasets, i.e. countries. Objects **t** illustrate timestamps.

3.3 Data Schema

Figure 1 gives an overview of our data schema. The TKG comprises three primary node types: nodes representing countries, economic sectors (GTA), actions within the context of ACLED events. These nodes are connected by 19 distinct relations. Consequently, the quadruples in the TKG can depict various scenarios: countries being affected by or implementing trade interventions on trade sectors; countries engaging in actions imposed by different actors (e.g., actions by civilians) in other countries; different actors (e.g., civilians) from countries conducting specific actions (e.g., protests); and countries being impacted by events of particular types (e.g., protests, armed clashes).

3.4 Dataset Properties

In this section, we analyze our TKG dataset in comparison to the existing TKG datasets introduced in Section 2. Our analysis aims to determine two key aspects: First, whether the dataset is suitable to apply TKG Forecasting to it with the goal of predicting upcoming global trade alert and crisis events based on historic events, i.e., its similarity to existing datasets; And second, whether it can serve as a valuable additional resource for TKG researchers in evaluating their methods.

Therefore, in Table 1, we report properties of our dataset, as well as of the datasets introduced in Section 2 for comparison, where our TKG is denoted **GTA_ACLED**. Besides commonly used characteristics such as number of entities (#N), relations (#R), and quadruples, and the timestep-based data splitting (short: #Tr/Val/Te TS), we compute the Recurrency Degree (Rec) and the Direct Recurrency Degree (DRec) introduced by [25]. The Recurrency Degree denotes the fraction of test quadruples (s, r, o, t) for which there exists a $k < t$ such that $(s, r, o, k) \in G$. Similarly, the Direct Recurrency Degree describes the fraction of test quadruples (s, r, o, t) for which it holds that $(s, r, o, t-1) \in G$.

GTA_ACLED exhibits a similar amount of relations and quadruples as the WIKI dataset. As compared to all other datasets, GTA_ACLED has less distinct nodes. Further, we highlight the very high Recurrency Degree value for GTA_ACLED, indicating that almost all temporal triples from the test set have occurred at previous timesteps. However, the Direct Recurrency Degree is notably lower as compared to the Rec, plummeting from 99.4 to 47.8. This contrasts with YAGO and WIKI, where the values for the DRec remain similarly high. This disparity signifies that for most triples in the test set, the same triple was not true at the direct previous step $t-1$. YAGO and WIKI consist mainly of facts which are true for certain time intervals, i.e. for multiple consecutive timesteps, with relations such as *works at* and *is married to*. In contrast, the quadruples in GTA_ACLED describe events such as Conflict Actions (Attacks, Protests) and Announcements of Trade Interventions.

Overall, the provided TKG exhibits notable differences from datasets commonly used in TKG Forecasting research. Notably, it features a lower number of nodes, resulting in a comparatively higher graph density. This denser structure may pose unique challenges and opportunities for modelling and forecasting techniques. Furthermore, while the triples within the dataset demonstrate a high level of recurrency, akin to YAGO and WIKI, they primarily describe individual events, like the ICEWS14, ICEWS18, and GDELT datasets.

Dataset	#N	#R	#Train	#Valid	#Test	Time Int.	#Tr/Val/Te TS	DRec [%]	Rec [%]
Rel. Work:									
ICEWS14	7128	230	74845	8514	7371	24 h	304/30/31	10.5	52.4
ICEWS18	23033	256	373018	45995	49545	24 h	239/30/34	10.8	50.4
GDELT	7691	240	1734399	238765	305241	15 m	2303/288/384	2.2	64.9
YAGO	10623	10	161540	19523	20026	1 y	177/5/6	92.7	92.7
WIKI	12554	24	539286	67538	63110	1 y	210/11/10	85.6	87.0
Ours:									
GTA_ACLED	451	19	515082	58430	36639	24 h	292/36/37	47.8	99.4

Table 1: We report dataset statistics like number of nodes and relations, the timestep interval, the specifics of the data splitting, the Recurrency Degree (Rec), and the Direct Recurrency degree (DRec) for the dataset introduced in this paper (GTA_ACLED) compared to datasets in related work.

Evaluating TKG Forecasting models on our dataset, in addition to established datasets, holds the potential to yield valuable insights for researchers in the field. By leveraging the unique characteristics of our dataset, researchers can explore new avenues for improving forecasting accuracy and understanding TKG dynamics. Despite the named differences, our analysis suggests that existing models designed for TKG Forecasting can be applied to the given dataset. In the subsequent section, we proceed to verify our expectation by applying existing TKG Forecasting models to our dataset and analyzing the obtained results.

Additional dataset statistics can be found in the appendix: Appendix A.1.1 provides information on relations, Appendix A.1.2 covers the graph's evolution over time, and Appendix A.1.3 presents further details on the distribution of the most prominent node types.

4 Experimental Setup

Over the past years, several models for TKG forecasting have emerged, employing diverse approaches such as combining Deep Graph Networks with sequential methods, harnessing Reinforcement Learning, and adopting Rule-based Approaches, as well as a baseline predicting facts based on their recurrence. We selected one representative model from each of these categories, choosing the models that demonstrate the best performance according to the evaluation conducted by [26]. Additionally, we compare the results to two static baselines, modifying the training loop to fit the temporal setting. This leads to the following selection:

- RE-GCN [11]: Combines Deep Graph Networks with sequential methods.
- TimeTraveler [12]: Uses Reinforcement Learning.
- TLogic [27]: Employs a rule-based approach.
- Recurrency Baseline [25]: A temporal baseline that predicts based on recurrency of facts.
- DistMult [28] and ComplEx [29]: As static Baselines.

Descriptions of the models utilized for our experiments are provided in Appendix A.1.4. For evaluating the models we follow the evaluation protocol introduced in [26]. We run experiments in single-step setting. We report the time-aware filtered Mean Reciprocal Rank (MRR), computing the average of the reciprocals of the ranks of the first relevant item in a list of results, as well as the Hits at k (H@k), with $k = \{1, 3, 10\}$, the proportion of queries for which at least one relevant item is among the top k ranked results. We run all experiments with one Nvidia Tesla V100 (32 GB) GPU, 128 GB Memory, and an Intel Xeon Silver 4114 CPU with 10 cores (20 threads). If not stated otherwise, for each of the above models we select the values of the hyperparameters based on the performance on the validation set. Find more information on hyperparameters in Appendix A.3. Dataset and Code are in our repository: `https://github.com/JuliaGast/GTA_ACLED_TKG/`.

5 Experimental Results

In this section we report Experimental results. We first show overall results on the full dataset, and then provide more detailed insights.

5.1 Overall Results

First, we report test scores for all models on the full test set in Table 2. The reported metrics serve as quantitative measures to assess the effectiveness of the predictive models on the given task. When comparing the individual models predictions, we see that a notable discrepancy emerges between the top-performing methods and the remainder. The Recurrency Baseline (RecB), which makes predictions solely based on fact recurrence, achieves the highest MRR of 67.8. RE-GCN, which combines structural and temporal information to compute temporal node embeddings, follows with an MRR of 65.0. In contrast, TLogic and Timetraveler, both of which rely on traversing temporal paths, perform notably worse on this task. Remarkably, the static Baselines DistMult and ComplEx achieve also high test scores, likely due to the frequent recurrence of facts within the dataset.

