Towards an Extensible, Heterogeneous Graph-based Modeling of Conflict

David Spitzenberg
Center for Crisis Early Warning
University of the Bundeswehr Munich
david.spitzenberg@unibw.de

ABSTRACT

Forecasting armed conflict remains challenging, and many approaches emphasize spatio-temporal patterns while under-representing human actors and the relations through which violence propagates. We reinterpret conflict event datasets as heterogeneous, time-indexed graphs. These graphs jointly encode actors, civilians, and countries alongside their mutual interactions. Using the UCDP Georeferenced Event Dataset, we show that centrality- and embedding-based analyses of the resulting graphs' topology capture meaningful patterns, including (i) actor groupings by aggression neighborhood, and (ii) a separation between countries that subsequently experience violence and those that do not. We conclude by outlining extensions to our representation incorporating additional data sources and findings from process-oriented research, and by discussing its future use as a foundation for graph-based predictive models, with the potential to improve the performance of conflict forecasts.

Keywords

Political Violence, Conflict Prediction, Knowledge Representation, Heterogeneous Graphs

INTRODUCTION

Armed conflict emerges from the complex interplay of multiple, distinct elements. Process-oriented conflict research has identified a wide range of mechanisms shaping the onset and dynamics of political violence, including socio-economic factors (Collier & Hoeffler, 2004; Fearon & Laitin, 2003), environmental conditions (Burke et al., 2009), and transnational linkages (Gleditsch, 2007). In parallel, conflict prediction seeks to forecast future (de-)escalation of conflict. Here, accurate predictions enable practitioners to timely identify future violence and allocate scarce mitigation resources most effectively. Yet, conflict prediction does not pay equal attention to all relevant elements, including human actors, location, and time. In fact, recent work heavily centers around spatio-temporal patterns, mostly under-representing human actors who instigate and sustain violence. This under-representation contrasts process-oriented research, where the link between political violence and actor-related variables including e.g. opportunity for rebellion (Collier & Hoeffler, 2004), group-level inequality (Buhaug et al., 2014) as well as group-level injustices (Sakstrup & Bartusevičius, 2024) has been studied extensively. Moreover, current research typically focus on a single type of element at a time (e.g. only countries) and seldom put multiple elements into mutual relation, leaving out questions such as "who operates in a given country in a given month?". As a consequence, the complex interplay of, for example, actors reacting to an opponents' actions is largely lost during forecasting. In parts, this limitation is grounded in a lack of expressive knowledge representations tailored to forecasting tasks. We propose graph-based knowledge modeling to overcome these limitations in conflict prediction: encoding conflict data as heterogeneous knowledge graph(s) allows scholars to (i) represent diverse elements explicitly, and (ii) model their mutual relations. By encoding individual actors directly into the graphs, this approach allows scholars to address the driving forces behind conflict events with the necessary attention. Here, actor-informed analyses might contribute to policy making in two distinct ways: (i) improved predictions resulting from the incorporation of above actor-related effects, and (ii) identification of candidate allies for policy implementation, improving implementation efficiency (Metternich et al., 2019).

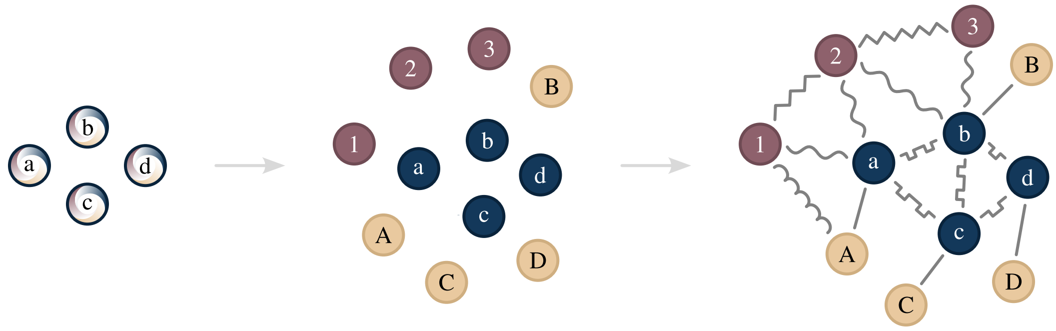

Figure 1. Left: conventional analysis subsume characteristics of different real-world elements (e.g. actors, civil populations and countries) under a single, fixed type of element of analysis (e.g. countries; depicted as nodes). Furthermore, individual elements are processed mostly independent from each other, leaving out mutual relations. Center: to disentangle this convolved representation of information, real-world elements have to be encoded individually (again depicted as nodes) — potentially using multiple, distinct types (e.g. actors vs. civil populations vs. countries; depicted in different colors). Right: to allow for a transfer of conclusions drawn on individual elements between one another, mutual relations among real-world elements (e.g. "an actor opposes another actor") have to be encoded as well (depicted as edges). Again, relations can be of multiple types (e.g. "an actor opposes another actor" vs. "a country neighbors another country"; depicted using different markers). For a more detailed break-down, also see Figure 2.

Contemporary conflict prediction casts all the information available into a single, fixed type of element — be it event, country or actor. In doing so, it does not account for the heterogeneity inherent to conflict data. This "one-size-fits-all" approach falls short of modeling multifaceted features encountered in the real world: an actor does not convey a per capita GDP nor does a spatial territory deal fatalities. With its focus on homogeneous elements (that is, all elements share a common type), questionable feature assignments become unavoidable in current research (cf. Figure 1 (left)). Additionally, individual elements are rarely put into relation with one another. Studying elements individually again does not represent the real world particularly well: actors form alliances and mutually opposing coalitions, where actions taken by one actor impact future actions taken by allies and opponents. Similarly, spill-over effects propagate conflict spatially among neighboring countries.[1] Focusing on individual elements keeps current predictive models from incorporating conclusions drawn on one element into predictions for another element.

