# Classifying Dyads for Militarized Conflict Analysis

**Niklas Stoehr**[ς] **Lucas Torroba Hennigen**[ɘ] **Samin Ahbab**
**Robert West**[ŋ] **Ryan Cotterell**[ς,ð]
[ς]ETH Zürich   [ɘ]MIT   [ŋ]EPFL   [ð]University of Cambridge
niklas.stoehr@inf.ethz.ch lucastor@mit.edu saminahbab0@gmail.com
robert.west@epfl.ch ryan.cotterell@inf.ethz.ch

## Abstract

Understanding the origins of militarized conflict is a complex, yet important undertaking. Existing research seeks to build this understanding by considering bi-lateral relationships between entity pairs (dyadic causes) and multi-lateral relationships among multiple entities (systemic causes). The aim of this work is to compare these two causes in terms of how they correlate with conflict between two entities. We do this by devising a set of textual and graph-based features which represent each of the causes. The features are extracted from Wikipedia and modeled as a large graph. Nodes in this graph represent entities connected by labeled edges representing ally or enemy-relationships. This allows casting the problem as an edge classification task, which we term dyad classification. We propose and evaluate classifiers to determine if a particular pair of entities are allies or enemies. Our results suggest that our systemic features might be slightly better correlates of conflict. Further, we find that Wikipedia articles of allies are semantically more similar than enemies.[1]

## 1 Introduction

Researchers have long sought to understand the underlying causes of militarized conflict. The origins of conflict can be broadly categorized as either **dyadic** or **systemic**. Dyadic pertains to entity-specific idiosyncrasies, competing ideologies (Leader Maynard, 2019), e.g., dissimilar political systems (Rousseau et al., 1996), and power differentials, e.g., economic and demographic capabilities (Geller, 1993). The *Mali War* is an example of a conflict to which a dyadic cause has been attributed: It was spawned from differing cultural and ideological identities between the *Azawad Liberation Movement* and the *Malian government* (Chauzal and Damme, 2015). Throughout this

---

[1]Our dataset can be explored in an interactive dashboard: https://conflict-ai.github.io/conflictwiki. Data, code and documentations are provided at https://github.com/conflict-ai/conflictwiki.
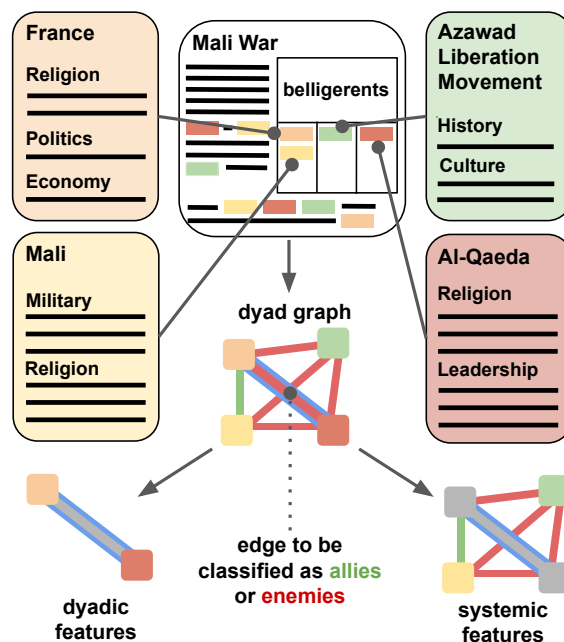


Figure 1: Overview of our approach. From Wikipedia articles on conflicts, we extract the belligerents table in the infobox. This allows constructing a dyad graph in which nodes represent entities connected by labeled edges representing ALLY or ENEMY-relationships. We then compare dyadic and systemic features in terms of how effective they are at classifying two entities as allies or enemies. We term this task dyad classification.

paper, we use the term dyad to not only denote conflictual entity pairs (ENEMIES) but also cooperative pairs (ALLIES) in a conflict (Geller, 1993).

A systemic cause, on the other hand, is a cause based on the broader relationship network involving a larger set of entities (Sweeney, 2004; Rasler and Thompson, 2010). For instance, the intervention of *France* in the *Mali War* may be said to have had a systemic cause as its origins may be partly attributed to NATO's close diplomatic ties with the Economic Community of West African States (ECOWAS; Francis, 2013). Determining a systemic cause may be aided by the analysis of a graph that encodes the relationships between the entities in the conflict. Another example of a

```
{{Infobox military conflict
| conflict   = Mali War
| place      = northern Mali
| result     = ongoing
| combatant1 = {Government of Mali,
                France,...}
| combatant2 = {National Movement...}
| combatant3 = {Al-Qaeda,...}
| date       = 16 January 2012 –
                present
```

Figure 2: Example of the template used for displaying belligerents in the infobox of Wikipedia conflict articles. The left hand side shows what the Wikipedia infobox template and the included metadata look like. Note that there may be two or more combatant tags indicating opposing conflict parties. The right hand side of the figure shows the relevant section of the infobox that contains the belligerents.

systemic origin of conflict is the ancient proverb "the enemy of my enemy is my friend", which is also known as the **structural balance theory** (Heider, 1946; Cartwright and Harary, 1956).

