Dynamic Representations of Global Crises: Creation And Analysis of a Temporal Knowledge Graph For Conflicts, Trade and Value Networks

Julia Gastinger[1,2,*], Nils Steinert[3], Sabine Gründer-Fahrer[4] and Michael Martin[4]

[1]NEC Laboratories Europe GmbH, Kurfuersten-Anlage 36, 69115 Heidelberg, Germany
[2]University of Mannheim, Data and Web Science Group, B 6, 26, 68159 Mannheim, Germany
[3]Implisense GmbH, Spiekermannstraße 31a, 13189 Berlin, Germany
[4]Institute for Applied Informatics (InfAI), Goerdelerring 9, 04109 Leipzig, Germany

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). Combining data from the Armed Conflict Location & Event Data Project (ACLED) and Global Trade Alerts (GTA), the tKG provides a comprehensive view of the intersection between worldwide crises and global trade over time. The paper details the process of creating the tKG, including the aggregation and merging of information from multiple sources. Additionally, the paper offers insights into the analysis of the tKG and its potential applicability to data-driven Resilience Research. As an initial application, the tKG can be used to predict global trade events, such as trade sanctions across various categories and countries, based on global conflict events, to identify potential trade disruptions and anticipate the economic impact of global conflicts.

Keywords

Temporal Knowledge Graphs, Knowledge Graphs, Resilience Research, Crisis Research

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 climate change, supply chain disruptions, and healthcare availability, define a new era where "managing disruptions defines sustainable growth more than managing continuity" [1]. 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.

Second International Workshop on Linked Data-driven Resilience Research (D2R2'23) co-located with ESWC 2023, May 28th, 2023, Hersonissos, Greece
*Corresponding author.
julia.gastinger@neclab.eu (J. Gastinger); nils.steinert@implisense.com (N. Steinert); gruender-fahrer@infai.org (S. Gründer-Fahrer); martin@infai.org (M. Martin)

Resilience research is a challenging but urgently needed scientific field which will contribute to solving urgent societal issues. In response to this urgent need, researchers in various fields, including information and communications technology, data science, and artificial intelligence (compare e.g., [2], [3], [4], [5]), have made significant contributions to resilience and crisis research.

The CoyPu project [6] aims to increase the transparency of value chains and the understanding of complex mechanisms of crisis factors at a global scale by using semantically represented data and AI analytics. Through a large consortium of partners, the project integrates, models, and analyzes huge amounts of data to build a new basis for situational awareness and decision making, as well as for the elaboration of advanced resilience strategies. In the context of a future CoyPu platform, semantic technologies such as RDF, OWL, and SPARQL combine data interoperability and "cross-silo" queries with decentralized storage. The CoyPu Knowledge Graph provides macro-economically relevant and market-specific data, as well as information on current global crisis and conflict events, which can be integrated with external data on an ad-hoc basis. This paper focuses on the subset of trade-related policy measures, sanctions, and political violence and conflicts within the CoyPu Knowledge Graph.

Temporal Knowledge Graphs (tKG) are Knowledge Graphs (KG) where facts occur, recur, or evolve over time [7]. Triples are extended with timestamps to indicate that they are valid at a given time, allowing to hold time-evolving multi-relational data [8]. Because they are not only able to represent the interconnectivity of systems, but also their dynamic evolvement, 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. The research field of tKG forecasting predicts facts at future timesteps based on a history of a KG [3]. 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 2021, providing a comprehensive view of the dynamic relationships of these events.

1.1. Use Case

By using the presented tKG for downstream analysis and learning tasks, we can identify patterns and predict future developments in the global landscape of crisis and trade sanctions. Specifically, we propose the use of tKG forecasting (see e.g., [9], [10], [8]) to predict upcoming global trade alert events and their links based on previous crisis events. This research opens up new possibilities for understanding the complex interactions between global crises and trade sanctions and lays the groundwork for future studies in this field.

1.2. Structure and Contribution

This paper first provides an overview of existing work on KG and vocabularies for resilience research and prevalent tKG datasets (Section 2), along with an overview of the utilized dataset resources (Section 3). Further, we describe our approach for creating the tKG (Section 4). To understand the properties of the created tKG, we perform a technical analysis and visualize a selected graph snapshot, providing insights for tKG Forecasting (Section 5). We conclude with an outlook describing the potential usage of this tKG in further research, as well as specifically in the CoyPu project (Section 6). We publish the tKG dataset and associated code for creating and analysing the tKG[1].

