Measuring Spatio-Temporal Civil War Dimensions Using Community-Based Dynamic Network Representation (CoDNet)

Didier A. Vega-Oliveros and Ore Koren

Abstract

Civil war exhibits complex geospatial trends over time, which may be missed by models that rely on count-based operationalizations. Here, the spatial and temporal correlation values of monthly civil war events are transformed into their influence degree symbol, which measures geospatial concentration, spread, and intensity of civil war. We then measure variation in these degrees over time to identify relevant spatio-temporal civil war aspects. The network model is constructed using 0.5 degree grid locations as nodes, counting nearby and over time connections. We then extract the temporal community structure behind the data. We use ground truth data to visualize how our measures correlate with observed patterns, thereby illustrating our method provides accurate depictions of geospatial civil war dynamics. We also evaluate the impact of several indicators highlighted by past research and our community-based spatio-temporal measures and comparing it to the preprocessed count indicator. Our findings indicate that the relationship between state capacity and climate stress show opposite correlations with civil war as those identified by studies that use count based indicators. Counterintuitively, our results show that conflict intensifies and spreads in locations where the state is stronger and where climate conditions are improved.

Index Terms

Spatiotemporal events, civil war, community-based network analysis

Revise & Resubmit at

IEEE Transactions on Computational Social Systems

D. A. Vega-Oliveros is with the Artificial Intelligence Lab. (Recod.ai), Institute of Computing, University of Campinas, Brazil, and with the Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA. Email: davo@unicamp.br

O. Koren is an assistant professor at the Department of Political Science, Indiana University Bloomington. Email: okoren@iu.edu. Koren's research was supported by the Henry Frank Guggenheim Foundation and by the National Science Foundation under Grant No. SAP-2149053.

I. INTRODUCTION

The growing availability of fine-grained data is transforming the social science [1]. The complexity and size of such datasets privileged state-of-the-art unsupervised approaches that could identify and predict patterns across a variety of social and behavioral domains [2, 3]. A particular and growing impact is that of network analysis, which – by allowing researchers to explores links across individuals, groups, societies, and states – is especially useful for social science research [1, 4].

One area that still lags with respect to the application of network based models is civil war analysis. Detailed conflict data are now widely available [e.g., 5, 6], with vast implications for the way scholars study civil war.[1] Yet, researchers did relatively little in terms of leveraging this information to develop and deploy effective *operationalization and measurement* approaches. Rarely pre-processing these data, scholars often use one of two empirical approaches. The first is to rely on the number of casualties to define or, more often, dichotomize a measure of civil war, for instance by defining a civil war as an armed conflict with at least 25 combatant casualties [7], where intensity is defined by the number of deaths. Another approach is to use the number of armed conflict incidents where, e.g., there was at least one casualty to create a count measure of civil war, where intensity is defined by the number of incident recorded in a given country or location [8].

Incident counts and locations, however, reflect only one (observable) aspect of civil war. Like other social phenomena, civil war exhibits complex trends over space and time, which count and location of the actual events cannot effectively measure [see, e.g., 9]. Some geographic locations may have particularly high strategic importance, yet involve relatively low levels of combatant casualties (government and rebel strongholds, hard-to-reach areas, etc.). Moreover, civil war can be reflected in ways other than the number of attacks or deaths, such as the spatio-temporal clustering of violence in some regions.

A key aspect of such civil war dimensions is that they are *implicit* – aspects like strategic calculations, troop movements, and general importance of a given location cannot be directly observed in the casualty or event data. Identifying and measuring these dimensions can have important implications for our ability to better explain the causes of civil war and improve our ability to predict civil war, locally and globally. For example, recent studies have explored how complex phenomena such as environmental stressors [e.g., 10], the distribution of state power and infrastructure [e.g., 11], and migration patterns [e.g., 12], impact conflict, and applied machine-learning techniques to create effective forecasting models and identify perpetrators of violence [13, 14]. The inherent complexity of the underlying causal mechanisms in these cases suggests that a better understanding of their role will be achieved if more attention is given to the complexity of civil war itself and its underlying spatio-temporal dynamics that may not be directly inferred from the data.