In this paper, we construct textual and graph-based features that encode dyadic and systemic correlates of conflict. We take this approach since establishing causality from our data is a highly complex endeavor, so we focus on correlates instead. Our approach uses Wikipedia data to compare the ability of classifiers trained using these features to predict whether two entities are allies or enemies. This is illustrated at a high level in Fig. 1. We then perform an ablation study by systematically leaving out dyadic and systemic features to ascertain to what degree these features correlate with whether two entities are enemies or allies in a conflict. Our systemic model obtains an F1 score of $0.917$ and our dyadic model obtains an F1 score of $0.873$. If one believes our features to be representative dyadic and systemic correlates, then this result provides support for the claim that, in aggregate, systemic causes may play a slightly larger role. Moreover, we also find that articles of allies are semantically more similar than enemies.

## 2 Dyadic and Systemic Features

The larger scientific mission of this paper is to investigate whether, when analyzing a large corpus of conflicts, dyadic or systemic correlates of conflict are more prominent. To carry out such a study, we construct features that encode the notions of dyadic and systemic and train classifiers to predict whether two entities are enemies or allies using these features. The classifiers are designed to operate on a **dyad graph**, an undirected graph where each node corresponds to an entity and each edge corresponds to the relationship between two entities. This allows casting the problem as an edge classification task, which we term **dyad classification**. In this section we describe the features accessible by both models, deferring the construction of the dyad graph from Wikipedia to §3 and the actual technical implementation of the models to §4.

**Notation.** Let $G = (\mathcal{N}, \mathcal{E})$ be the dyad graph consisting of entity nodes $\mathcal{N}$ and labeled relationship edges $\mathcal{E}$. This graph can be equivalently represented by the set of all dyads $D = \{d_i\}_{i=1}^{|D|}$. Each dyad $d_i = (u_i, v_i, e_i, y_i)$ consists of two entities $u_i, v_i \in \mathcal{N}$ connected by an edge $e_i = (u_i, v_i) \in \mathcal{E}$, which is labeled as $y_i \in \{\text{ALLIES}, \text{ENEMIES}\}$.

**Dyadic features.** The adjective **dyadic** is derived from the noun dyad, which is the basic unit of a militarized conflict and describes a pair of warring entities (Harbom et al., 2008). Throughout this paper, we expand the use of the term dyad to not only denote conflictual entity pairs (ENEMIES) but also cooperative pairs (ALLIES) in a given conflict (Geller, 1993). Dyadic correlates pertain to idiosyncrasies of two entities and their bilateral relationship. This suggests that suitable dyadic features would be any information directly associated with a dyad. Particularly, as dyadic features, we consider the representations of both entities, $u_i$ and $v_i$, and the unlabeled edge between them, $e_i$.

**Systemic features.** Systemic correlates are contained within the wider network of relationships two entities are embedded in. Hence, we can think of systemic features as those that are exposed by a **restricted dyad graph** $G_{\setminus d_i}$, defined as the dyad
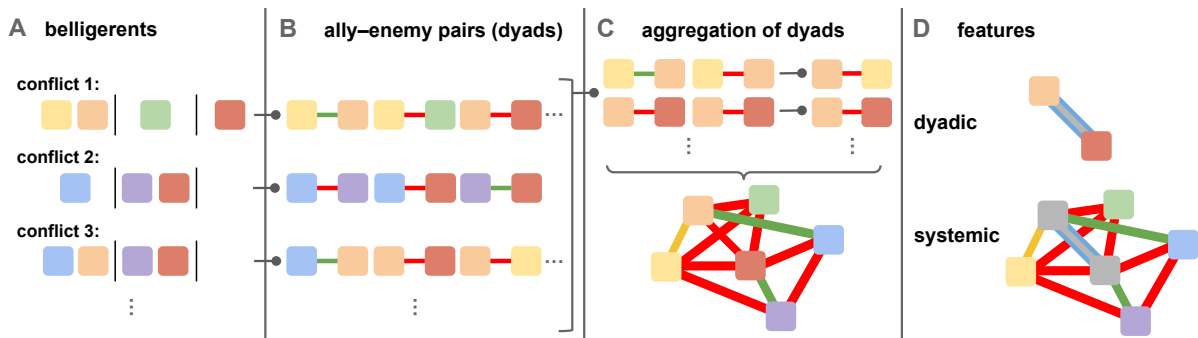
Figure 3: Construction of dyad graph; (A) Entities in each conflict are partitioned into belligerents; (B) We construct entity pairs (dyads) from all combinations of belligerents in a conflict; (C) We aggregate dyads across conflicts into a graph where nodes are entities and edges are conflicts; (D) When considering dyadic features, we only expose to the model the dyad that is meant to be classified, but when considering systemic features, we expose the graph information of everything but the dyad to be classified.

graph $G$ minus the dyad $d_i$ that is to be classified:

$$G_{\backslash d_i} = (\mathcal{N}_{\backslash d_i}, \mathcal{E}_{\backslash d_i}) \qquad (1)$$
$$= (\mathcal{N} - \{u_i, v_i\}, \mathcal{E} - \{(u_i, v_i, y_i)\})$$

Specifically, our systemic features are the representations of neighboring entities $\mathcal{N}_{\backslash d_i}$, the representations of their relationships $\mathcal{E}_{\backslash d_i}$, and the labels of those relationships, which we denote $\mathcal{E}_{\backslash d_i}^{\text{lab}}$.