2. Related Work

2.1. Knowledge Graphs and Vocabularies for Resilience Research

The use of Knowledge Graphs in resilience and crisis research has gained increasing attention in recent years [11]. KG offer a flexible and comprehensive approach to modeling and analysing complex systems [12], making them suitable for a wide range of domains, including macroeconomical analysis [13], which is a main focus of the CoyPu research project.

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 [14] 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, in the CoyPu project a custom ontology - the CoyPu COY ontology [15] - was developed to model the KG.

2.2. Temporal Knowledge Graph Datasets

Zhang et al. [16] provide a comprehensive overview of existing temporal RDF models. We follow the work of Trivedi et al. [7], Li et al. [10], Han et al. [8], and others, who represent tKG as sequences of timestamped KG. A timestamped KG, or KG snapshot, denoted as $G_t = \{V, R, \mathcal{E}_t\}$, captures the state of the tKG at a specific timestep t, where V is the set of entities, R is the set of relations, and $\mathcal{E}_t$ is the set of quadruples [10]. A quadruple consists of four elements, such as *(Event A, hasActor, French Police Forces, 2021-07-01)*.

In the domain of tKG analysis, six datasets have been published and utilized, including different versions of the Integrated Crisis and Early Warning System (ICEWS) [17]: ICEWS05-15 [18], ICEWS14 [18], and ICEWS18 [9] (the numbers describe the respective years); GDELT [19]; YAGO [20]; and WIKI [21] (preprocessed by Jin et al. [9]). Notably, the three versions of ICEWS cover the crisis topic, demonstrating the applicability of tKG to crisis research. However, to the best of our knowledge, no tKG currently exists that describes trade relations and sanctions over time. Additionally, none of the existing tKG merge data from multiple sources to provide a comprehensive view or analyse the interconnection of different event types. Finally, to our knowledge, no other study has analysed the evolution of graph properties over time for tKG.

[1]https://github.com/GastJulia/TKG-ACLED-GTA-Dataset

3. Resources

3.1. GTA

The Global Trade Alerts (GTA) dataset [22] is a comprehensive database that tracks trade-related policy measures implemented by nation-states around the world since 2008. The dataset 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[2].

One of the key strengths of the GTA dataset is its focus on the affected jurisdictions, providing details on both the implementing and the affected jurisdiction for each measure. Additionally, it covers measures that impact the flow of goods and services across borders, such as taxation and exim quotas. Moreover, GTA provides information on the broader context of each measure, including the sectors and industries that are most affected by their implementation, as well regulatory political and economic factors that may be driving changes in trade policy. 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.

3.2. ACLED

The Armed Conflict Location & Event Data Project (ACLED) [23] is a non-profit organization that collects and analyses data on political violence and protest events across the world.

ACLED 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 covers more than 200 countries and provides information on the actors involved in each conflict, as well as the location, date, type, and intensity of the violence. The dataset is updated weekly and can be accessed via an API or downloaded as a data dump[3].

3.3. Relationships between GTA and ACLED

There are several possible relationships and 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 civil war, other countries may decide to stop trading with it. 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 sometimes exacerbate political conflicts or tensions between countries. E.g., the trade restrictions that one country imposes on another could lead to economic hardship and political instability, which could

[2] https://www.globaltradealert.org/
[3] https://acleddata.com/data-export-tool/

in turn lead to conflicts. 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. 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 the ACLED and GTA datasets 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.

4. Method and Implementation

The creation of the present temporal Knowledge Graph, which comprises a subset of the larger CoyPu KG, involves several steps to integrate the data into a structured and standardized framework.

First, the source data is retrieved manually from the corresponding web services in a machine-readable format. Next, the data is converted 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 contain the specific semantic specifications of transforming the source data into triples and form an extension of the central CoyPu COY ontology [15]. We provide both the ontology OWL files, as well as the RML mapping rules used for the graph creation process in our repository for reproducibility.

Simplification We aim to simplify the ACLED and GTA datasets to extract relevant information for our use case, while minimizing noise from irrelevant information. Thus, for GTA, we exclude triples that contain labels, intervention and state acts IDs, and event types, as these triples do not provide any additional information. For ACLED, we exclude triples with comments and labels, and only consider the country location of each event.

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 modeling the impact of political violence or other events on trade flows, as it helps to identify the most affected sectors or products. As this data reduction is helpful in our use case but could be harmful in others, it is a configurable step during dataset generation.