Finding clusters in the data is no trivial task. Supervised methods such as k-means account for the density or distribution of points in the feature space, but they necessitate pre-defining parameters like the number of desired clusters or centroids as inputs. As several mechanisms could impact conflict clustering, this approach is problematic for modeling civil war. On the other hand, methods such as the Leiden algorithm are designed for

[1]We often use the terms “conflict” and “civil war” interchangeably throughout this study.

finding communities (or clusters) in a network without the need to pre-define features such as the number of clusters. These methods use the topology structure of the network to optimally infer the community structure behind the data. [15] discuss some of the advantages of using networks to identify underlying spatio-temporal clusters, comparing DBSCAN with a network-based community detection approach.

Similarly, we propose to quantify implicit dimensions of civil war using a community-based dynamic (temporal) network representation (CoDNet). Our approach allows for a network-based modeling of the geospatial aspects of a particular phenomenon of interest over time and is hence ideal for identifying key dynamics of armed conflict. Briefly, using CoDNet we are able to identify longitudinal-geopolitical civil war "communities," i.e., clusters of civil war events that are *directly related over space and time*, and highlight particular relevant geospatial dimensions of civil war as they are reflected in these communities. Using a dataset of location and timing information on conflict events, we identify and discuss the CoDNeT-based operationalization procedure behind three spatio-temporal civil war dimensions (for each one of these dimensions, we obtain annual quantities): (i) *community average degree*, or **average intensity** of civil war within a specific community over time; (ii) *the number of unique nodes*, or **dispersion**, which measures the number of unique cell-years nodes within a community experiencing active civil war; and (iii) *number of unique near related civil war links*, or **intraconnectivity**, which captures the level of connectivity across all unique cells within a community over time. Especially considering that network-based representation is still an under-analyzed problem in the extant research [16, 17], these measures provide an important extension in our ability to measure and understand civil war.

To illustrate the theoretical importance of focusing on these topological network-based civil war spatio-temporal dimensions to our understanding of civil war and its determinants, we first validate the results based on observed civil war patterns in several historical and ongoing civil war cases. We systematically evaluate a series of subnational civil war predictors at the global level for the years 1989-2012. In the latter, we consider key indicators highlighted by past research on civil war, and estimate a series of statistical models corresponding to each of our implicit civil war operationalizations as the dependent variable, compared to the standard (count based) conflict event indicators. We find that several key conflict determinants change their signs and significance as we switched from testing their impact on an unprocessed count indicator of conflict events to examining how they shape degree variations across our three spatio-temporal civil war indicators. The results hence illustrate that focusing only on observed civil war incidents can yield divergent, potentially inaccurate inferences about civil war.

II. Why Implicit Spatio-Temporal Dimensions Matter?

In most civil war studies, civil war is operationalized using one of two empirical approaches. The first relies on the *number of casualties* to define or, more often, create a binary measure of civil war, for instance by dichotomizing civil war as an armed conflict with at least 1,000 or 25 combatant casualties [e.g., 7, 11, 18, 19]. Yet, like other social phenomena, civil war exhibits complex trends over space and time. Different civil war-affected areas might share similarities related to the underlying nature of the conflict itself, or experience spatio-temporal dependencies related to the spread of warfare. In particular interest to scholars, especially in recent years, is the *geography* of civil war – how and where civil war spreads, clusters, intensifies, and declines [9, 19]. Understanding these patterns and their determinants will greatly improve our ability to explain and ultimately mitigate civil war risk. Improving our understanding of these geographical aspects begins with the availability of detailed and geolocated data. Unsurprisingly, then, this aim stimulated major data collections, including the Armed Conflict Location and Event Data (ACLED) and the Georeferenced Event Dataset (GED) [5, 6], as well as data frameworks constructions such as the PRIO-Grid [20] and AfroGrid [21], which provide researchers with a variety of geolocated political and socioeconomic indicators to explore the determinants and impacts of civil war.

Creating detailed high-resolution data, however, is only the first step in improving our ability to understand and explain the geography of civil war. The next necessary phase is to improve the *geographic operationalization* of civil war and its features, using these data to identify important trends and dependencies that cannot be immediately recognized. While scholars have adopted new ways of modeling and incorporating geolocated event data in their efforts to test the determinants of civil war (right-hand side variables), improving the viability of the *geography* of civil war as the dependent variable (left-hand side) necessitates a more comprehensive way of modeling its spatial and temporal dynamics, and their specific properties (e.g., conflict concentration and spread within and across locations) therein, as scholars came to recently recognize [8, 9, 14].