## 3 Constructing the Dyad Graph

We now turn to the problem of extracting a dyad graph from Wikipedia. We first retrieve conflict, e.g., the *Mali War*, and entity, e.g., *France*, *Al-Qaeda*, articles from Wikipedia[2] (§3.1) and pre-process the articles to obtain vector representations of articles and their sections (§4.1). The resulting ConflictWiki dataset is a collection of articles on militarized conflict and their involved entities. Data, code and documentation are publicly available in an interactive dashboard. A subset of the dyad graph is shown in Fig. 4.

### 3.1 Data Retrieval

To obtain conflict articles, we first extract all articles from the Wikipedia subcategory Category:21st-century conflicts by year, offering a collection of all conflict articles from 2001 to 2021. We then recursively extract all articles in all of its subcategories up to a depth of 4 levels. While this procedure ensures wide coverage, it includes various articles which do not describe militarized conflicts but instead conflict entities, political figures, movements or geographic locations. For this

reason, we filter our selection based on a precise militarized conflict criterion—we discard all articles that do not feature at least two belligerents as indicated by the tags combatant1 and combatant2 in their infobox (see Template:Infobox military conflict and Fig. 2). Due to inconsistencies in the usage of tags, our extraction steps require a considerable number of regular expressions.[3] The whole procedure leaves us with 1145 annotated militarized conflicts over a period of 20 years.

To obtain the Wikipedia articles for all entities involved in a conflict, we consider the combatant tags in each conflict article's infobox (see Template:Infobox military conflict and Fig. 2).[4] There may be two or more combatant tags, indicating opposing conflict parties which are displayed as belligerents in Wikipedia. Each belligerent comprises one or more entities (states, militias, etc.) that are united as allies in a particular conflict. Entities assigned to different belligerents are enemies in that conflict. Almost all entities are hyperlinked to their own Wikipedia articles, which we retrieve. All together, we gather 1245 articles of entities that are involved in at least one conflict.[5]

---

## 3.2 Dyad Graph Construction

The 3-step process for building a dyad graph is depicted in Fig. 3A,B,C. The retrieved data yields a set of conflicts. Each one of these conflicts can be thought of as a group of warring factions, which we call **belligerents**. This is illustrated in Fig. 3A. For example, for the *Mali War* conflict, we have three sets of belligerents: (1) {*Mali, France*}, (2) {*Azawad Liberation Movement*}, and (3) {*Al-Qaeda*}. Next, we construct a set of ally–enemy pairs for each conflict by taking the Cartesian product of all entities involved in a conflict. For each pair of entities in a conflict, we take them to be enemies in that conflict if the two entities are in different belligerent sets, and as allies otherwise. This is displayed in Fig. 3B, where green edges represent allies and red edges represent enemies.

Next, we aggregate all ally–enemy pairs across conflicts to construct a graph $G = (\mathcal{N}, \mathcal{E})$ as displayed in Fig. 3C.[6] The set of nodes $\mathcal{N}$ represents the set of all entities and the set of edges $\mathcal{E}$ represents the bilateral relationships between all entities that have been engaged in at least one conflict together, where multiple conflicts between a pair of entities are aggregated into a single edge. We label an edge as allies if, across all conflicts they partook together, the two entities have been allies strictly more often than enemies; otherwise the edge is labeled as enemies. Note that entities that do not co-occur in a conflict have no edge between them. The resulting dyad graph contains a total of 26,536 ally–enemy edges, with 55% of them being labeled as allies. A subset of it is displayed in Fig. 4.

## 4 Experimental Setup

What remains to be discussed is the conversion of articles to a machine-readable representation (§4.1), the technical implementation of each of these models, and the baselines and setup of the experiments (§4.2).

### 4.1 Article and Section Features

Having collected the raw data from Wikipedia and constructed the dyad graph, we need to pre-process ConflictWiki so that it can be used as input to our models. However, there are two challenges associated with pre-processing the data we retrieved from Wikipedia. The first is that Wikipedia articles for
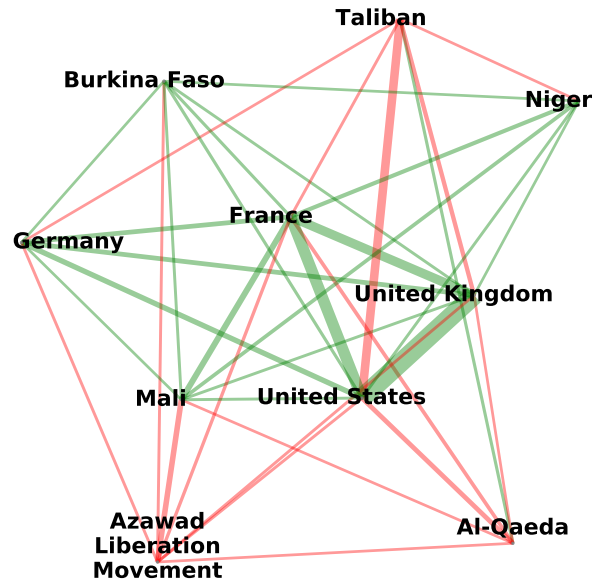
Figure 4: A small subset of the aggregated dyad graph. Each node represents an entity, each edge represents the relationship (green indicates allies and red indicates enemies) of two entities that participated together in at least one conflict. The edge line width is proportional to the number of conflicts shared by both entities; note that we do this only for illustrative purposes, to emphasize the aggregation of multiple conflict edges into one. The models do not have access to this information.