In the following we consider the results of the best performing model (RecB). The Hits@10 value 85.7% means that the correct answer is within the top 10 predictions 85.7% of the time. An MRR of 67.8% means that, on average, the reciprocal of the rank of the correct answer across all queries is 0.678. This implies that, on average, the correct answer is ranked relatively low (i.e. good) among the predictions made by the model. Overall, these results demonstrate a strong model performance in forecasting crisis events and trade interventions. However, the best-performing model is a baseline relying solely on fact recurrence. This suggests that the tested TKG models struggle to capture these simple recurrence patterns and hints at a strong need for performance improvement. Similar findings were reported in [25].

Method	MRR	H@1	H@3	H@10
DistMult	63.6	53.2	69.0	84.1
ComplEx	64.1	53.8	69.6	84.4
Timetraveler	48.4	43.0	50.7	57.5
TLogic	53.0	40.5	59.3	77.6
RE-GCN	<u>65.0</u>	<u>54.8</u>	<u>70.6</u>	<u>85.0</u>
RecB	**67.8**	**58.6**	**72.9**	**85.7**

Table 2: Test scores, where highest/second highest values are **bold**/<u>underlined</u>.

5.2 Detailed Results

Subsequently, we focus on detailed results. We report performance on GTA_ACLED vs. Other Datasets and example queries in Appendix A.2.

Type Consistency: We first evaluate whether the models can predict the correct node types. Specifically, we examine whether their top-1 predictions match the ground truth node type (e.g., predicting a country when the correct type is *country*). This serves as a sanity check for model predictions. We find that TLogic and RE-GCN predict the correct type for each test query, DistMult predicts a wrong type for one test query (ground truth *country*), and Timetraveler predicts wrong types for 281 test queries (ground truth *classification*).

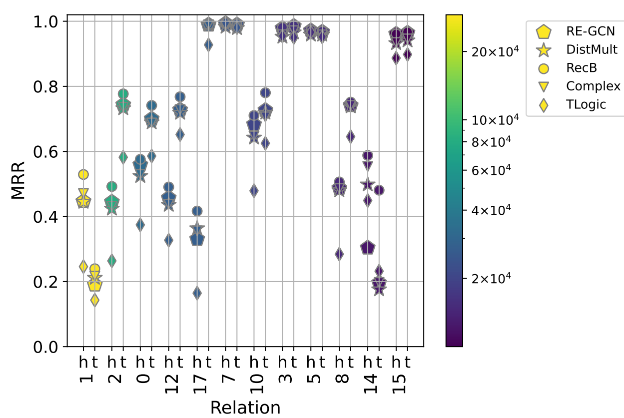

Figure 2: MRR per relation in head (h) and tail (t) direction for the five best-performing methods and 12 most prominent relations. The color indicates the number of triples with that relation. Relations 1 and 14 originate from the GTA dataset.

Performance per Relation: Figure 2 presents the MRR per relation for the 12 most frequent relations. A mapping from relation IDs to respective descriptions is in Figure 4 in Appendix A.1.1. In this analysis, the *head direction* refers to queries like (?, isAffectedByInterventionOnSector, cpc_v2.1/469, t^+), where the goal is to predict the affected or implementing country. In contrast, the *tail direction* refers to queries like (USA, isAffectedByInterventionOnSector, ?, t^+), aiming to predict the sectors affected by interventions in specific countries. A notable finding is the significant performance difference between head and tail prediction for many relations. For relation 1 (isAffectedByInterventionOnSector), models perform much better at predicting heads (i.e., affected countries) compared to tails (i.e., sectors). This is similar for relation 14 (isImplementingInterventionOnSector), the second relation from the GTA dataset. Conversely, in several ACLED-relations, such as relation 2

Method	(a) With ACLED MRR	H@1	H@3	H@10	(b) Without ACLED MRR	H@1	H@3	H@10
DistMult	<u>30.4</u>	<u>18.7</u>	<u>33.4</u>	<u>54.5</u>	29.8	17.4	33.8	54.8
ComplEx	29.5	17.8	32.3	53.3	<u>29.9</u>	<u>18.2</u>	<u>32.6</u>	<u>54.1</u>
Timetraveler	12.6	5.3	12.3	24.7	15.5	9.6	14.2	27.2
TLogic	19.9	7.6	23.0	43.1	19.9	7.5	23.3	43.1
RE-GCN	25.6	13.5	29.7	48.1	28.5	16.1	32.5	52.3
RecB	**36.0**	**24.4**	**39.5**	**60.3**	**36.0**	**24.4**	**39.5**	**60.3**

Table 3: Prediction scores for each method for relations of interest, i.e. for predicting GTA events only with (a) and without (b) information from ACLED triples.

(isAffectedByEventOfType), models perform better at predicting tails (i.e., event types) rather than heads (i.e., affected countries). This trend continues for relations 0 (StateForcesAction) and 12 (CiviliansAction), where predicting the action type yields higher MRRs than predicting the involved country. Queries related to action types (Country, XActIn, Country), with X representing different actions, achieve high MRRs across all relevant relations (relations 3, 5, 7, 15).

The performance across relations varies significantly between models. For instance, TLogic performs relatively well on relation 14 (isImplementingInterventionOnSector) but poorly on relation 1 (isAffectedByInterventionOnSector). In contrast, RE-GCN performs well on heads for relation 1 but struggles with heads for relation 14. These differences suggest that combining models with complementary strengths could improve performance. Nonetheless, the Recurrency Baseline consistently ranks among the top performers for most relations, indicating room for improvement in other methods.

Use Case Study: Predicting Trade Events On the given Use Case we are especially interested in predicting upcoming trade events. Therefore, we want to to assess the predictive capability of models regarding triples from the GTA dataset, specifically those with relations *isAffectedByInterventionOnSector* (1) and *isImplementingInterventionOnSector* (14) in both directions. For this, we train the models for the given Use Case: For methods with loss computation (DistMult, ComplEx, RE-GCN, Timetraveler), the loss is computed only on triples with the above relations. For TLogic, we only consider rules with these relations in their head. For the Recurrency Baseline no specific modifications are necessary, because predictions for each relation are made independently.

Results in Table 3(a) show scores for predicting links solely for the GTA relations. Compared to Table 2, there's a notable decrease in prediction scores, indicating the difficulty of predicting trade relations. Method performance aligns consistently across experiments, with those performing well overall also excelling in the targeted relations. Overall, the gap between MRR on the full dataset and on the GTA relations underscores the need for improved predictive capabilities in anticipating trade events.

Cross-dataset Dependencies: In addition to evaluating model performance on targeted relations, we assess the impact of ACLED data when predicting GTA events. Table 3 (b) presents the results for predicting GTA-related links after removing ACLED triples from the dataset. Interestingly, we observe varying effects on prediction scores across different methods. While DistMult benefits from the inclusion of ACLED data, omitting ACLED triples results in improved MRR scores for RE-GCN, ComplEx, and Timetraveler. The increase in MRR for the better-performing method indicates that integrating ACLED data alongside GTA contributes positively to prediction accuracy and underscores the value of synergizing both sources.