Network analysis offers important advantages in geospatially formalizating and operationalizing civil war patterns over time. In recent years, network-based tools have become increasingly popular in civil war research [e.g., 22, 23]. In particular, scholars deployed these tools to create new *conceptual* measures of civil war, for instance as a means of operationalizing adverse or friendly relationships between warring actors or evaluating whether deteriorating relationships between different warring actors impact dyadic relationships between other actors in the system [22, 24, 25]. While such conceptual network-based operationalizations yield important new insights, very little has been done to create network-based *geospatial* operationalizations of war. Geospatial networks – especially ones that add temporal components – can provide important new ways of operationalizing and measuring civil war as a dynamic process, which adds to the way civil war is understood, incorporating its geographic aspects more directly, in line with current research trajectories and interests. Accordingly, in this study we illustrate the effectiveness of using geospatial dynamic network tools to operationalize civil war and identify some of its implicit aspects, rather than rely on the raw (count) as the sole tool of analysis.

III. Measuring Spatio-Temporal Civil War Dimensions Using CoDNet

A. Preprocessed data

The civil war event data used in this study come from the Uppsala Conflict Data Program Georeferenced Event Data (UCDP GED) [6] dataset, which is considered one of the most comprehensive geolocated political violence datasets in existence. The UCDP GED codes a comprehensive set of civil war-related conflict events with at least one death (combatant or civilian) at the local level over the entire globe between 1989 and 2014, excluding Syria. We retain only information for the 1992-2012 period, due to limited temporal availability on some of our independent variables, explained in Section V. We focus on incidents coded as “state-based” (i.e., between a government and a rebel group) to ensure we only capture civil war events involving armed combatants (i.e., not violence against civilians) involving official state actors and anti-state groups. We also kept only incidents measured at the second (district) administrative level or below to ensure geographic compatibility across our cell-based indicators [26]. A total of 67,798 such events were recorded during the 1992-2012 period analyzed here. To ensure effective comparison between our CoDNet-preprocessed indicators and the non-modified GED data, we merge all our civil war data to a yearly PRIO-GRID framework [20] for the entire terrestrial globe. The $0.5^{\circ} \times 0.5^{\circ}$ PRIO-GRID corresponds to the lowest "grid cell" level of geographic aggregation recommended for conflict event data analyses based on past validation assessments [26], and in our case yields state-based event counts as well as different CoDNet based indicator values for each grid cell cell between 1992 and 2012. Accordingly, we estimate the effect of key predictors on a count indicator that simply sums up all annual conflict events, and compare these estimates to those related to our CoDNet based indicators.

B. Modeling civil war geospatially using networks

Because the event data are not directly represented as a network but rather as a collection of spatio-temporal events, a suitable network construction process is necessary for geospatial operationalization. Some of the benefits that network-based representation provide are in describing sub-manifold in dimensional space and capturing dynamic and topological structures – hierarchical structures (communities, motifs) and global or local patterns – regardless of the distribution of the underlying data [15, 27].

Functional or correlational approaches are well-known methods used for constructing time-series networks [28–30]. The construction process connects (highly) correlated time series among pairs of nodes. However, correlation measures may have significant problems and can generate spurious values with short-length time series such as the ones used here. Additionally, our event-based data have a large number of zero values, with many (over 99%) of our grid cells experiencing no registered civil war events over time, which also adversely affects correlation-based methods.

As an alternative, [15] formalized a method that connects chronologically spatio-temporal events that co-occurred in a given region. Indeed, spatio-temporal networks of co-occurring events have been relevant tools explored in several other domains [e.g., 31–35], although these methods have not (to our knowledge) been applied to geospatial civil war analysis. One characteristic of such approaches is their reliance on very fine grained temporal disaggregation, with information being recorded at the hour or minute level. event data, however, do not offer such a high level of temporal precision because of delayed reports, missing data, or human coding decisions. While, there is sufficient level of confidence in reporting to ensure that a given event happened within a given year, month or (sometimes) week or day, this results with large number of parallel armed events co-occurring at the same grid month, which affects our precision in connecting different civil war impacts.