both conflicts and entities can contain explicit mentions of ally–enemy relationships. For instance, the conflict article *Mali War* states[7] that the *Azawad Liberation Movement* "began fighting a campaign against" the *Malian government*, and the entity article *France* says that the country "intervened to help the *Malian Army*". This poses a dilemma: How can we guarantee that a model that takes as input a Wikipedia article is not simply using superficial linguistic cues to regurgitate these relationships? After all, if this is all that a model does, we could hardly attribute its success to whether it is dyadic or systemic! Therefore, for a fair comparison, we need to ensure that our data does not contain explicit mentions of such relationships.[8]

The second challenge is that we must not inadvertently provide more information to our models than the information in our features (§2). For example, pre-trained representations have been shown to encode a plethora of information that may skew predictions (Kutuzov et al., 2017; Petroni et al., 2019; Bouraoui et al., 2020). Hence, for our exper-

---

[6]Most graph processing is done with the help of the *networkX library* (Hagberg et al., 2008).

[7]as of 14 September 2021

[8]We believe that a model that exploits explicit mentions of ally–enemy relationships in the text is better analyzed through the lens of information extraction and machine reading.

| Conflict | Top 10 unigrams |
|----------|-----------------|
| Mali War | soldier, town, attack, troop, rebel, group, city, conflict, northern, hostage |
| Mali | woman, country, coup, population, president, align, region, control, popular, rate |
| France | world, country, large, region, territory, department, nuclear, language, population, tourist |
| Al-Qaeda | attack, group, bombing, organization, militant, member, muslim, leader, senior, government |
| Azawad | movement, city, army, government, control, independence, military, northern, force, fighter |

Table 1: Top 10 unigrams by tf-idf weighting of different Wikipedia articles; *Mali War* conflict article and four involved entities *Mali*, *France*, *Al-Qaeda* and *Azawad Liberation Movement*.

iments we steer away from pre-trained representations and instead learn all parameters of the model from scratch, using only the data they should have access to.

Due to the challenges listed above, we use term frequency inverse-document frequency (tf-idf; Manning and Schütze, 1999) to compute vector representations of each section[9] of every entity and conflict article. We construct two separate corpora for unigram tokens appearing in conflict and entity articles. Next, we pre-process the corpora following several steps: we filter all tokens that are neither nouns nor adjectives and lemmatise all tokens. The last pre-processing steps pertain to the removal of all named entities using spaCy (Matthew et al., 2020). Particularly, we remove context-indicative tokens such as locations (e.g., Mali), dates (e.g., 2012), nationalities (e.g., French), political groups (e.g., Democrats) and organisations (e.g., Al-Qaeda), but keep world religions. Finally, we transform the unigram distribution of each article and each article section into a 500-dimensional tf-idf feature vector. To this end, we filter tokens appearing in more than 40% and less than 1% of articles.

---

[9]Since the first section of each article usually has no title, we denote it as *Summary*. Article sections with headers such as *See also*, *Bibliography*, *References*, *Further reading*, *Sources*, *Literature*, *External links*, *Citations*, *Footnotes* and *Notes* are removed.

Then, we select the top 500 tokens based on absolute term frequency across the corpus. Tab. 1 shows the top 10 unigrams by tf-idf weighting of conflict and entity articles associated with the *Mali War*.

### 4.2 Model Implementation

We implement the two main models—one that exploits the dyadic features and one with systemic features—alongside a combined model that has access to both features. Recall that, in the dyad graph, every node represents an entity, and edges between entities represent their enemy or ally relationship across one or more conflicts (§3.2). We use the tf-idf vectors of entity articles as **node embeddings**, and the average tf-idf vectors across all conflict articles associated with an edge as **edge embeddings**.

i) **Dyadic model** (D): The dyadic model has access to dyadic features only (see top half of Fig. 3D). It takes the node and edge embeddings of a dyad and passes them through multilayer perceptron (MLP) node and edge encoders, respectively. Then, the node and edge embeddings are mean-aggregated at both nodes of the dyad. The averaged embeddings are passed through another MLP and combined through a dot product, which is finally passed through a sigmoid function.

ii) **Systemic model** (S): The systemic model has access to systemic features only (see bottom half of Fig. 3D). Concretely, it passes all node and edge embeddings through MLP node and edge encoders, except those of the dyad to be classified. Next, the node embeddings are used to initialize a graph isomorphism network (GIN; Xu et al., 2019) with learnable parameters. In the GIN, edges representing an enemy relationship are weighted by $-1$ and allies by $+1$, with the edge of the dyad being excluded. After a fixed number of message passing steps with the GIN, the resulting node embeddings are mean-aggregated with the edge embeddings, passed through an MLP, and combined as in the dyadic model.