Graph-based representations allow us to overcome most of the aforementioned shortcomings. In heterogeneous graphs, elements like actors or country territories are represented individually as nodes (c.f. the nodes depicted in Figure 1 (center) as well as Figure 2). Here, nodes do not have to share a common type. Instead, different element types are made apparent by encoding them as distinct node sets. In this regard, contrast red actor, yellow civilian population and blue country nodes in the aforementioned figures. However, representing elements heterogeneously as individual nodes is of limited use. Representing the elements' mutual interplay is of equal importance. Graphs readily allow for putting individual elements into mutual relation. To this end, relations are encoded as edges (c.f. the edges depicted in Figure 1 (right) and Figure 2). Here, edges do not have to share a common type either. Once more, different edge types — ranging semantically from "an actor deals fatalities to another actor" to "a civil population resides in a country" — are encoded as distinct edge sets (c.f. the different edge markers in Figure 1 (right) and Figure 2).

Our contributions are three-fold. We first show the feasibility of re-interpreting existing conflict event datasets from the viewpoint of heterogeneous graphs. To this end, we present an encoding of the UCDP Georeferenced Event Dataset (UCDP GED) as collection of knowledge graphs. In doing so, we incorporate (i) different types of elements of analysis (namely actors, civil populations and countries) as well as (ii) heterogeneous mutual relations among elements (namely actors opposing each other, actors victimizing civil populations, actors operating in countries, civilians residing in countries and countries neighboring each other) all within a unified, holistic representation. Second, we discuss some of the immediate advantages of a graph-based representation of conflict. Here, we show the increased interpretability of conflict data by deriving insightful visuals from our graphs. Additionally, by computing node level indices and embeddings, we show that our graphs' topology carry both semantic and predictive meaning beyond the mere encoded elements' types. Third, we conclude with an discussion of potential extensions of our work. Being work in progress, we do not turn to forecasting future conflict dynamics based on our

[1] We acknowledge that a small number of models does capture spatial diffusion of conflict occurrence (Brandt et al., 2022; Li et al., 2021; Maase, 2025).

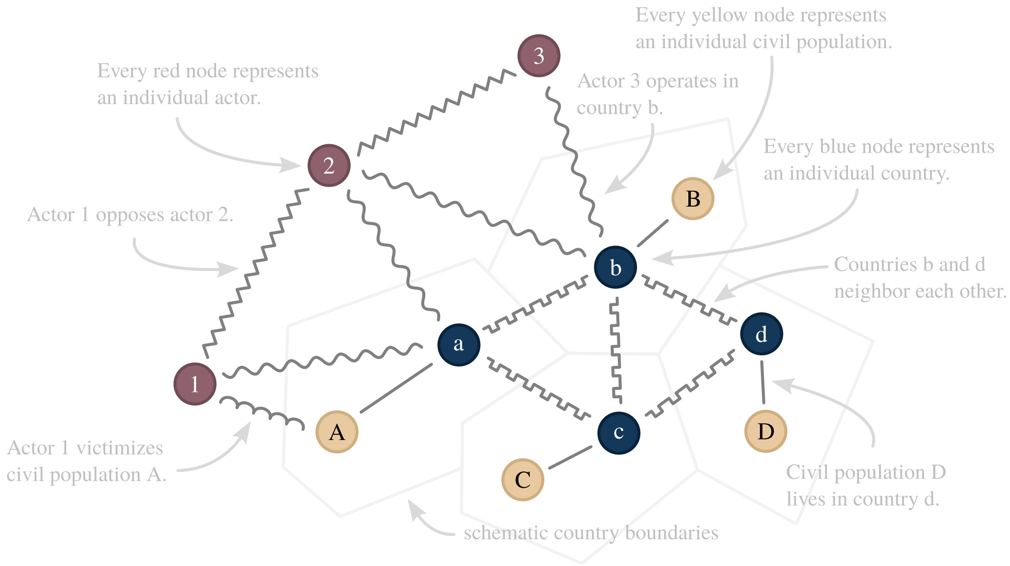

Figure 2. Artificial knowledge graph consisting of actors, civilian populations and countries. Individual elements are encoded as individual nodes within the graph. The graph also encodes relations: "actors opposing each other" (), "actors operating in some country" (), "actors victimizing civilians" (), "civilians residing in some country" (—) as well as and "countries neighboring each other" (). Again, each relation type is encoded as distinct edge set and individual relations are encoded as individual edges.

constructed graphs just yet. Besides building predictions on top of the constructed graphs, we also see room for (i) incorporation of both additional datasets and (ii) constitutional effects from process-oriented research within a single framework, as well as (iii) supporting knowledge representation in both human- and AI-driven analyses. In summary, we take the first steps towards introducing a heterogeneous graph-based framework to the field of conflict prediction. In doing so, we hope to advance both future models' predictive expressiveness and performance.

RELATED WORK

Political violence has been studied from various viewpoints in the past. Before turning to our own contributions, we briefly review streams of research we consider most relevant to the present body of work.

Conflict prediction

As coined by the ViEWS prediction challenge (Hegre et al., 2023; Vesco et al., 2022), conflict prediction deals with the forecasting of dynamics in political violence. Within this area, only few studies are actor-informed or graph-based.