Performance over Time: Figure 3 (top) shows the MRR per timestep for the five best models. A consistent performance gap is evident, with TLogic underperforming at

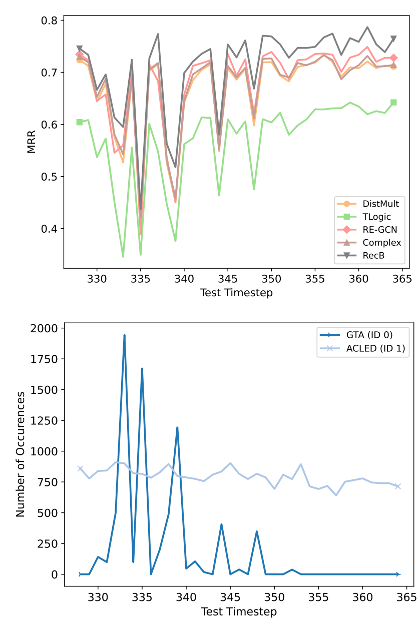

Figure 3: MRR (top) and Number of GTA and ACLED triples (bottom) per timestep of test dataset.

every timestep, which aligns with the overall results in Table 2. Figure 3 (bottom) illustrates the number of test triples from ACLED and GTA events at each timestep. We observe that at timesteps with a high concentration of GTA triples (e.g., timesteps 333 and 335), the performance of all models drops significantly. This pattern confirms earlier observations that models generally perform worse on GTA triples compared to ACLED ones. The irregularity of GTA interventions adds complexity, potentially making them more difficult for models to predict.

5.3 Result Summary

Our findings reveal that while the tested TKG models demonstrate potential in predicting future links, there is still significant room for improvement, particularly when compared to the Recurrency Baseline. Despite its simplicity, the baseline—predicting based on fact recurrence—outperforms more advanced models, underscoring the need for better refinement of TKG methods.

Among the non-baseline models, RE-GCN performs best, achieving an MRR of 65.0% and Hits@10 of 85%. However, the Recurrency Baseline, with an MRR of 67.8%, surpasses all tested models, including RE-GCN, highlighting the models' difficulty in capturing simple recurrence patterns critical to this task. Path- and rule-based approaches like TLogic and Timetraveler perform worse, with Timetraveler notably struggling to predict correct node types, suggesting considerable methodological flaws that warrant further investigation. We leave the investigation into these shortcomings for future research.

Furthermore, our analysis reveals that predicting links associated with upcoming trade events (GTA) poses a greater challenge compared to ACLED events, hinting at the need for further optimization of predictive capabilities. Interestingly, our results suggest that incorporating information from ACLED has the potential to enhance the predictive capabilities for forecasting links on GTA events.

In conclusion, while the Recurrency Baseline sets an unexpectedly high benchmark, models like RE-GCN display promise. Future work should focus on addressing the current models' inability to capture recurrence patterns as effectively as the baseline, while also exploring more advanced methods to improve performance, especially on challenging triples.

6 Conclusion and Future Work

In this paper we presented a novel approach to the analysis and prediction of global crises and trade patterns through the creation and exploration of a Temporal Knowledge Graph, and the application of TKG Forecasting. By combining data from the Armed Conflict Location & Event Data Project (ACLED) and Global Trade Alerts (GTA), the TKG serves as a basis for understanding the interplay between worldwide crises and trade dynamics over time. We detailed the TKG creation process, including the integration of data from multiple sources. Moreover, we conduct a dataset analysis to offer insights into its potential application in data-driven Resilience Research.

Future work could enhance these predictions by incorporating additional datasets that encompass various dimensions of global crises and trade, such as environmental and geographic data, thereby providing a more holistic understanding of influencing factors. Integrating further information on conflict severity could also be beneficial. Analyzing the advantages of incorporating longer-range data (e.g., spanning multiple decades) is another valuable direction, which could be facilitated by modifying the timespans in the provided scripts.

Enhancing model capabilities to predict entire future graph snapshots and implementing anomaly detection algorithms within the TKG would yield valuable insights for addressing fluctuations in data dynamics effectively. Furthermore, developing models that specifically account for long-range dependencies, as well as predicting time steps alongside future entities [30] are crucial future research directions. Ultimately, implementing a Real-Time Monitoring System using the TKG and forecasted outcomes could serve as a practical tool for policymakers and businesses to make informed decisions amid emerging global crises.

References

[1] Julia Gastinger, Nils Steinert, Sabine Gruender-Fahrer, and Michael Martin. Dynamic representations of global crises: Creation and analysis of a temporal knowledge graph for conflicts, trade and value networks. In *Second International Workshop on Linked Data-driven Resilience Research (D2R2'23) co-located with ESWC 2023*, 2023. 1

[2] World Economic Forum McKinsey & Company. Resilience for Sustainable, Inclusive Growth - White Paper, 2022. URL https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/resilience-for-sustainable-inclusive-growth. 1

[3] Leysia Palen, Sarah Vieweg, Jeannette Sutton, Sophia Liu, and Amanda Hughes. Crisis informatics: Studying crisis in a networked world. 01 2007. 1

[4] Margunn Indreboe Alshaikh and Helena Puig Larrauri. Building resilience through crisis mapping, community engagement and recovery planning in sudan. In *ISCRAM*, 2012. 1

[5] Robert Soden, Nama Budhathoki, and Leysia Palen. Resilience-building and the crisis informatics agenda: Lessons learned from open cities kathmandu. In *International Conference on Information Systems for Crisis Response and Management*, 2014. 1

[6] Robert Soden and Leysia Palen. Informating crisis: Expanding critical perspectives in crisis informatics. 2(CSCW), 2018. doi: 10.1145/3274431. URL https://doi.org/10.1145/3274431. 1

[7] Rakshit Trivedi, Hanjun Dai, Yichen Wang, and Le Song. Know-evolve: Deep temporal reasoning for dynamic knowledge graphs. In Doina Precup and Yee Whye Teh, editors, *ICML*, 2017. 2

[8] Zhen Han, Peng Chen, Yunpu Ma, and Volker Tresp. Explainable subgraph reasoning for forecasting on temporal knowledge graphs. In *ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021. 2

[9] Jon McClure, Simon Webb, and Grant McCool. The war in ukraine is fuelling a global food crisis., 2022. URL https://www.reuters.com/graphics/UKRAINE-CRISIS/FOOD/zjvqkgomjvx/. Accessed on Date: 15.03.2024. 2

[10] Woojeong Jin, Meng Qu, Xisen Jin, and Xiang Ren. Recurrent event network: Autoregressive structure inference over temporal knowledge graphs. *arXiv preprint arXiv:1904.05530*, 2019. preprint version. 2, 3

[11] Zixuan Li, Xiaolong Jin, Wei Li, Saiping Guan, Jiafeng Guo, Huawei Shen, Yuanzhuo Wang, and Xueqi Cheng. Temporal knowledge graph reasoning based on evolutional representation learning. In *SIGIR '21, Virtual Event, Canada*, pages 408–417. ACM, 2021. 2, 6, 17