iii) **Combined** (C): The combined model has access to both, dyadic and systemic features. Concretely, it passes all node and edge embeddings, including those of the dyad, through the node and edge encoders and uses all node embeddings to initialize the GIN. The edge of the dyad is of course not weighted to

| | | Dyadic features | | | Systemic features | | | |
|---|---|---|---|---|---|---|---|---|
| | | $u_i$ | $v_i$ | $e_i$ | $\mathcal{N}_{\backslash d_i}$ | $\mathcal{E}_{\backslash d_i}$ | $\mathcal{E}^{\text{lab}}_{\backslash d_i}$ | F1 score ($\mu\pm$s.d.) |
| Main models | Ⓓ | ✓ | ✓ | ✓ | | | | $0.873 \pm 0.009$ |
| | Ⓢ | | | | ✓ | ✓ | ✓ | $0.917 \pm 0.006$ |
| | Ⓒ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | $0.926 \pm 0.008$ |
| | MAJ | | | | | | | $0.649 \pm 0.009$ |
| Ablation study | ① | ✓ | ✓ | | | | | $0.836 \pm 0.013$ |
| | ② | | | | ✓ | | | $0.828 \pm 0.007$ |
| | ③ | | | | | ✓ | | $0.779 \pm 0.009$ |
| | ④ | | | | | | ✓ | $0.871 \pm 0.005$ |

Table 2: Mean results and standard deviation of dyad classification task over 10 runs. The top half of the table shows the results of our main comparison; the bottom half shows the results of our feature ablation study (§5.1). We find that the systemic model (F1 = 0.917) outperforms the dyadic model (F1 = 0.873).

hide the enemy or ally relationship. The rest of the model is identical to the dyadic model.

iv) **Majority class** (MAJ): This is a majority-class baseline which always predict that two entities are allies.

**Hyperparameter settings.** All models are implemented using PyTorch (Paszke et al., 2019) and the Deep Graph Library (DGL; Wang et al., 2019). We use the Adam optimizer with $\eta = 0.001, \beta_1 = 0.9, \beta_2 = 0.999$, which have been shown to work well in a variety of settings (Kingma and Ba, 2015). We train our models for 30 epochs, with early stopping with a patience of 3, and a batch size of 512. Based on preliminary experiments, we use 2 message passing steps for the GIN. We use ReLU activations for all MLP non-linearities in the network. We ran a grid search to determine the dimensionality of the final layers of node encoder, edge encoder, and edge classifier.[10]

**Data split and training procedure.** We randomly split the 26,536 labeled edges of our graph $G$ into a training (60%), validation (30%) and testing set (10%). During training, the entire graph is presented to the model in subgraph batches, but the loss is computed only on the training set edges. This is a form of transductive learning (Hamilton et al., 2017) that eliminates the challenging task of splitting the graph into a separate training and testing graph through sampling (as required by the inductive setting). Moreover, we believe that the transductive setting represents a more realistic

scenario, where new entities and conflicts are added to the graph as time progresses and new conflicts erupt.

# 5 Results

The results of our main comparison are shown in the top half of Tab. 2. We evaluate results in terms of the F1 score, which is the weighted average of the precision and recall. We observe a higher binary F1 score with the systemic model Ⓢ (F1 = 0.917) than with the dyadic model Ⓓ (F1 = 0.873). This difference is significant at $p < 0.05$ under a permutation test. We also find that the combined model Ⓒ achieves F1 = 0.926, slightly outperforming the models that use only dyadic or systemic features. This asserts that, if our features are to be taken as good representatives of dyadic and systemic correlates, then our results would suggest that conflicts may be better explained by systemic causes rather than dyadic ones.

We conduct two additional analyses to gain further insight into our results. The first is an ablation study of features, to shed light onto the strongest dyadic and systemic correlates (§5.1). The second is a comparative analysis of the article sections that are most similar between allies and enemies (§5.2).

## 5.1 Ablation Study of Features

We conduct an ablation study on the individual dyadic and systemic features we defined in §2. Specifically, we ask the question: out of all dyadic and systemic features, which ones are stronger correlates of militarized conflict? The results are shown in the bottom half of Tab. 2.