Actor-informed conflict prediction

Metternich et al. (2019) compare space-centric with actor-centric approaches in conflict prediction. The authors argue in favor of actor-centric approaches as these (i) may capture large-scale processes beyond small-scale spatial diffusion, (ii) reduce zero-inflation in input data, and (iii) have a stronger link to applied policy making, because actor-centric approaches readily allow for telling apart candidate allies from opponents for e.g. policy enforcement. Metternich et al. derive actor-centric features from the UCDP GED and apply random forests to obtain predictions on future conflict dynamics.

Croicu and Maase (2025) predict future conflict dynamics by means of large language models (LLMs) from automatically generated, monthly digests covering individual actor dyads. Here, the monthly digest is built from the corpus of reportings referenced in the UCDP GED enriched by surrounding, non-event-related coverage (so called contexts). The authors make the case for (i) putting more research attention to actors (being the driving forces behind conflict), and (ii) leveraging more conflict- but not conflict event-related, unstructured inputs (i.e. raw texts as opposed to tabular data) to increase the predictive power of models.

Both bodies of work study only a single type of element (actor dyads) and do not put elements into relation with one another (neither with elements of the same type nor with elements of different type).

Graph-based conflict prediction

Brandt et al. (2022) and Li et al. (2021) apply graph neural networks (GNNs) to conflict prediction. More precisely, the authors employ graph convolutional layers (GCLs) alongside ordinary one-dimensional convolutional layers to process spatial and temporal dependencies respectively. Using these two types of layers, the authors build a spatio-temporal two-step approach to predict future fatality counts per PRIO-grid cell.[2]

However, the authors do not take advantage of graphs' expressive power. Due to the study of PRIO-grid cells, the spatial graph is highly regular. I.e. each cell has exactly eight neighboring cells akin to how pixels are arranged in images. Furthermore, only a single type of elements is studied (PRIO-grid cells). In consequence, the authors apply GNNs' to regular, homogeneous graphs — equivalent to the application of ordinary "plain" two-dimensional convolutional layers.

Network effects in political violence

In a more process-oriented stream of research, Dorff et al. apply graph approaches to the study of constitutional effects of political violence. The authors' studies include higher-order dependencies among actors in Nigeria and the effect of network competition on the victimization of civilians (Dorff et al., 2020, 2023). In both bodies of work, the authors make the case for violence emerging less from the mere presence of violent actors but rather from the mutual relations among actors. Despite theorizing about the importance of e.g. territorial control and civilian support (i.e. the interplay between actors, civilians and territories), the studies are based on homogeneous graphs consisting of actors only.

REINTERPRETING CONFLICT EVENT DATASETS AS HETEROGENEOUS GRAPHS

We now detail our approach of re-interpreting conflict event datasets as heterogeneous knowledge graphs. Following our reasoning in the introduction, we want our knowledge graph to (i) incorporate the "human dimension" of conflict, i.e. be actor-informed, and (ii) put individual elements into mutual relation. To give us some exploratory room, we thus encode the three distinct node sets (i) actors, (ii) countries, and (iii) civil populations alongside the five distinct edge sets (i) "an actor deals fatalities to another actor", (ii) "an actor deals fatalities to a civil population", (iii) "an actor operates in a country", (iv) "a country neighbors another country", and (v) "a civilian population resides in a country". Months establish the the de-facto standard temporal resolution for analysis in the literature. For this reason, we derive a single knowledge graph per month.

Used datasets

We center our work around one of the two most common conflict event datasets: the Georeferenced Event Dataset released by the Uppsala Conflict Data Program (UCDP GED; Davies et al., 2025; Sundberg and Melander, 2013). UCDP GED is a tabular dataset reporting on conflict events. In its latest version,[3] the dataset covers the entire world from January 1, 1989 through December 31, 2024. The events included in the dataset are extracted from different types of reportings including news, NGO and IGO reports as well as social media. Here, an event is defined as lethal use of armed force by an organized actor. Events have to be attributable to both location and time. For the precise definition of an event, see Högbladh (2025). The dataset includes a multitude of per-event features of which the (i) involved actors, (ii) event location, and (iii) event timing are relevant for this present work. Note that there is another widespread conflict event dataset: the Armed Conflict Location & Event Data dataset (ACLED; Raleigh et al., 2023). ACLED does also report the required entities so a transfer of our method to the latter is possible.

Some processing steps require access to spatial representations of countries' boundaries. To this end, we employ the Comprehensive Global Administrative Zones dataset (CGAZ; Runfola et al., 2020). CGAZ captures administrative zone boundaries globally up to the second level (municipalities; though we only employ the zero-th level in our analysis) at an appropriate level of detail. We use CGAZ in favor of e.g. the list of independent states released by Gleditsch and Ward (GW; Gleditsch and Ward, 1999), because in contrast to the latter CGAZ includes the countries' boundaries as polygons.

Node set construction

From the datasets outlined above, we derive the three node sets (i) actors, (ii) countries, and (iii) civil populations.

[2] PRIO-grid cells establish a commonly used, regular tesselation of the earth's surface. For details, see the original publication (Tollefsen et al., 2012).
[3] Version 25.1 as of writing.