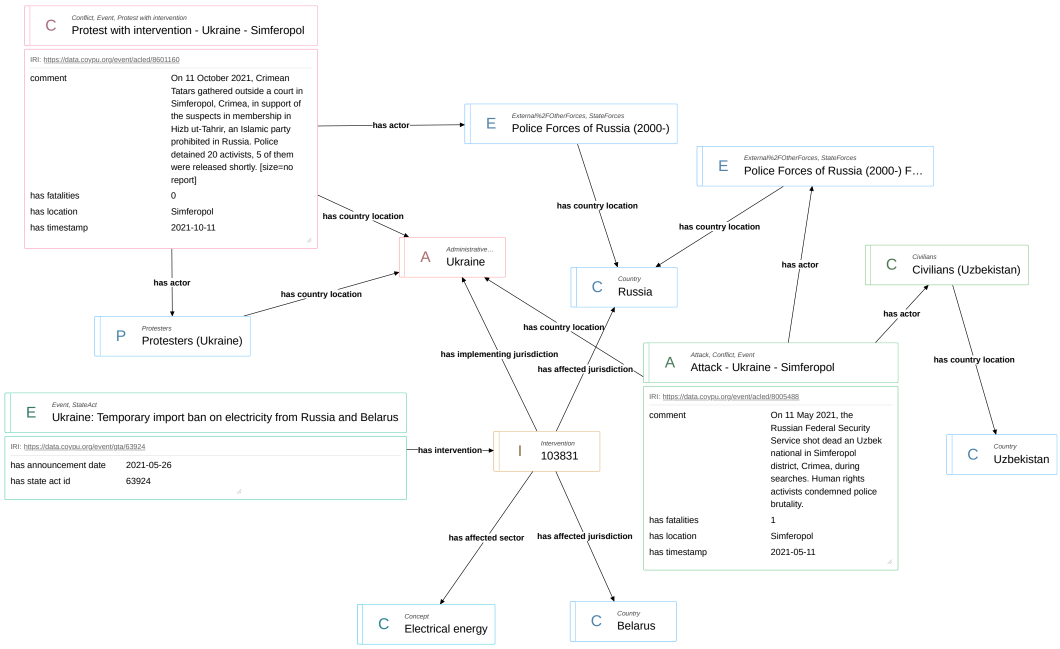

Figure 1: Link between two ACLED events and a GTA event via locations in 2021. The ACLED events (*Attack - Ukraine - Simferopol*, green, and *Protest with intervention - Ukraine - Simferopol*, pink) and the GTA intervention (*intervention 103831*, yellow) link via Russia and Ukraine.

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. We illustrate such a connection in Figure 1, depicting two ACLED events and a GTA intervention in Ukraine and Russia in 2021. It is important to note that this link does not necessitate causality, but rather serves as a foundation for further analysis.

Temporal Information From the given graph in RDF format, we create a tKG with daily granularity, containing quadruples for each day in the year 2021. We have opted to utilize quadruples as our chosen representation, as they are widely employed in the field of tKG forecasting research, see e.g., [9], [10], [8]. 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 (see Section 1.1) aims to predict upcoming global trade events, we focus on the earliest available date for each GTA event, i.e. the announcement date. Therefore, we add the announcement date timestamp to all triples belonging to a state act or intervention. The output of this step is a tKG with quadruples in TXT format for further analysis.