---

[10]Details on the grid search and the final hyperparameter values are available on the repository.
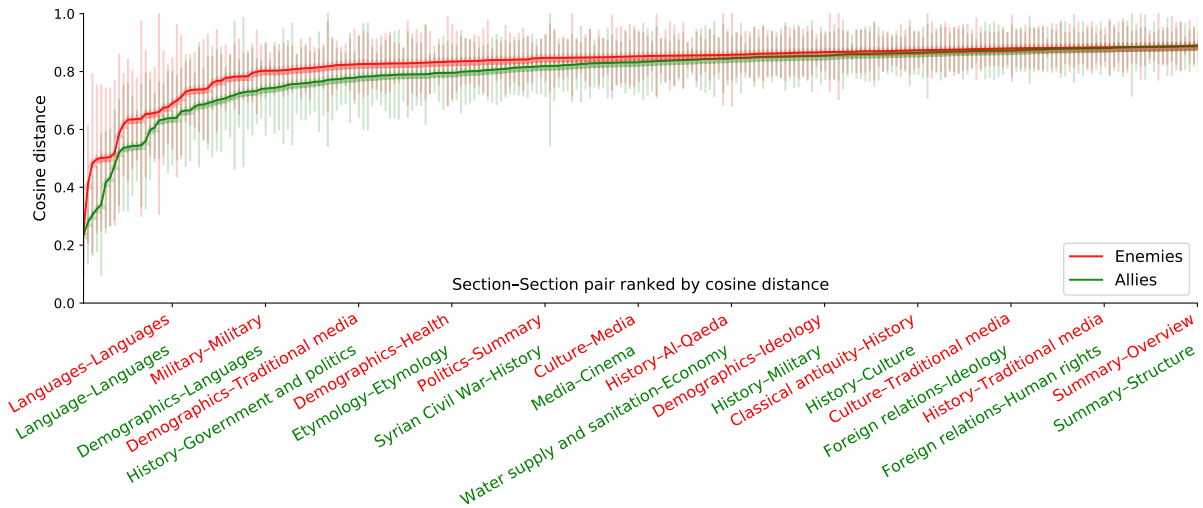
Figure 5: Top 250 most similar pairs of sections of Wikipedia articles between allies (green) and enemies (red), ranked by average cosine distance between tf-idf embeddings (standard deviation in error bars).

When leaving out the edge features of the dyad in the dyadic model ablation ① , the F1 score drops from 0.873 to 0.836. This drop suggests that the information contained within the conflict articles is complementary to entity information. Among the systemic features, we find that the model exploiting only neighboring edge labels ④ (F1 = 0.871) outperforms both the systemic model that only has access to the node features ② (F1 = 0.828), and the systemic model that only has access to the edge features ③ (F1 = 0.779). All in all, the results of our systemic feature ablations indicate that, among systemic features, the edge labels appear to be most strongly correlated with conflict; this may give some weight to structural balance theory, which seeks to understand conflict using only the labels of these edges. That said, a stronger correlate is obtained by coupling these labels with other systemic information (as evidenced by the results of Ⓢ ), which seems to indicate that conflict is very much multi-dimensional and cannot be condensed to analyzing binary relationships between entities; in particular, other systemic factors seem to also play a role in conflict.

## 5.2 Analysis of Textual Similarity

Our results suggest that the dyadic and systemic features we extracted from Wikipedia correlate, to some extent, with whether a pair of entities are allies or enemies. Indeed, some preliminary experiments show that the tf-idf representations of articles and sections are more similar among allies than enemies (for an example, see Fig. 6 where the allies *Mali* and *France* are closer to each other than to any enemy). To gain further insights into the semantic similarity of allies and enemies, we select the 1000 pairs of section titles that most frequently co-occur among allies and enemies and compute the cosine distance of their representations (e.g., Summary–Summary). We plot this in Fig. 5, with ally section pairs shown in green, and enemy ones in red. We find that distance is, on average, lower between allies (mean distance: 0.905, standard deviation: 0.063) than between enemies (mean distance: 0.912, standard deviation: 0.060) at a significance level of $p < 0.05$ under a $t$-test. This means that entities with similar articles are statistically less likely to appear as enemies in a conflict.

## 6 Related Work

**Entity relationship classification.** Most work on entity relationship classification is focused on modeling multi-dimensional relations in knowledge bases and ontologies (e.g., Riedel et al., 2010; Miwa and Bansal, 2016). The focus of our work is more similar to person-to-person sentiment analysis (West et al., 2014) since dyadic relationships are binary. There exist expert-based conflict-cooperation scales such as the Goldstein Scale (Goldstein, 1992). Structural balance theory (Heider, 1946; Cartwright and Harary, 1956) has been extended to status theory (Leskovec et al., 2010) and studied in online discussions by combining signed graphs with sentiment analysis (Hassan et al., 2012a,b). Friend and enemy relations have been studied in novels (Iyyer et al., 2016;
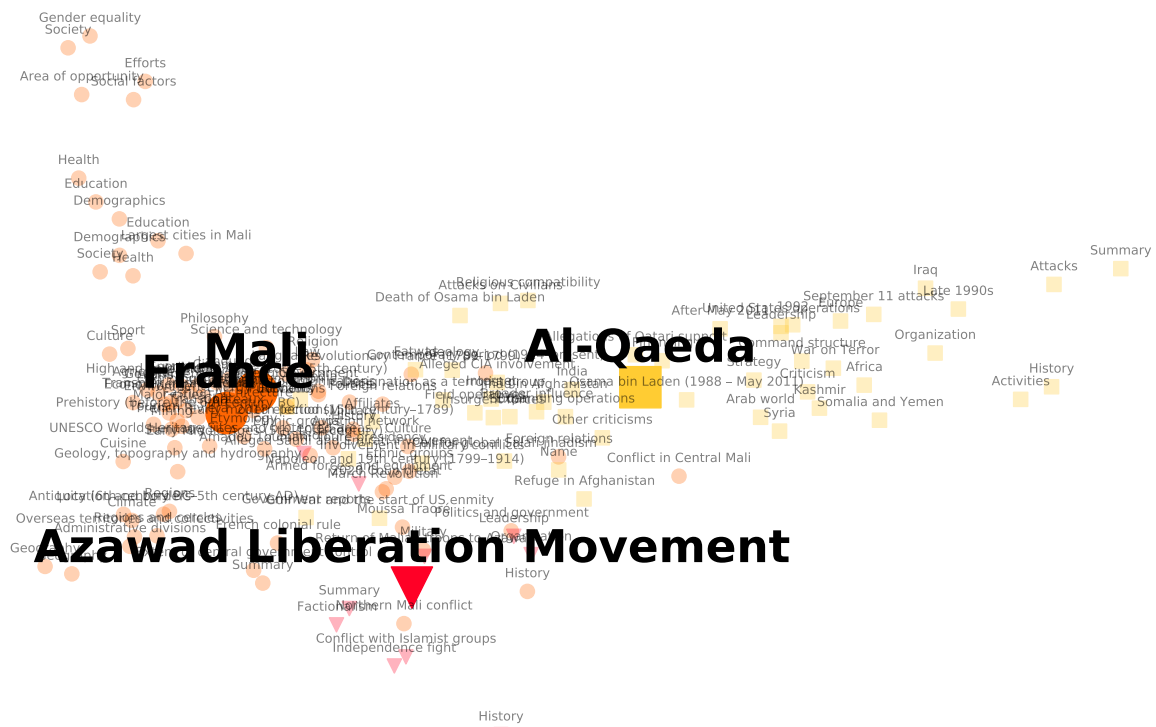
Figure 6: Top 2 principal components of tf-idf representations of the four entity articles involved in the Mali War; each belligerent is shown with the same symbol. We observe that the allies *Mali* and *France* are semantically more similar than enemies.