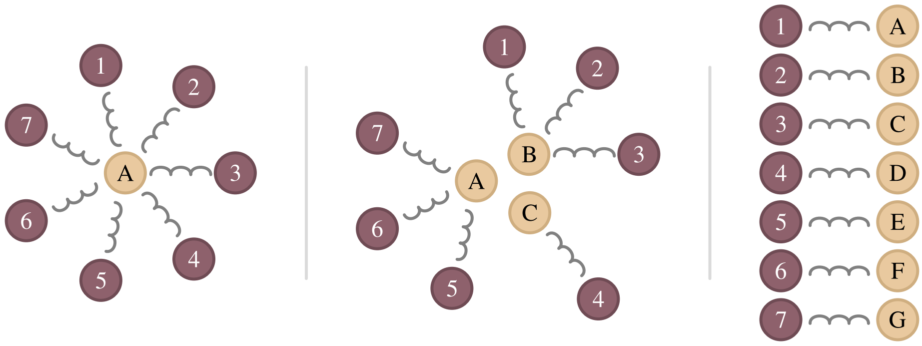

Figure 3. Ways of disaggregating the civil substitute actor and putting it into relation with "true" actors. Left: direct representation of UCDP's single substitute actor as single, global node. This representation puts all civilian-targeting actors in close neighborhood. Center: disaggregation of UCDP's single substitute actor based on e.g. countries (not shown). Actors get grouped based on the civil population they target. Right: disaggregation of UCDP's single substitute actor based on "true" actors. Actors do not get connected indirectly via any civilian-representing node.

Actors

For each month, we encode a set of actors as individual nodes (c.f. red nodes in Figure 4). To reduce zero-inflation, we do not encode the entire set of actors recorded in UCDP GED. Instead for a given month, we only encode actors instigating or being targeted by violence in the current and upcoming month. The tight coupling between actors encoded in a graph and actors involved in violence occurrence results in a bias towards violence occurrence compared to the overall average across time. In a predictive setting, where data leakage is to be avoided, this bias can be reduced by e.g. extending the set of encoded actors. We leave this extension to future work once we actually turn to predicting future violence from the constructed knowledge graphs.

As reported by UCDP GED, approximately 5 % of the reported events reference groups of actors as conflicting sides. As we want to encode actors individually in our graphs (i.e. one node per actor and one actor per node), we exclude such events from our analysis. One way to resolve actor groups is to incorporate them as cliques (in the graph sense) of allying actors. As we currently do not target alliances explicitly, we again leave this study of actor groups for future work. Similarly, we exclude events of ambiguous month of occurrence (approximately 2 % of the events).

Countries

To incorporate conflict locations, we encode a single node per country as reported by CGAZ in our constructed graphs (c.f. blue nodes in Figure 4). Here, the set of encoded countries is constant along time. As UCDP GED reports GW state numbers, we have to map GW state numbers to CGAZ country codes. We perform the mapping by matching state/country names as well as 3-letter codes in a multi-step procedure, manually taking care of ambiguities.

Civilians

Besides actor-actor violence, UCDP GED also includes one-sided events. In one-sided events, an organized party deliberately targets the civil population. Here, UCDP GED reports the instigator of violence as *side a* and a "substitute actor" as *side b*. The substitute actor representing the civil population is shared among all one-sided events. This shared nature implies a certain complexity in the choice of representation of civilians in our graphs. Several choices exist, including (i) encoding a single civil actor, (ii) encoding per-instigator disaggregated civil actors, and (iii) encoding per-country disaggregated civil actors.

Encoding civilians as single civil node directly maps UCDP GED's representation into the domain of knowledge graphs. However, we consider this representation unsuitable for our modeling. Encoding civilians as single node subsequently introduces very short paths connecting otherwise unrelated actor nodes. I.e. two unrelated actors targeting the civil population in some events — no matter how distant spatially, temporally or ideologically — become indirect neighbors within our constructed graphs (c.f. Figure 3 (left)).

The above issue of introducing unintended paths in our constructed graphs is resolved by disaggregating the single civilian substitute actor into multiple nodes. An easy to implement approach is to introduce one civil counterpart per "true" actor node. However, we consider this method of disaggregation somewhat artificial. While the above single-node encoding of civilians potentially puts actors in too close neighborhood, this per-instigator encoding of civilians removes any connection between actors targeting similar groups of civilians (c.f. Figure 3 (right)). E.g. this method of disaggregation falls short of representing civilians becoming a plaything of conflicting actors.

As a middle ground, we represent civilians based on a per-country disaggregation. I.e. we encode a single civilian node per country included in the CGAZ dataset. Here, we consider countries as both suitable and feasible proxy for disaggregation. In our constructed graphs, actors targeting the civil population in the same country are kept in close neighborhood while actors targeting the civil population in different countries are left unrelated (c.f. center of Figure 3 as well as Figure 4). Nevertheless, this representation implies the debatable assumption of civil populations being somewhat homogeneous within countries. Other ways of disaggregation based on e.g. ethnicity would also be possible.[4]

Edge sets

As argued in the introduction, the disaggregated representation of individual, heterogeneous elements on its own is of limited use. The representation of mutual relations is of equal importance. We encode five distinct types of relations by means of individual edge sets: (i) "an actor deals fatalities to another actor", (ii) "an actor deals fatalities to a civil population", (iii) "an actor operates in a country", (iv) "a country neighbors another country", and (v) "a civilian population resides in a country". Here, we consider edges to be directed.

Actor activity

Edges of the types "an actor deals fatalities to another actor" (*AFA*), "an actor deals fatalities to a civil population" (*AFC*) and "an actor operates in a country" (*AOC*) all result from an analysis of the events reported in UCDP GED. For each month, we encode an edge $a \rightarrow x$ whenever there is at least one event for said month indicating activity by actor a against or in element x.