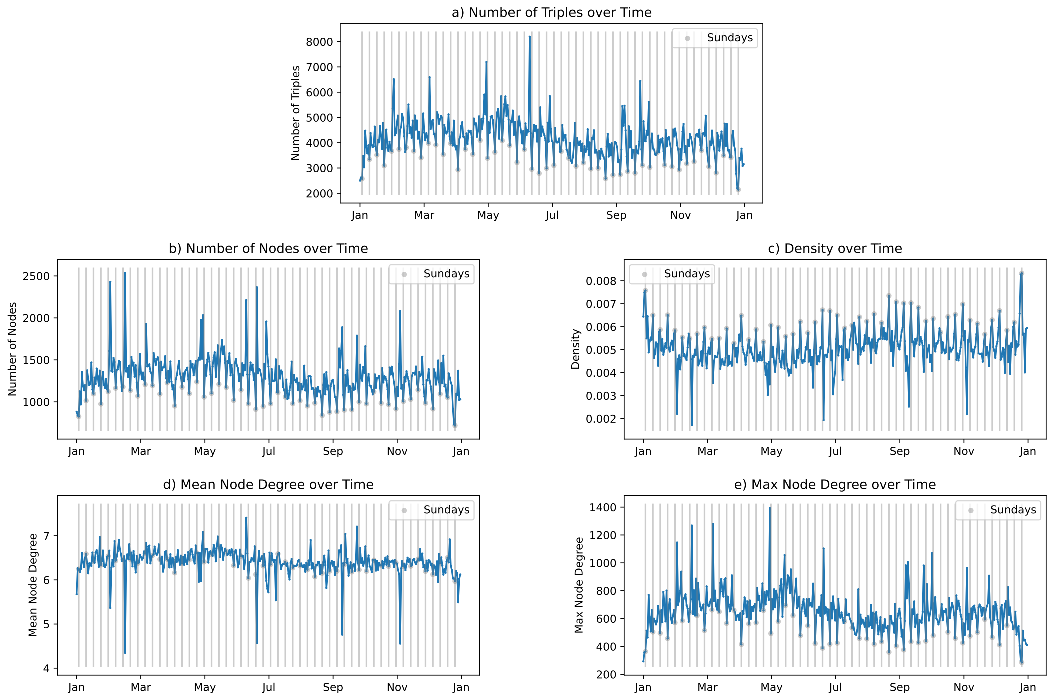

Figure 2: Graph properties over time, one entry per graph snapshot. All Figures show the year 2021. Grey lines and dots mark Sundays. (a) number of triples, (b) number of nodes, (c) density, (d) mean node degree, (e) max node degree.

5. Analysis and Results

Following the steps outlined in Section 4 results in a tKG comprising 1,513,398 quadruples across 365 timesteps, with 290,457 distinct nodes and 13 distinct relations. In this tKG, 1,424,956 quadruples originate from the ACLED dataset, 88,442 quadruples originate from the GTA dataset, containing information on in total 3,677 GTA interventions. In the following, we describe this tKG and provide insights from the conducted analysis.

5.1. Temporal Knowledge Graph Analysis

We analyse the resulting tKG by computing and observing its graph properties over time. Figure 2 illustrates these properties of the KG snapshots per timestep, including the number of triples (a), the number of nodes (b), the density (c), the mean node degree (d), and the maximum node degree (e).

In the following, we describe some key observations:

Number of triples per timestep With more than 2,500 triples in each timestep, this tKG is larger than the tKG datasets described in section 2.

Number of nodes per timestep The tKG contains 290,457 unique nodes, but only a small subset of these nodes (< 1%) is present in each timestep, implying that many nodes do not appear frequently.

Density The density varies between 0.002 and 0.007, indicating a relatively sparse graph[4].

Mean and Maximum Node degree The maximum node degree is significantly higher ($>$ 400%) than the mean node degree, indicating the existence of hub nodes with comparatively very high node degree.

Seasonality The time series in (a) - (d) exhibit weekly seasonality. Sundays contain the lowest number of triples/nodes, the lowest node degree, and the highest density.

Outliers Figure 2 shows five peak days, containing a high number of nodes ($> 2,100$), high number of triples ($> 6,000$), low density (< 0.003), low mean node degree (< 5), and high maximum node degree (> 1000). These days contain hub nodes that have more neighbors than hub nodes in other timesteps.

5.2. Visualisation

We show an exemplary tKG snapshot for the first timestep[5] in Figure 3. Nodes in orange are from GTA triples, blue nodes are from ACLED triples, and green nodes appear in both datasets.

The figure illustrates that the majority of triples are from the ACLED dataset. These blue triples contain a small number of hub nodes, linking to a significant amount of other nodes. These hub nodes consist of nodes representing event types such as *Peaceful Protest*, nodes for prominent actors like *State Forces* or numeric values like a node that denotes the number 1 (connected via the relation *Number of Fatalities*). Further, the figure depicts orange hub nodes, i.e. hub nodes for the GTA dataset. This graph snapshot comprises 13 distinct GTA interventions across 7 GTA state acts. Each intervention has a varying amount of affected jurisdictions (ranging from 1 to 50) and has unique properties, such as affected products and sectors. The orange hub nodes are interventions with a large number of affected jurisdictions.