[12] Haohai Sun, Jialun Zhong, Yunpu Ma, Zhen Han, and Kun He. Timetraveler: Reinforcement learning for temporal knowledge graph forecasting. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2021. 2, 6, 17

[13] Seonhyeong Kim and Young-Woo Kwon. Construction of disaster knowledge graphs to enhance disaster resilience. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 6721–6723, 2022. 2

[14] Sonja Tilly and Giacomo Livan. Macroeconomic forecasting with statistically validated knowledge graphs. *Expert Systems with Applications*, 186:115765, 2021. ISSN 0957-4174. 2

[15] Yucheng Yang, Yue Pang, Guanhua Huang, and Weinan E. The knowledge graph for macroeconomic analysis with alternative big data, 2020. URL https://arxiv.org/abs/2010.05172. 2

[16] Definition of the cidoc conceptual reference model version 5.0.4. Technical report, ICOM/CIDOC CRM Special Interest Group, 11 2011. 3

[17] Coypu coy ontology, Feb 2023. URL https://gitlab.com/coypu-project/coy-ontology. 3, 4

[18] Elizabeth Boschee, Jennifer Lautenschlager, Sean O'Brien, Steve Shellman, James Starz, and Michael Ward. ICEWS Coded Event Data, 2015. URL https://doi.org/10.7910/DVN/28075. 3

[19] Alberto García-Durán, Sebastijan Dumančić, and Mathias Niepert. Learning sequence encoders for temporal knowledge graph completion. In *EMNLP 2018*, pages 4816–4821, Belgium, 2018. 3

[20] Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, pages 1–49. Citeseer, 2013. 3

[21] Farzaneh Mahdisoltani, Joanna Asia Biega, and Fabian M. Suchanek. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*, 2015. 3

[22] Julien Leblay and Melisachew Wudage Chekol. Deriving validity time in knowledge graph. In *Companion proceedings of the the web conference 2018*, pages 1771–1776, 2018. 3

[23] Simon J. Evenett. Global trade alert: Motivation and launch. *World Trade Review*, 8(4):607–609, 2009. doi: 10.1017/S1474745609990061. 3

[24] C. Raleigh, A. Linke, H. Hegre, and J. Karlsen. Introducing acled: An armed conflict location and event dataset. *Journal of Peace Research*, 47(5):651–660, 2010. URL `https://doi.org/10.1177/0022343310378914`. 3

[25] Julia Gastinger, Christian Meilicke, Federico Errica, Timo Sztyler, Anett Schuelke, and Heiner Stuckenschmidt. History repeats itself: A baseline for temporal knowledge graph forecasting. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI 2024, Jeju, South Korea*, 2024. 5, 6, 7, 17, 18

[26] Julia Gastinger, Timo Sztyler, Lokesh Sharma, Anett Schuelke, and Heiner Stuckenschmidt. Comparing apples and oranges? On the evaluation of methods for temporal knowledge graph forecasting. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD)*, pages 533–549, 2023. 6, 17, 18

[27] Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. TLogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In *36th Conference on Artificial Intelligence (AAAI)*, pages 4120–4127, 2022. 6, 17

[28] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. In *3rd International Conference on Learning Representations, ICLR*, 2015. 6, 17

[29] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. In *International conference on machine learning*, pages 2071–2080. PMLR, 2016. 6, 17

[30] Roxana Pop and Egor V Kostylev. Inductive future time prediction on temporal knowledge graphs with interval time. In *NeSy*, pages 233–240, 2023. 9

A Appendix

A.1 Additional Dataset Statistics

In this section, we provide further insights into the dataset. First, we present information on the relations within the graph, followed by statistics on the graph over time, and conclude with details about the node types.

A.1.1 Relation Analysis

Figure 4 presents a pie chart illustrating the distribution of relations in the dataset. It shows that 48% of the triples contain the relation gta#isAffectedByInterventionOnSector. As discussed in Section 5.2, certain timesteps exhibit an exceptionally high number of interventions. The figure also indicates that there are significantly more triples of type gta#isAffectedByInterventionOnSector compared to gta#isImplementingInterventionOnSector. This is because a single country can implement the same intervention across multiple countries simultaneously, such as interventions targeting all EU member states. Additionally, the different types of ACLED actions (StateForcesAction, CiviliansAction, ...) show a relatively small variation in distribution, ranging from 4% down to about 1%.

Figure 5 illustrates the number of occurences per timestep per relation. The numbers are aggregated to show the mean, maximum, and minimum values for each block. The figure highlights that the high percentage of triples for the relation gta#isAffectedByInterventionOnSector (Relation 1) is driven by four outlier timesteps with exceptionally high numbers of triples. A similar outlier is observed for gta#isImplementingInterventionOnSector (Relation 14). For most other relations, the number of triples remains relatively stable over time, with significantly smaller variations between the minimum and maximum values. Relations 10 (acled#ExternalOtherForcesAction) and 15 (acled#ExternalOtherForcesActIn) show a slight increase in the number of triples over time.

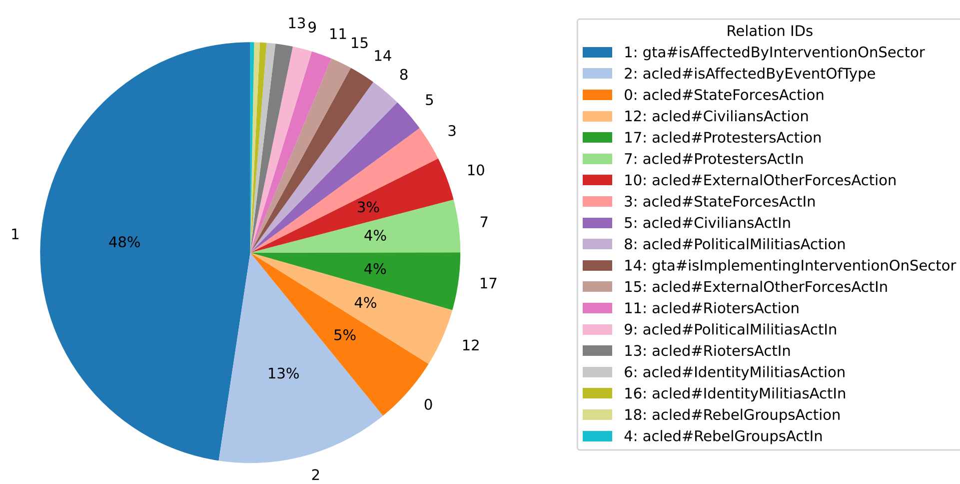

Figure 4: Distribution of relations in the dataset: pie chart.