Country neighborhood

To derive edges in the "a country neighbors another country" edge set (*CNC*), we analyze country pairs. To this end, we compute the pair-wise distance between country boundaries as included in CGAZ. Country pairs (c_1, c_2) featuring a minimal distance below an empirically chosen threshold get encoded as two edges $c_1 \rightarrow c_2$ and $c_1 \leftarrow c_2$. Here, encoding country pairs as pairs of inverse edges captures the neighborhood relation's symmetry. We consider the resulting edge set constant along time, i.e. for each month, the constructed graph contains the same edges in the *CNC* edge set.

Civilian residency

As detailed above, we represent a country's civil population using a single node per country within the constructed graphs. We capture this correspondence between country and civilian nodes by including a single edge per country c and its civil population p_c connecting the two: $p_c \rightarrow c$. Again, we consider the resulting edge set constant along time. In the following, we refer to this edge set as *CRC*.

DERIVING INSIGHTS FROM GRAPHS OF CONFLICT

We apply a set of first analyses to showcase different notions of information captured within our constructed knowledge graphs. First, the constructed knowledge graphs allow for insightful visuals. Figure 4 depicts the graph constructed for October 2024. The graph contains a total of 218 countries alongside the respective civil populations (accounting for the same number of nodes) as well as 157 actors involved in violence in either October 2024 or November 2024. In total, there are 1357 directed edges of varying type (c.f. Table 1 for details). IS is the actor involved with the most opponents (21 actors). IS also is the actor operating in the most countries (eleven countries). For the depicted month, civilians living in the Democratic Republic of the Congo suffer from the highest number of violent actors victimizing the civil population (six actors).

On top of these descriptive readings, Figure 4 also surfaces more intricate patterns of conflict. One of these patterns is the difference in area of operations for individual actors. While actors in e.g. the Americas tend to operate on an intra-national level (see e.g. Brazil, Colombia, Ecuador, and Mexico), actors from Africa operate more internationally. E.g. IS' area of operation extends even beyond Africa from Mali and Burkina Faso in the West to

[4]As UCDP GED does not report on e.g. ethnicity, disaggregations based on such groupings require the incorporation of additional datasets. To keep this body of work focused, we thus leave exploring such disaggregations to future work. Also note that UCDP GED does not report on the exact reason why an actor victimizes civilians. Consequently, determining e.g. the specific ethnicity targeted in a given event might be infeasible or even impossible.

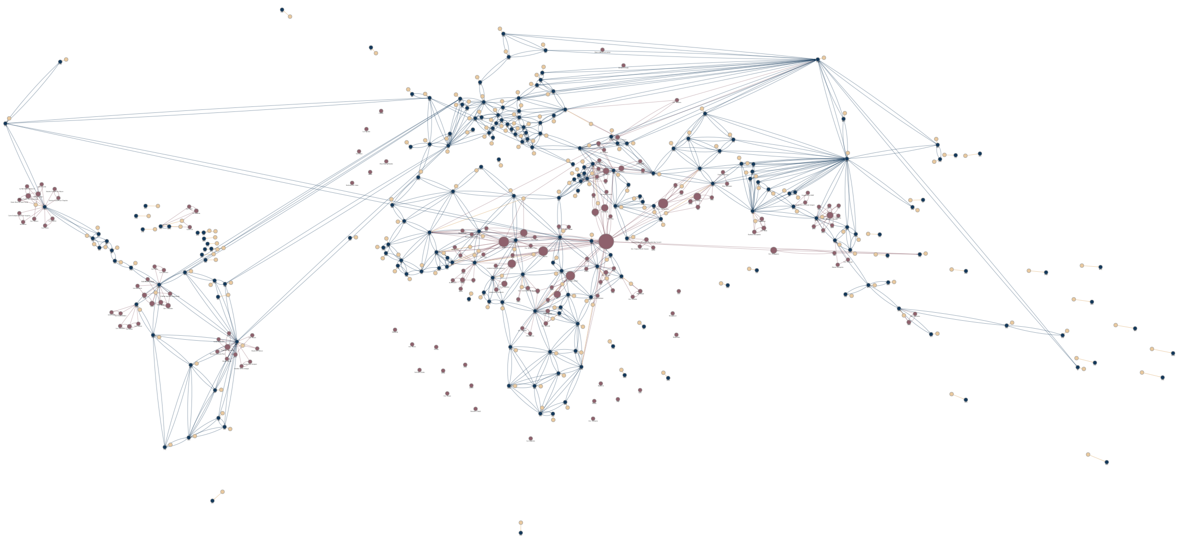

Figure 4. Graph constructed from UCDP GED for October 2024. The graph consists of actors, civil populations and countries. Additionally, the graph contains edges representing *AFA*, *AFC*, *AOC*, *CNC* and *CRC* semantics. Refer to Table 1 for node and edge counts. Actor nodes are scaled according to their betweenness centrality index computed on the $AF*$ -induced subgraph. Positioning of countries follows their centroids' positions as derived from CGAZ. Some countries have been relocated for visual purposes. Figure done using *graph-tool* (Peixoto, 2014).