5.3. Challenges and Considerations for tKG Forecasting: Analysis Insights

The insights gained from the analysis help to define requirements for tKG Forecasting for the use case in Section 1. Compared to the datasets in Section 2, a tKG forecasting model for the given tKG dataset needs to handle a larger number of triples, and a significantly larger number of nodes. Moreover, the model must be capable of distinguishing hub nodes with a very high node degree from nodes with a low node degree, and of differentiating between these. Further, the model must be able to account for peak days and incorporate them into its predictions. An additional challenge is the capturing of seasonal information. For this reason, the forecasting model should have the capability to incorporate seasonal variations in its predictions.

[4]A fully connected graph has a density of 1.

[5]To view the dynamic visualisation of the remaining timesteps, please run the script provided in our GitHub repository and adjust the dedicated slider.

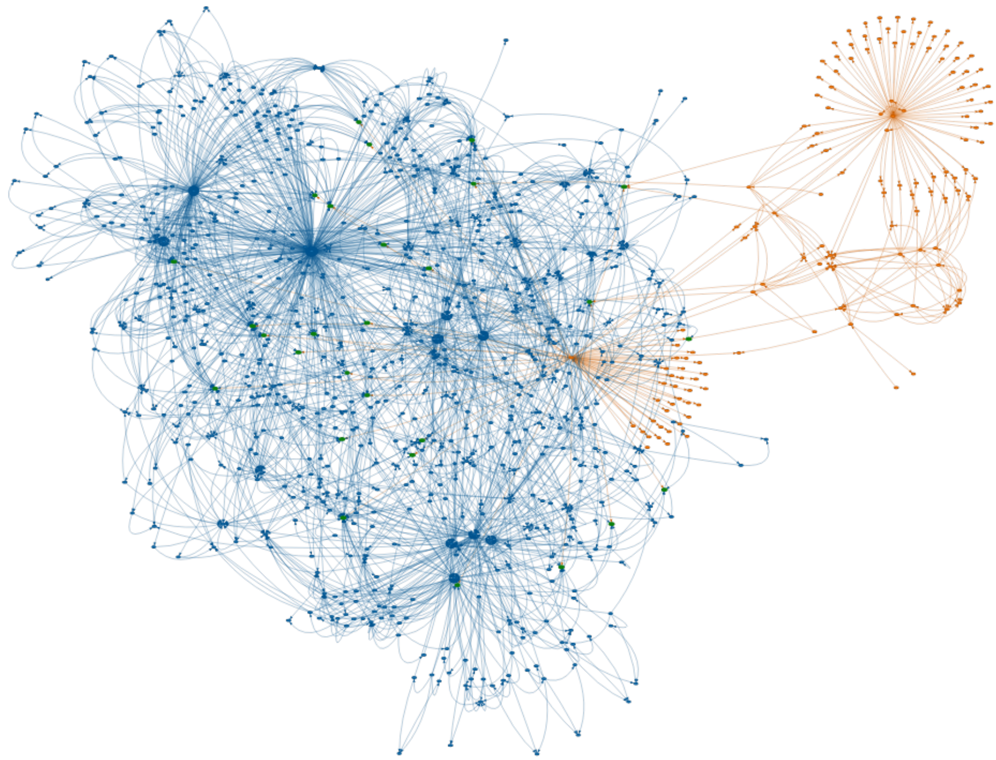

Figure 3: Exemplary KG snapshot for the first timestep. Orange nodes are from GTA, blue nodes are from ACLED, and green nodes appear in both datasets. The majority of triples are from ACLED. Both, ACLED and GTA, contain a small number of hub nodes, linking to a significant amount of other nodes.

6. Conclusion and Future Work

We have presented a novel approach to understanding global crises and trade patterns. For this, our paper outlines the curation process of a tKG from publicly available dynamic data and includes a comprehensive analysis of this tKG. Additionally, we have defined requirements for tKG forecasting models to be used with this dataset.

Leveraging tKG is a promising way to understand the intersection between global crises and trade data over time. In the future, we plan to apply tKG forecasting models within the CoyPu project to predict future trade alert events based on the historical global trade and crisis data, taking into account the key takeaways highlighted in Section 5.3. Our ultimate goal is to enhance our understanding of crises and trade patterns and to contribute to a more effective decision-making process in response to emerging crises.

Acknowledgments

The authors acknowledge funding by the Federal Ministry for Economics and Climate Action of Germany in the project CoyPu (project number 01MK21007[A-L]).

References

[1] W. E. F. 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.