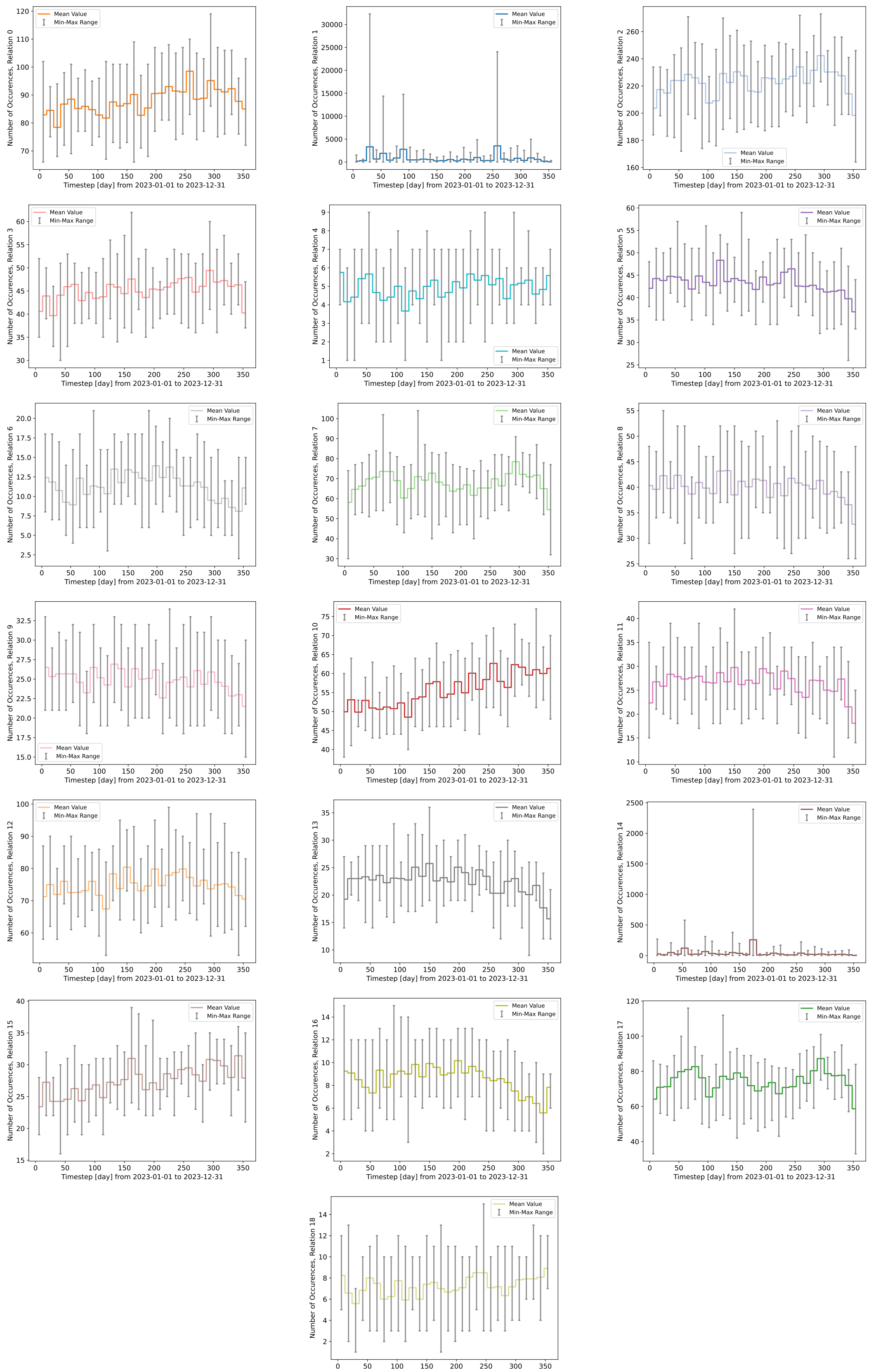

Figure 5: Number of Occurences over time per relation.

A.1.2 Graph Analysis over time

Figures 6, 7, and 8 present the number of nodes, triples, and the average node degree per timestep in the dataset. The values are aggregated, displaying the mean, minimum, and maximum for each block. Figure 9 illustrates the percentage of occurrences for ACLED and GTA triples. The number of nodes ranges from approximately 100 to 400 per timestep, while the graph typically contains around 1000 triples per timestep on average. However, four outlier timesteps exhibit exceptionally high values, with up to 34000 triples and an average node degree of 170. This observation aligns with the findings in Section A.1.1, where we note that these timesteps correspond to a large number of GTA interventions. Methods applied to this graph must be capable of handling the high number of triples during these outlier timesteps.

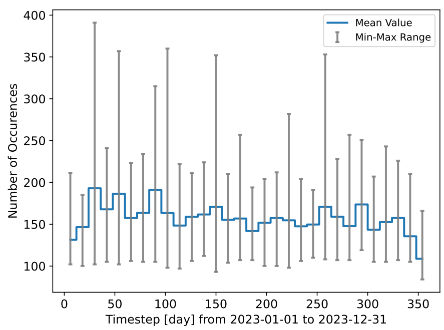

Figure 6: Number of nodes over time, aggregated.

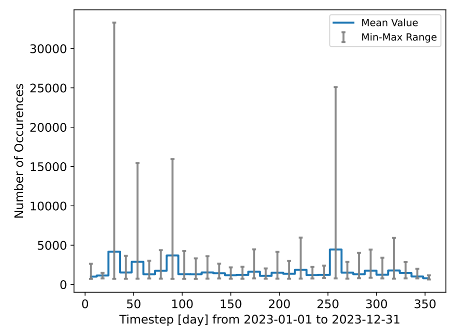

Figure 7: Number of triples over time, aggregated.

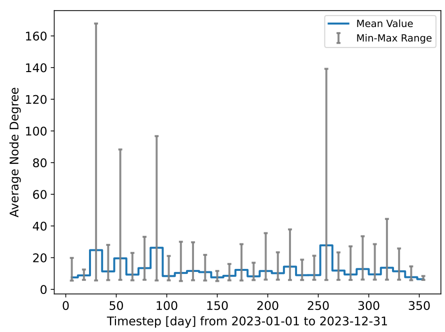

Figure 8: Average Node Degree over time, aggregated.

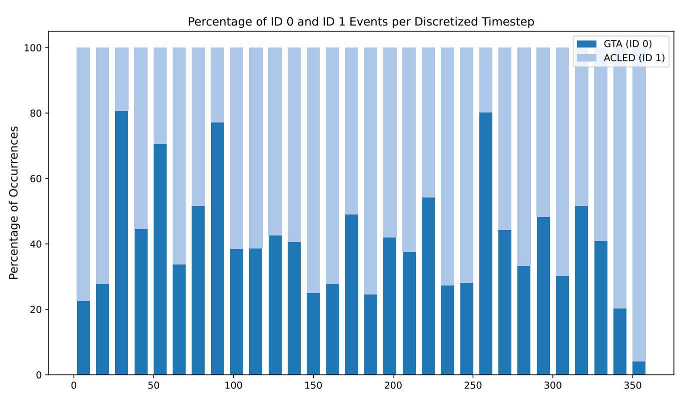

Figure 9: Percentage of ACLED/GTA triples over time, aggregated.