Table 1. Node and edge counts for the graph constructed for October 2024 from UCDP GED (see Figure 4).

element	*number of nodes*
actor	157
civil population	218
country	218
total	*593*

relation	*number of edges*
country neighbors country	738
actor causes fatalities to civilians	57
actor causes fatalities to actor	179
civilians live in country	218
actor operates in country	165
total	*1357*

Mozambique in the South to Syria and Afghanistan in the North to the Philippines in the East. Similarly, JNIM and JAS operate in five (Benin, Burkina Faso, Mali, Niger, and Togo) and three (Cameroon, Chad, and Nigeria) countries respectively. Grasping the above descriptive readings and the resulting patterns visually is more intuitive compared to reading the raw underlying event dataset tables. The generation of such visuals in turn requires the encoding of the dataset as heterogeneous graph.

Second, a graph-based representation of conflict allows for application of methods from social network analysis (SNA). In SNA, betweenness centrality is a common index measuring the frequency of a node v being part of a shortest path from node s to node t. Applied to the subgraph induced by $AF*$ edges, actors featuring a high betweenness centrality represent actors connecting otherwise unrelated conflicts. As can be seen from Figure 4, e.g. IS features a high betweenness centrality because of connecting conflict in Africa with conflict in the Middle East. In practice, actors from Africa and the Middle East may form alliances against IS — alliances that are less probable in the absence of the shared opponent.

Third, the graph-encoding of conflict allows us to derive graph-based country or actor embeddings. Such embeddings carry a meaning even for non-graph-based downstream analyses. To support our intuition detailed in the introduction, we apply the metapath2vec algorithm (MP2V; Dong et al., 2017) to the constructed graphs. Within MP2V, embeddings for a heterogeneous graph's nodes are learned. Here, the learned embeddings maximize the probability of encountering neighborhoods as seen in a given graph. Neighborhoods are constructed by means of random walks emerging from a given node adhering to a given metapath. By construction, the obtained embeddings are fully feature-agnostic and only draw from the underlying graph's topology. In other words, embeddings do not capture e.g. past fatality counts, socio-economic indicators or alike, but only information on adjacency. We employ MP2V by applying it to a metapath and subsequently reducing the resulting embeddings' dimensionality by means of a principal component analysis (PCA).

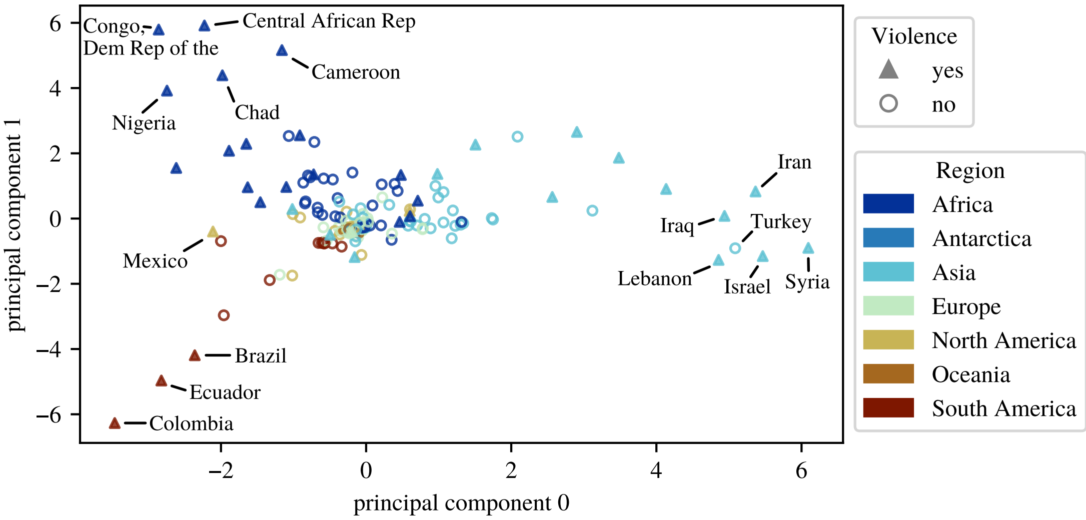

Figure 5. Projected MP2V-based country embeddings. Each marker represents a country from the graph constructed for October 2024 (c.f. Figure 4). Countries not facing violence in November 2024 form a cluster towards the origin while countries facing conflict in the same month spread out.

Figure 5 depicts country embeddings obtained for October 2024. Here, we apply MP2V to the metapath $\langle AOC, CNC, \mathrm{AOC}_{\mathrm{reversed}}, AFA \rangle$. Note how (i) countries group according to their respective region of world, and (ii) countries not facing violence in the following month form a cluster towards the origin — while countries facing violence are more spread out. Recall that while not perfect, this separation of peace and violence draws from the underlying graph's topology only. In a sense, this separation is a direct consequence of what is known as "conflict trap": country nodes facing violence now (and likely seeing conflict in the future) are adjacent to actor nodes whereas country nodes not facing violence do not neighbor actor nodes.

Figure 6 depicts actor embeddings obtained for October 2024. Here, we apply MP2V to the metapath $\langle CNC, \mathrm{AOC}_{\mathrm{reversed}}, AFA, AFA, AFA \rangle$. Note how actors cluster according to their respective aggression-neighborhood. This neighborhood coincides with countries for actors operating on an intra-national level (cf. e.g. Brazil, Mexico or Myanmar). But also for internationally operating actors, this clustering applies (cf. e.g. the grouping of governments of Iran and Israel as well as Fatah, Hamas, and Hezbollah). Furthermore, note how actors featuring a high number of attacked opponents tend to be embedded off-center (c.f. e.g. Comando Vermelho, the Government of Myanmar, IS, or JNIM). More precisely, the actors' respective projected embedding's L_2-norm correlates with the number of opponents.