[2] L. Palen, S. Vieweg, J. Sutton, S. Liu, A. Hughes, Crisis informatics: Studying crisis in a networked world (2007).

[3] M. I. Alshaikh, H. P. Larrauri, Building resilience through crisis mapping, community engagement and recovery planning in sudan., in: ISCRAM, 2012.

[4] R. Soden, N. Budhathoki, L. 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.

[5] R. Soden, L. Palen, Informating crisis: Expanding critical perspectives in crisis informatics 2 (2018). URL: https://doi.org/10.1145/3274431. doi:10.1145/3274431.

[6] Coypu, Cognitive economy intelligence plattform für die resilienz wirtschaftlicher Ökosysteme (coypu), 2021. URL: https://coypu.org.

[7] R. Trivedi, H. Dai, Y. Wang, L. Song, Know-evolve: Deep temporal reasoning for dynamic knowledge graphs, in: D. Precup, Y. W. Teh (Eds.), ICML 2017, Sydney, Australia, PMLR, 2017. URL: http://proceedings.mlr.press/v70/trivedi17a.html.

[8] Z. Han, P. Chen, Y. Ma, V. Tresp, Explainable subgraph reasoning for forecasting on temporal knowledge graphs, in: ICLR 2021, Virtual Event, Austria, May 3-7, 2021, 2021. URL: https://openreview.net/forum?id=pGIHq1m7PU.

[9] W. Jin, M. Qu, X. Jin, X. Ren, Recurrent event network: Autoregressive structure inference over temporal knowledge graphs, arXiv preprint arXiv:1904.05530 (2019). Preprint version.

[10] Z. Li, X. Jin, W. Li, S. Guan, J. Guo, H. Shen, Y. Wang, X. Cheng, Temporal knowledge graph reasoning based on evolutional representation learning, in: SIGIR '21, Virtual Event, Canada, ACM, 2021, pp. 408–417. URL: https://doi.org/10.1145/3404835.3462963.

[11] S. Kim, Y.-W. Kwon, Construction of disaster knowledge graphs to enhance disaster resilience, in: 2022 IEEE International Conference on Big Data (Big Data), 2022, pp. 6721–6723. doi:10.1109/BigData55660.2022.10021017.

[12] S. Tilly, G. Livan, Macroeconomic forecasting with statistically validated knowledge graphs, Expert Systems with Applications 186 (2021) 115765. URL: https://www.sciencedirect.com/science/article/pii/S0957417421011386.

[13] Y. Yang, Y. Pang, G. Huang, W. E, The knowledge graph for macroeconomic analysis with alternative big data, 2020. URL: https://arxiv.org/abs/2010.05172. doi:10.48550/ARXIV.2010.05172.

[14] Definition of the CIDOC Conceptual Reference Model Version 5.0.4, Technical Report, ICOM/CIDOC CRM Special Interest Group, 2011.

[15] Coypu coy ontology, 2023. URL: https://gitlab.com/coypu-project/coy-ontology.

[16] F. Zhang, Z. Li, D. Peng, J. Cheng, RDF for temporal data management - a survey, Earth Sci. Informatics 14 (2021) 563–599. URL: https://doi.org/10.1007/s12145-021-00574-w.

[17] E. Boschee, J. Lautenschlager, S. O'Brien, S. Shellman, J. Starz, M. Ward, ICEWS Coded Event Data, 2015. URL: https://doi.org/10.7910/DVN/28075. doi:10.7910/DVN/28075.

[18] A. García-Durán, S. Dumančić, M. Niepert, Learning sequence encoders for temporal

knowledge graph completion, in: EMNLP 2018, Brussels, Belgium, 2018, pp. 4816–4821. URL: https://aclanthology.org/D18-1516.

[19] K. Leetaru, P. A. Schrodt, Gdelt: Global data on events, location, and tone, 1979–2012, in: ISA annual convention, Citeseer, 2013, pp. 1–49.

[20] F. Mahdisoltani, J. A. Biega, F. M. Suchanek, Yago3: A knowledge base from multilingual wikipedias, in: CIDR, 2015.

[21] J. Leblay, M. W. Chekol, Deriving validity time in knowledge graph, in: Companion proceedings of the the web conference 2018, 2018, pp. 1771–1776. URL: https://doi.org/10.1145/3184558.3191639.

[22] S. J. EVENETT, Global trade alert: Motivation and launch, World Trade Review 8 (2009) 607–609. doi:10.1017/S1474745609990061.

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