Srivastava et al., 2016) and international relations extracted from news (O'Connor et al., 2013; Tan et al., 2017; Han et al., 2019).

**Quantitative conflict studies.** Consistent with our work, existing empirical studies find evidence for coalescing dyadic and systemic conflict causes (de Mesquita and Lalman, 1988; Midlarsky, 1990; Geller, 1993). However, empirical studies are limited by availability of text- and graph-based data (Harbom et al., 2008). Many machine-extracted (e.g., Europe Media Monitor (EMM; Atkinson et al., 2017)) and human-curated (e.g., The Armed Conflict Location & Event Data Project (ACLED; Raleigh et al., 2010)) conflict event datasets are collections of news articles covering events of daily granularity. Associating events with their overarching long-term conflict and mentioned entities requires complex co-reference resolution (Radford, 2020). The UCDP Global Event Dataset (GED; Sundberg and Melander, 2013), UCDP Dyadic Dataset (Harbom et al., 2008) and Correlates of War (CoW; Reid and Wayman, 2010) are among the few datasets that associate individual events with overarching conflicts. The UCDP Dyadic Dataset is closest to our dataset, but limited to 3000 dyads and does not feature textual descriptions. Re-

lated militarized conflict analyses focus on news coverage (West and Pfeffer, 2017), interpretable topic models (Mueller and Rauh, 2018) and graph neural networks for event detection (Nguyen and Grishman, 2018; Cui et al., 2020).

## 7  Conclusion

This work explores the extent to which dyadic and systemic features correlate with whether two entities are allies or enemies. Our results suggest that both features are correlated, although, if one is to believe our featurizations and models, systemic features appear to be more correlated. We conduct an ablation study to identify the overall contribution of individual dyadic and systemic features, and a textual similarity study which shows that articles of allies exhibit more similarity than those of enemies.

## Impact Statement

The authors foresee no ethical concerns with the research presented in this paper.

# References

Martin Atkinson, Jakub Piskorski, Hristo Tanev, and Vanni Zavarella. 2017. On the creation of a security-related event corpus. In *Proceedings of the Events and Stories in the News Workshop*, pages 59–65, Vancouver, Canada. Association for Computational Linguistics.

Zied Bouraoui, Jose Camacho-Collados, and Steven Schockaert. 2020. Inducing relational knowledge from BERT. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):7456–7463.

Dorwin Cartwright and Frank Harary. 1956. Structural balance: A generalization of Heider's theory. *Psychological Review*, 63(5):277–293.

Gregory Chauzal and Thibault van Damme. 2015. The roots of Mali's conflict: Moving beyond the 2012 crisis. *CRU report*.

Shiyao Cui, Bowen Yu, Tingwen Liu, Zhenyu Zhang, Xuebin Wang, and Jinqiao Shi. 2020. Edge-enhanced graph convolution networks for event detection with syntactic relation. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2329–2339, Online. Association for Computational Linguistics.

David Francis. 2013. The regional impact of the armed conflict and French intervention in Mali. Technical report, Norwegian Peacebuilding Resource Center.

Daniel Geller. 1993. Power differentials and war in rival dyads. *International Studies Quarterly*, 37(2):173.

Joshua Goldstein. 1992. A conflict-cooperation scale for WEIS events data. *The Journal of Conflict Resolution*, 36(2):369–385. Publisher: Sage Publications, Inc.

Aric Hagberg, Daniel Schult, and Pieter Swart. 2008. Exploring network structure, dynamics, and function using NetworkX. *Proceedings of the 7th Python in Science Conference (SciPy2008)*.

William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 1025–1035, Red Hook, NY, USA. Curran Associates Inc. Event-place: Long Beach, California, USA.