In conclusion, the constructed knowledge graphs' topology carries both semantic and predictive meaning beyond the type of encoded nodes and edges. For countries, this includes e.g. (i) geo-spatial arrangement on a region of world level, and (ii) future violence or peace. For actors, this includes (i) the role being played in the surrounding network of actors, (ii) (geo-spatial) clustering within the network, and (iii) aggressiveness in terms of attacked opponents. To leverage the underlying topology, a graph encoding of conflict data is a strict requirement. The derived meaning can be made apparent also to non-graph-based analyses by means of e.g. SNA indices or graph-based node embeddings. Nevertheless, to fully process the information captured by the heterogeneous graph-encoding of violence, graph-based analyses have to be studied in closer detail in future work.

OUTLOOK

Political violence results from the complex interplay of human actors, space and time. Current research on conflict prediction is often centered around tabular encodings of spatio-temporal patterns of violence. Here, human actors are mostly overlooked — despite their central role in conflict. Similarly, relations among individual elements as e.g. "who operates in a given country in a given month?" often remain disregarded. The key role of actors and interactions has been studied in the process-oriented literature, but is hard to incorporate in forecasts when using tabular encodings. Consequently, key aspects of political violence are not incorporated in current analysis.

To make such aspects apparent, we detailed an approach to re-interpreting proven conflict event datasets (exemplified using UCDP GED) as collection of per-month heterogeneous graphs. In doing so, we incorporated dedicated node sets for (i) actors, (ii) countries, and (iii) civil populations. We linked individual nodes by heterogeneous edges representing the semantics of (i) actors opposing actors, (ii) actors victimizing civil populations, (iii) actors

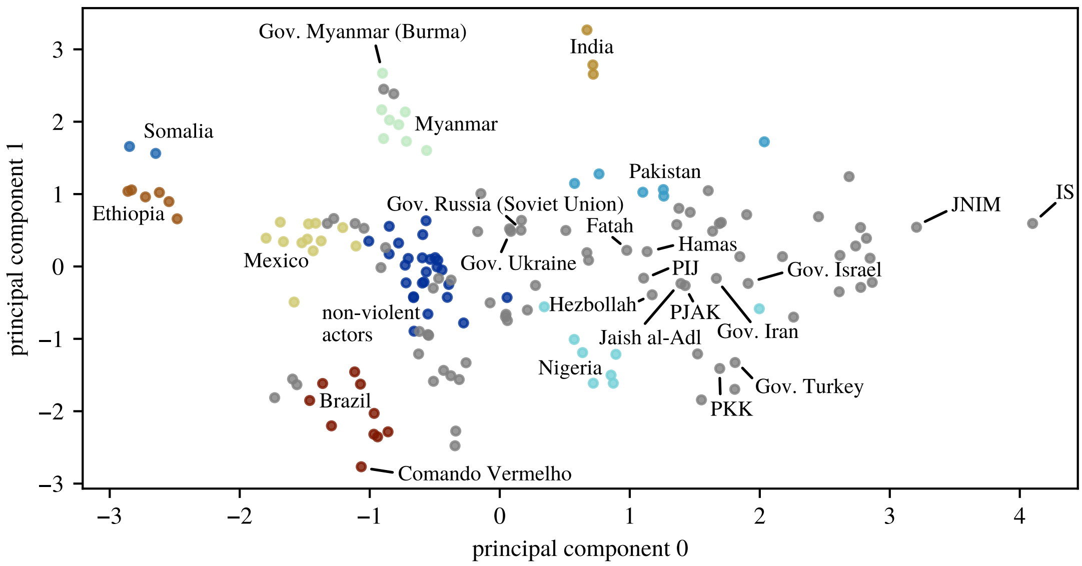

Figure 6. Projected MP2V-based actor embeddings. Each marker represents an actor from the graph constructed for October 2024 (c.f. Figure 4). Actors cluster according to their respective aggression-neighborhood. Furthermore, actors facing a high number of opponents tend to be located off-center.

operating in countries, (iv) civilians residing in countries, as well as (v) countries neighboring countries. Building on this graph-based encoding of conflict, we provided evidence for the encoding's immediate advantages. To this end, we derived insightful visuals surfacing patterns of conflict. Additionally, we computed purely topology-based node level indices and embeddings showing that topology (i) captures semantic meaning: indices and embeddings pinpoint human actors depending on their respective role taken within the network of mutual aggression; (ii) captures predictive meaning: embeddings separate future peaceful from violent countries; and (iii) can be translated to non-graph based analyses by means of said embeddings.

The present study lays the groundwork for the introduction of heterogeneous graphs for knowledge modeling to the area of conflict prediction. Building on this groundwork, future research may pursue several directions. First, heterogeneous graphs lend towards a holistic encoding of various types of data. Historical conflict data can readily be attached to actor nodes. Country nodes provide an adequate point of incorporation for structural data like socio-economic indicators. Text data can be included in the constructed graphs by e.g. introducing an additional distinct node set representing textual reports. In doing so, graphs' expressiveness allows for blending previously disparate types of input data in an holistic way while accounting for the inherent heterogeneity. Second, findings from process-oriented research can be incorporated. The notion of e.g. territorial control can be accounted for by means of a node set of higher spatial resolution (e.g. ADM1 territories) and an edge set linking actors to their controlled territories. Here, a careful examination of which constitutional effects to incorporate is mandatory, as statistical significance does not always imply predictive benefit (Ward et al., 2010). Third, we see room for evolving the constructed graphs towards a conflict-encoding graph database to support knowledge discovery for both humans and artificial intelligence in e.g. retrieval-augmented generation setups. Last but not least, the framework of heterogeneous graphs is not limited to the modeling of knowledge. So far, the framework is motivated by observation of conceptual limitations of present forecasting approaches as well as findings from process-oriented research. To further support our work empirically, we will evaluate our approach in a predictive setting in future work. Here, our eventual goal is to build a predictive model operating directly on our constructed graphs. Drawing inspiration from e.g. the ViEWS prediction challenge (Hegre et al., 2023; Vesco et al., 2022), the idea is to predict future fatality counts from a graph encoding. Leveraging graph neural networks, this opens the area of conflict prediction for application of recent advances in deep learning. Eventually, we hope to advance models' predictive performance to enable practitioners to mitigate the harms associated with political violence in the best manner possible.