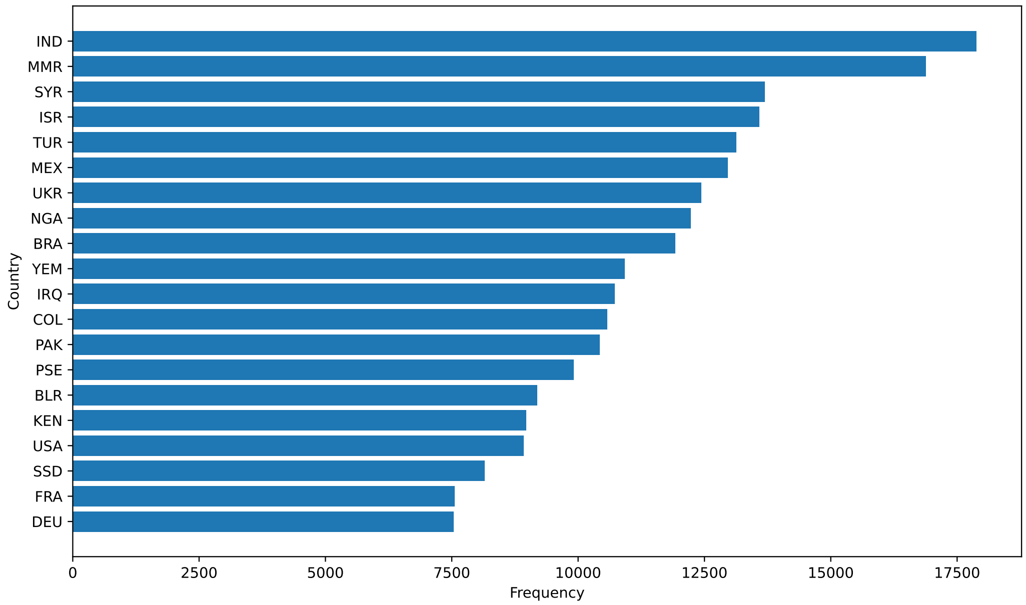

Figure 10: The 20 most Prominent Countries and their triple counts.

A.1.3 Node Analysis

In the following, we illustrate the distribution of nodes in different categories. Figure 10 illustrates the 20 most prominent countries in the graph, along with their frequencies, i.e., the number of triples they appear in. We observe that India and Myanmar are the most prominent, each with over 15000 occurrences, followed by Syria and Israel. Figure 11 shows the 20 most prominent GTA intervention sectors and their frequencies. The sectors "Other fabricated metal products" and "Cereals" are the most prominent, each with over 6000 occurrences. Lastly, Figure 12 displays the 20 most prominent ACLED events and their frequencies. "PeacefulProtest" is the most frequent event, with over 45000 occurrences, followed by "Attack" and "ArmedClash." There is significant variation, with the highest frequency exceeding 45000 occurrences and the lowest falling below 4000 occurrences, indicating an unequal distribution of ACLED event types in the graph. While events like "NonStateActorOvertakesTerritory" and "SexualViolence" have relatively low occurrences, these types of events might represent critical moments of conflict and escalation. This disparity highlights the need for a nuanced understanding of conflict dynamics, as both frequent and infrequent events can have profound implications for global trade and conflict.

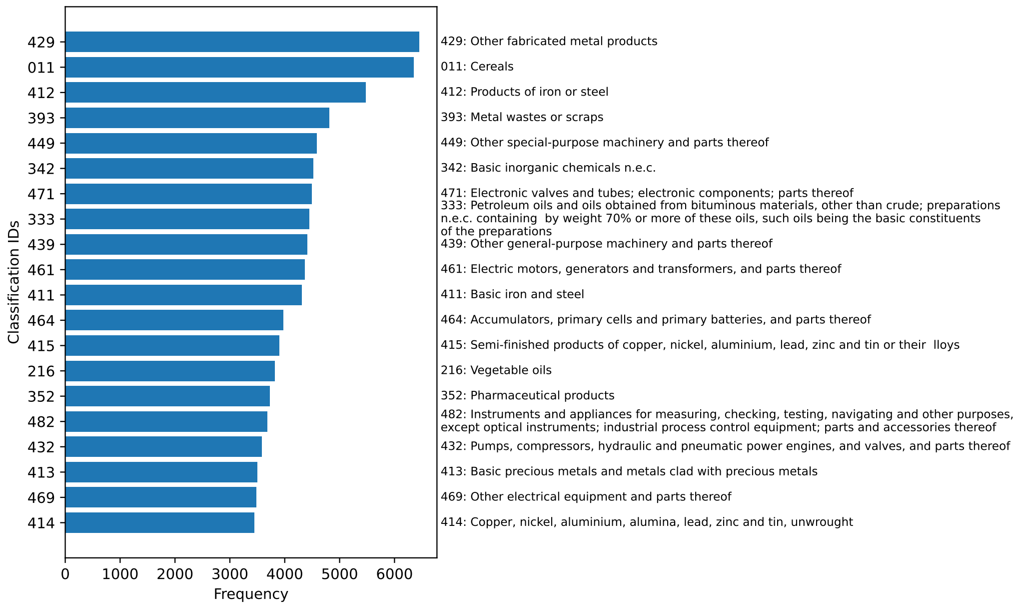

Figure 11: The 20 most Prominent GTA Intervention Sectors and their triple counts.

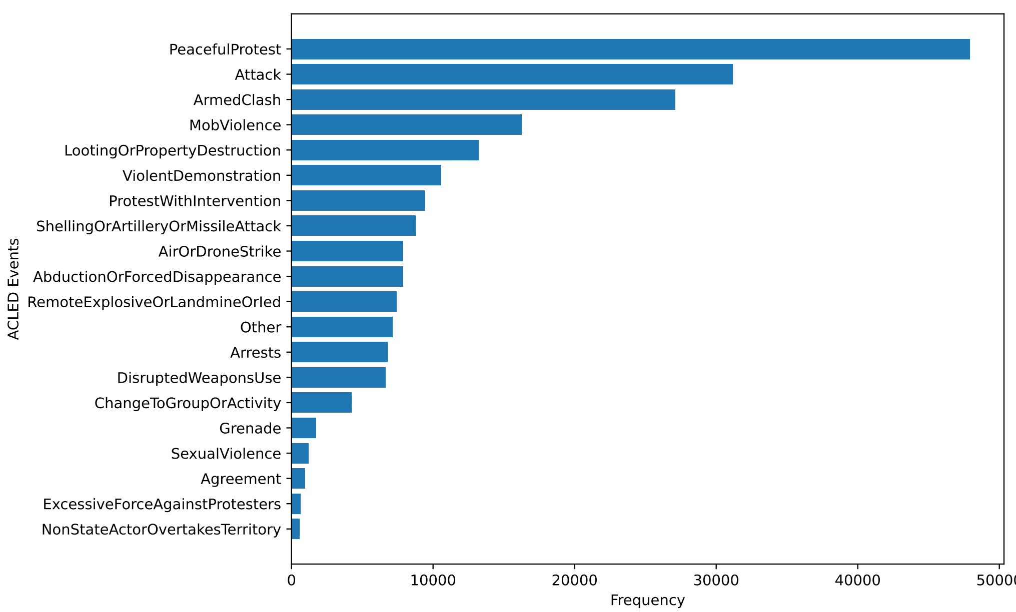

Figure 12: The 20 most Prominent ACLED events and their triple counts.