Xiaochuang Han, Eunsol Choi, and Chenhao Tan. 2019. No permanent friends or enemies: Tracking relationships between nations from news. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pages 1660–1676.

Lotta Harbom, Erik Melander, and Peter Wallensteen. 2008. Dyadic dimensions of armed conflict, 1946–2007. *Journal of Peace Research*, 45(5):697–710.

Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012a. Detecting subgroups in online discussions by modeling positive and negative relations among participants. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 59–70.

Ahmed Hassan, Amjad Abu-Jbara, and Dragomir Radev. 2012b. Extracting signed social networks from text. In *Workshop Proceedings of TextGraphs*, pages 6–14.

Fritz Heider. 1946. *The psychology of interpersonal relations.* John Wiley & Sons Inc, Hoboken.

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1534–1544, San Diego, California.

Diederik Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*, page 337, Ithaca, NY.

Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid. 2017. Tracing armed conflicts with diachronic word embedding models. In *Proceedings of the Events and Stories in the News Workshop*, pages 31–36, Vancouver, Canada.

Jonathan Leader Maynard. 2019. Ideology and armed conflict. *Journal of Peace Research*, 56(5):635–649.

Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1361, Atlanta, Georgia, USA. ACM Press.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing.* MIT Press, Cambridge, MA, USA.

Honnibal Matthew, Montani Ines, Van Landeghem Sofie, and Boyd, Adriane. 2020. spaCy: Industrial-strength natural language processing in Python.

Bruce Bueno de Mesquita and David Lalman. 1988. Empirical support for systemic and dyadic explanations of international conflict. *World Politics*, 41(1):1–20.

Manus I. Midlarsky. 1990. Systemic wars and dyadic wars: No single theory. *International Interactions*, 16(3):171–181.

Makoto Miwa and Mohit Bansal. 2016. End-to-end relation extraction using LSTMs on sequences and tree structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*

*(Volume 1: Long Papers)*, pages 1105–1116, Berlin, Germany. Association for Computational Linguistics.

Hannes Mueller and Christopher Rauh. 2018. Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, 112(2):358–375.

Thien Nguyen and Ralph Grishman. 2018. Graph convolutional networks with argument-aware pooling for event detection. In *Thirty-Second AAAI Conference on Artificial Intelligence*, volume 32.

Brendan O'Connor, Brandon Stewart, and Noah Smith. 2013. Learning to extract international relations from political context. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 1094–1104.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, and et al. 2019. *PyTorch: An imperative style, high-performance deep learning library*, page 8024–8035. Curran Associates, Inc.

Fabio Petroni, Tim Rocktäschel, Sebastian Riedel, Patrick Lewis, Anton Bakhtin, Yuxiang Wu, and Alexander Miller. 2019. Language models as knowledge bases? In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 2463–2473, Hong Kong, China. Association for Computational Linguistics.

Benjamin Radford. 2020. Seeing the forest and the trees: Detection and cross-document coreference resolution of militarized interstate disputes. In *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News 2020*, pages 35–41. European Language Resources Association (ELRA).

Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: An armed conflict location and event dataset: Special data feature. *Journal of Peace Research*, 47(5):651–660.

Karen A. Rasler and William R. Thompson. 2010. Systemic theories of conflict. In *Oxford Research Encyclopedia of International Studies*. Oxford University Press.

Sarkees Reid, Meredith and Frank Wayman. 2010. Resort to war: 1816–2007. *CQ Press*.

Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling relations and their mentions without labeled text. In *Machine Learning and Knowledge Discovery in Databases*, pages 148–163, Berlin, Heidelberg. Springer Berlin Heidelberg.

David L. Rousseau, Christopher Gelpi, Dan Reiter, and Paul K. Huth. 1996. Assessing the dyadic nature of the democratic peace, 1918–88. *American Political Science Review*, 90(3):512–533.

Shashank Srivastava, Snigdha Chaturvedi, and Tom Mitchell. 2016. Inferring interpersonal relations in narrative summaries. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 2807–2813.

Ralph Sundberg and Erik Melander. 2013. Introducing the UCDP georeferenced event dataset. *Journal of Peace Research*, 50(4):523–532.

Kevin Sweeney. 2004. *A dyadic theory of conflict: Power and interests in world politics*. Ph.D. thesis, Ohio State University.

Chenhao Tan, Dallas Card, and Noah A. Smith. 2017. Friendships, rivalries, and trysts: Characterizing relations between ideas in texts. In *Proceedings of the Association for Computational Linguistics*, pages 773–783, Vancouver, Canada. Association for Computational Linguistics.

Minjie Wang, Da Zheng, Zihao Ye, Quan Gan, Mufei Li, Xiang Song, Jinjing Zhou, Chao Ma, Lingfan Yu, Yu Gai, Tianjun Xiao, Tong He, George Karypis, Jinyang Li, and Zheng Zhang. 2019. Deep graph library: A graph-centric, highly-performant package for graph neural networks. *arXiv preprint arXiv:1909.01315*.

Robert West, Hristo S. Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2.

Robert West and Jürgen Pfeffer. 2017. Armed conflicts in online news: A multilingual study. In *Proceedings of the International AAAI Conference on Web and Social Media*.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How powerful are graph neural networks? In *International Conference on Learning Representations*.