Disclaimer

The Center for Crisis Early Warning is funded by the German Federal Ministry of Defense and the German Federal Foreign Office. The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any agency of the German government.

REFERENCES

Brandt, P. T., D'Orazio, V., Khan, L., Li, Y.-F., Osorio, J., & Sianan, M. (2022). Conflict forecasting with event data and spatio-temporal graph convolutional networks. *International Interactions*, *48*(4), 800–822.

Buhaug, H., Cederman, L.-E., & Gleditsch, K. S. (2014). Square Pegs in Round Holes: Inequalities, Grievances, and Civil War. *International Studies Quarterly*, *58*(2), 418–431.

Burke, M. B., Miguel, E., Satyanath, S., Dykema, J. A., & Lobell, D. B. (2009). Warming increases the risk of civil war in Africa. *Proceedings of the National Academy of Sciences*, *106*(49), 20670–20674.

Collier, P., & Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford Economic Papers*, *56*(4), 563–595.

Croicu, M., & Maase, S. P. von der. (2025). *From Newswire to Nexus: Using text-based actor embeddings and transformer networks to forecast conflict dynamics*. arXiv: 2501.03928 [cs].

Davies, S., Pettersson, T., Sollenberg, M., & Öberg, M. (2025). Organized violence 1989–2024, and the challenges of identifying civilian victims. *Journal of Peace Research*, *62*(4), 1223–1240.

Dong, Y., Chawla, N. V., & Swami, A. (2017). Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 135–144.

Dorff, C., Gallop, M., & Minhas, S. (2020). Networks of Violence: Predicting Conflict in Nigeria. *The Journal of Politics*, *82*(2), 476–493.

Dorff, C., Gallop, M., & Minhas, S. (2023). Network Competition and Civilian Targeting during Civil Conflict. *British Journal of Political Science*, *53*(2), 441–459.

Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, Insurgency, and Civil War. *American Political Science Review*, *97*(1), 75–90.

Gleditsch, K. S., & Ward, M. D. (1999). A revised list of independent states since the congress of Vienna. *International Interactions*, *25*(4), 393–413.

Gleditsch, K. S. (2007). Transnational Dimensions of Civil War. *Journal of Peace Research*, *44*(3), 293–309.

Hegre, H., Vesco, P., Colaresi, M., & Vestby, J. (2023). The 2023/24 VIEWS Prediction competition Predicting the number of fatalities in armed conflict, with uncertainty.

Högbladh, S. (2025). UCDP GED Codebook version 25.1. *Department of Peace*, *Conflict Research, Uppsala University*.

Li, Y.-F., Dong, B., Khan, L., Thuraisingham, B., Brandt, P. T., & D'Orazio, V. J. (2021). Data-Driven Time Series Forecasting for Social Studies Using Spatio-Temporal Graph Neural Networks. *Proceedings of the Conference on Information Technology for Social Good*, 61–66.

Maase, S. P. von der. (2025). *Next-Generation Conflict Forecasting: Unleashing Predictive Patterns through Spatiotemporal Learning*. arXiv: 2506.14817 [stat].

Metternich, N. W., Çiflikli, G., Weber, S., Rickard, K., & Lomax, G. (2019). *Predicting the severity of civil wars: An actor-centric approach*.

Peixoto, T. P. (2014). The graph-tool python library. *figshare*.

Raleigh, C., Kishi, R., & Linke, A. (2023). Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications*, *10*(1), 74.

Runfola, D., Anderson, A., Baier, H., Crittenden, M., Dowker, E., Fuhrig, S., Goodman, S., Grimsley, G., Layko, R., Melville, G., Mulder, M., Oberman, R., Panganiban, J., Peck, A., Seitz, L., Shea, S., Slevin, H., Youngerman, R., & Hobbs, L. (2020). geoBoundaries: A global database of political administrative boundaries. *PLOS ONE*, *15*(4), e0231866.

Sakstrup, C., & Bartusevičius, H. (2024). Group-Based Injustice, but Not Group-Based Economic Inequality, Predicts Political Violence Across 18 African Countries. *Social Psychological and Personality Science*, *15*(7), 747–758.

Sundberg, R., & Melander, E. (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research*, *50*(4), 523–532.

Tollefsen, A. F., Strand, H., & Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, *49*(2), 363–374.

Vesco, P., Hegre, H., Colaresi, M., Jansen, R. B., Lo, A., Reisch, G., & Weidmann, N. B. (2022). United they stand: Findings from an escalation prediction competition. *International Interactions*, *48*(4), 860–896.

Ward, M. D., Greenhill, B. D., & Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, *47*(4), 363–375.