A.1.4 Details on Models

RE-GCN: This model [11] combines message-passing with a sequential neural network to integrate both structural and sequential information. It recurrently models the sequence of Knowledge Graph snapshots by combining a convolutional graph Neural Network with a sequential Neural Network model. RE-GCN also features a static graph constraint to incorporate additional information such as entity types.

TimeTraveler: TimeTraveler [12] utilizes a Reinforcement Learning model based on temporal paths. The agent starts walks through outgoing edges across timesteps. Actions in TimeTraveler are sampled based on transition probabilities. These probabilities are derived from dynamic embeddings of the query, history of paths, and candidate actions. The model employs a time-shaped reward.

TLogic: TLogic [27] is a symbolic framework that acquires temporal logic rules through temporal random walks, exploring edges in the graph backward in time. These rules are subsequently applied to events preceding the query. When scoring answer candidates, TLogic considers both the confidence levels of the rules and time differences.

Recurrency Baseline (RecB): The Recurrency Baseline![25] is a baseline method that makes predictions based on recurring facts, combining scores from two sources: strict recurrency, which factors in both the recency and frequency of facts, and relaxed recurrency, which considers partial recurrences within a query.

DistMult Static Baseline: We adapt and employ DistMult [28] as a model commonly utilized for link prediction in static KG completion.[5] We modify the training loop to fit the temporal setting, i.e., the training and validation losses for this model are computed on a per-snapshot basis.

ComplEx Static Baseline: Additionally, we adapt and employ ComplEx [29] as another model commonly utilized for link prediction in static KG completion. ComplEx makes use of complex valued embeddings. In the same way as DistMult, we modify the training loop to fit the temporal setting.

A.2 Additional Results

A.2.1 Example Queries

To highlight one of the top-performing model's predictive prowess, Table 4 showcases randomly selected predictions from the test set, with correct ones bolded. These exemplify RE-GCN's accuracy in forecasting diverse ACLED events, including Peaceful Protests and Protests with Interventions, as well as its adeptness in predicting event locations. Furthermore, the model demonstrates its capability to forecast countries affected by interventions in specific sectors. While these examples offer only a glimpse into the model's overall performance, they effectively illustrate its predictive ability.

A.2.2 Other Datasets

In Table 5, we show experimental results reported by [26] on the datasets used in related work (described in Section 2) for the methods employed in our experiments. We report mean reciprocal rank (MRR) and Hits@k (Hk) with $k = 1, 3, 10$ in a time-aware filter setting. The MRRs range from as low as 24.5 for the GDELT dataset to as high as 90.9 (YAGO) and 82.3 (WIKI). No single method outperforms the others across all datasets. We observe significant variance in model performance across datasets, indicating differing levels of difficulty. With an average MRR of 60.3, the GTA_ACLED dataset presents a medium level of challenge. Furthermore, there is no clearly superior method for the other datasets: RE-GCN excels on ICEWS18, while TLogic achieves the highest MRR on WIKI and ICEWS14. Notably, the competitive performance of the Recurrency Baseline (RecB.), which ranks among the top for three out of five datasets, is striking [25]. Additionally, the difference between the best and worst performing models for each dataset is significantly smaller than for GTA_ACLED. This suggests that GTA_ACLED presents unique challenges that only some methods can effectively address, unlike the other datasets where performance is more uniformly distributed across models.

[5] As of the present date, DistMult does not represent the state-of-the-art for static KG completion. However, we chose it due to its simplicity and adaptability to the temporal snapshot setting. Therefore, it serves as a practical static baseline against which the temporal models can be compared.

subject	relation	ground truth object	predicted object	date
country/SYR	acled#ProtestersAction	global#Peaceful-Protest	**global#PeacefulProtest**	2023-10-20
country/SSD	acled#StateForces-ActIn	country/SDN	**country/SDN**	2023-10-20
country/JPN	acled#ProtestersActIn	country/JPN	**country/JPN**	2023-10-20
country/IND	acled#PoliticalMilitias-Action	global#Disrupted-WeaponsUse	global#Attack	2023-10-20
country/FRA	acled#ExternalOther-ForcesAction	global#Protest-WithIntervention	**global#Protest-WithIntervention**	2023-10-20
classification /cpc_v2.1/464	inv_gta#isAffectedBy-InterventionOnSector	country/SGP	**country/SGP**	2023-10-22
classification /cpc_v2.1/469	inv_gta#isAffectedBy-InterventionOnSector	country/SWE	country/CHN	2023-10-22

Table 4: Example queries from the test set, comparing ground truth objects and RE-GCN's highest-ranked predicted objects. Bold entries mark correct predictions.

	GDELT				YAGO				WIKI			
	MRR	H1	H3	H10	MRR	H1	H3	H10	MRR	H1	H3	H10
RE-GCN	19.8	12.5	21.0	33.9	82.2	78.7	84.2	88.5	78.7	74.8	81.7	84.7
TLogic	19.8	12.2	21.7	35.6	76.5	74.0	78.9	79.2	**82.3**	78.6	86.0	87.0
Timetr.	20.2	14.1	22.2	31.2	87.7	84.6	90.9	91.2	78.7	75.2	82.0	83.1
RecB.	**24.5**	n.r.	n.r.	**39.8**	**90.9**	n.r.	n.r.	**93**.0	81.5	n.r.	n.r.	**87.1**

	ICEWS14				ICEWS18			
	MRR	H1	H3	H10	MRR	H1	H3	H10
RE-GCN	42.1	31.4	47.3	**62.7**	**32.6**	**22.4**	**36.8**	**52.6**
TLogic	**42.5**	**33.2**	**47.6**	60.3	29.6	20.4	33.6	48.1
Timetr.	40.8	31.9	45.4	57.6	29.1	21.3	32.5	43.9
RecB.	37.2	n.r.	n.r.	51.8	28.7	n.r.	n.r.	43.7

Table 5: Experimental results for single-step predictionwith datasets GDELT, YAGO, WIKI (top), and ICEWS14, ICEWS18 (bottom) [26], [25]. The best results for each dataset are marked in bold. "n.r." means, these values were not reported.

A.3 Hyperparameters

We report the selected hyperparameter values in Table 6. Further, our GitHub repository contains the hyperparameter ranges for each method. If not stated otherwise, for each of the above models we select the values of the hyperparameters based on the performance on the validation set, more specifically based on the highest validation MRR. We therefore, for each model, select among the hyperparameter ranges specified in the original works. For TLogic, we set the highest value for each of the hyperparameters that is feasible with our hardware without memory errors. We report the selected hyperparameters in Appendix A.3.

A.4 The CoyPu Coy Ontology

The CoyPu Ontology, a semantic framework designed to model environmental and socio-economic data relevant to supply chain sustainability, facilitates interoperability between datasets by providing a standardized structure for environmental indicators, emissions, crises and economic activities. It is used to describe countries or companies and their supply network, infrastructure, products, production materials, industries, events, and relations. Its ability to represent complex macro-analytical, multi-relational data over time complements the temporal dimensions of our TKG model. The full ontology is available at `https://gitlab.com/coypu-project/coy-ontology`.

Method	Hyperparameter Values
DistMult	$dropout = 0.3$, $embedding\_dim = 200$, $weight\_decay = 0.0002$, $learn\_rate = 0.0001$
TLogic	$rule\_lengths = 1$, $window = 0$, $top\_k = 20$
RE-GCN	$n\_hidden = 200$, $n\_layers = 1$, $dropout = 0.4$, $n\_bases = 100$, $train\_history\_len = 7$
Timetraveler	$max\_action\_num = 80$, $ent\_dim = 80$, $mu = 0.1$, $path\_length = 3$, $lr = 0.00001$

Table 6: Selected hyperparameter values for each method.

A.5 Details on SPARQL queries

In the following, we list and describe the SPARQL queries used to create the TKG from raw data. The queries can also be found in our GitHub repository.

A.5.1 Query 1: Retrieving ACLED Events and Related Triples for the Year 2023

```
PREFIX coy: <https://schema.coypu.org/global#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

CONSTRUCT {
  ?e ?p ?o .
  ?o ?p1 ?o1 .
} WHERE {
  {
    SELECT DISTINCT ?e WHERE {
      graph <https://data.coypu.org/events/acled/> {
        ?e rdf:type coy:Event ;
           coy:hasTimestamp ?t .
        BIND(year(?t) as ?year)
        FILTER(?year = 2023)
      }
    }
  }
  ?e ?p ?o .
  OPTIONAL { ?o ?p1 ?o1 }
}
```

This query retrieves all ACLED events that occurred in the year 2023 and constructs triples related to these events. The query first selects events from the ACLED dataset where the timestamp is in 2023. For each event, it retrieves all direct triples (`?e ?p ?o`) and optionally retrieves one level of connected data (`?o ?p1 ?o1`). The `CONSTRUCT` clause then assembles these triples.

A.5.2 Query 2: Retrieving ACLED Events with Actors and Specific Properties for 2023

```
PREFIX coy: <https://schema.coypu.org/global#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

CONSTRUCT {
  ?e a ?type .
  ?e ?p ?o .
  ?e coy:hasActor ?actor . ?actor ?p_actor ?o_actor .
} WHERE {
  {
    SELECT DISTINCT ?e ?type WHERE {
```

```
      graph <https://data.coypu.org/events/acled/> {
        ?e rdf:type coy:Event ;
           rdf:type ?type ;
         coy:hasTimestamp ?t .
        BIND(year(?t) as ?year)
        FILTER(?year = 2023)
        FILTER(?type not in (coy:Event, coy:Conflict))
      }
    }
  }

  ?e ?p ?o
  FILTER(?p IN (coy:hasCountryLocation, coy:hasFatalities, coy:hasTimestamp))
  OPTIONAL {
   ?e coy:hasActor ?actor . ?actor ?p_actor ?o_actor
  }
}
```

This query constructs a graph of ACLED events from 2023 with specific properties and actor data. It focuses on events that have a type other than the general `coy:Event` or `coy:Conflict`, narrowing the results to more specific event types (e.g., protests). Only certain properties are retrieved, such as the country location, number of fatalities, and timestamp. Additionally, for each event, data about the actors involved is fetched and related triples about the actors are included.

A.5.3 Query 3: Aggregating State Acts for 2023 with Product and Sector Data

```
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gta: <https://schema.coypu.org/gta#>

INSERT {
    GRAPH <https://data.coypu.org/gta/2023/> {
      ?a a gta:StateAct ;
       gta:hasAnnouncementDate ?date ;
       gta:hasIntervention ?i .

      ?i a gta:Intervention ;
            gta:hasAffectedCommercialFlow  ?i_flow ;
            gta:hasGTAEvaluation           ?i_eval ;
            gta:hasImplementationDate      ?i_date ;
            gta:hasImplementationLevel     ?i_level ;
            gta:hasImplementingJurisdiction ?i_jurisdiction ;
            gta:hasInterventionType        ?i_type ;
            gta:hasAffectedSector          ?i_sector ;
            gta:hasAffectedProduct         ?i_product .
    }
    } WHERE {
      ?a a gta:StateAct ;
         gta:hasAnnouncementDate ?date ;
         gta:hasIntervention ?i .
      FILTER(year(?date) = 2023)

      ?i a gta:Intervention ;
         gta:hasAffectedCommercialFlow  ?i_flow ;
         gta:hasGTAEvaluation           ?i_eval ;
         gta:hasImplementationDate      ?i_date ;
         gta:hasImplementationLevel     ?i_level ;
         gta:hasImplementingJurisdiction ?i_jurisdiction ;
         gta:hasInterventionType        ?i_type ;
```

```
            gta:hasAffectedSector          ?i_sector ;
            gta:hasAffectedProduct         ?i_product .

    # aggregate the products
    WITH <https://data.coypu.org/gta/2023/>
    DELETE {
    ?i gta:hasAffectedProduct ?p ;
    }
    INSERT {
        ?i gta:hasAffectedProduct ?pp
        } WHERE {
              ?i a gta:Intervention ;
                 gta:hasAffectedProduct ?p .
              GRAPH <https://data.coypu.org/sectors/hs2012/> {?p skos:broader ?pp .}
    };

    # aggregate the sectors
    WITH <https://data.coypu.org/gta/2023/>
    DELETE {
    ?i gta:hasAffectedSector ?s ;
    }
    INSERT {
        ?i gta:hasAffectedSector ?s
        } WHERE {
              ?i a gta:Intervention ;
                 gta:hasAffectedSector ?s .
              GRAPH <https://data.coypu.org/products/cpc21/> {?s skos:broader ?ss . }
    };
}
```

This query inserts state acts from 2023, focusing on interventions that affected commercial flows, sectors, and products. It constructs triples for state acts and their interventions, including properties such as implementation dates, jurisdictions, and evaluations. The purpose is to gather structured information about trade interventions and their economic impacts. It later aggregates product and sector data into higher-level categories using a hierarchy (`skos:broader`), reducing the dataset's granularity while preserving the overall structure.

A.5.4 Query 4: Annotating Interventions with Announcement Date

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX gta: <https://schema.coypu.org/gta#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

CONSTRUCT {
 ?edge1 gta:hasAnnouncementDate ?date .
 ?edge2 gta:hasAnnouncementDate ?date .
} WHERE {
  GRAPH <https://data.coypu.org/gta/> {
   ?a a gta:StateAct .
   ?a gta:hasAnnouncementDate ?date .
   ?a gta:hasIntervention ?i .
   ?i gta:hasAffectedSector ?s .
   GRAPH <https://data.coypu.org/products/cpc21/> {?s skos:broader ?ss . }

   ?i gta:hasImplementingJurisdiction ?ij .
   ?i gta:hasAffectedJurisdiction ?aj .

   BIND(<< ?ij gta:isImplementingInterventionOnSector ?ss >> AS ?edge1)
   BIND(<< ?aj gta:isAffectedByInterventionOnSector ?ss >> AS ?edge2)
```

```
  }
}
```

This query creates new triples linking interventions (and their impact on different sectors) with the corresponding announcement dates. Specifically, it generates: One edge indicating that a jurisdiction is implementing an intervention on a broader sector (?edge1) and one edge indicating that a jurisdiction is affected by an intervention on a broader sector (?edge2). Both edges are annotated with the announcement date of the state act.