To overcome this temporal limitation, we propose a construction method for sparse events building on co-occurrence geographical relatedness. Suppose a geographic region or area of interest. We seek to understand the graph topological and spatio-temporal properties of some reported events in this area occurring over a specific period of time $t = \{1, 2, ..., T\}$. In a hypothetical example, let us suppose several events were reported in five grid cells (or simply cells), i.e., $\{a, b, c, d, e\}$. Further assume that over the next year (t=year2), all the same cells

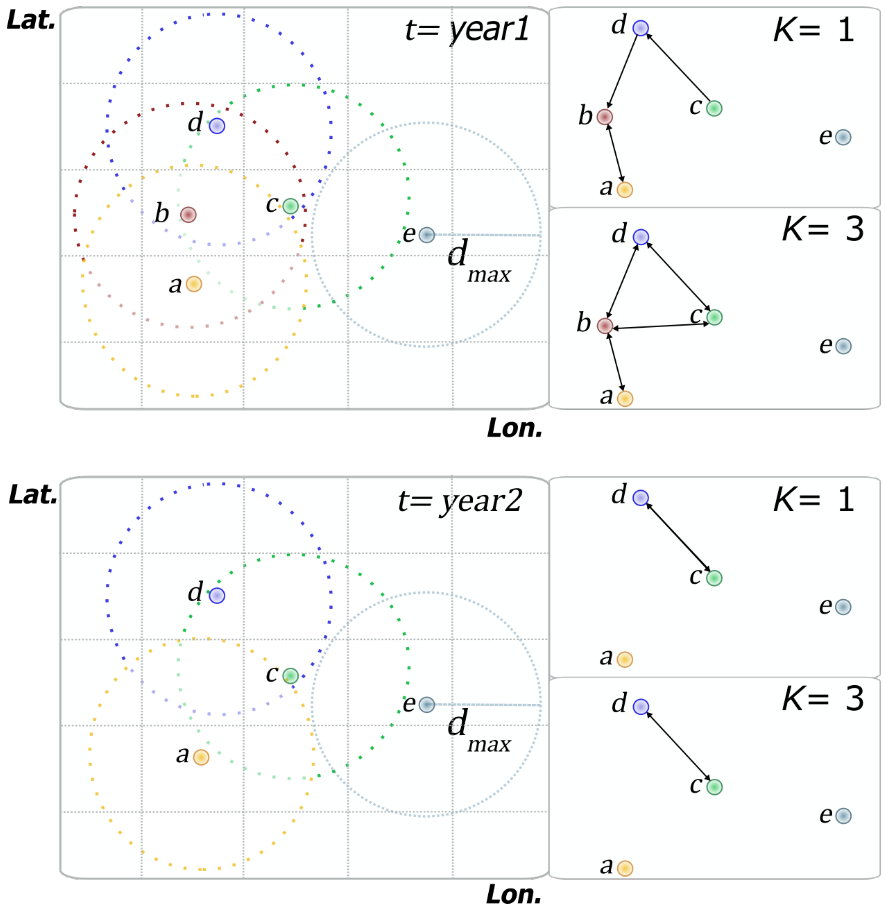


Fig. 1. A scheme of the proposed approach for modeling geospatial sparse events using network-based approaches. In the top panels, we include events that occurred within the time window corresponding to a hypothetical year. For each event, we consider the maximum distance of influence $d_{\max}$. In the network construction process, the grid-cell where the event happens becomes a node. We then connect all the k nodes located inside this perimeter. The bottom panels report the time window for events occurring during the next hypothetical year, in which events in cell b did not occur.

excluding cell b again experienced events. This pattern is illustrated in Figure 1. We define $d_{\max}$ as the maximum geographical distance for each cell, which defines the perimeter of connection or influence among the cells.

The network at time t is then specified as $G^{(t)} = (V^{(t)}, E^{(t)})$ composed by the set $V^{(t)}$ with size $|V^{(t)}| = n^{(t)}$ nodes and the set $E^{(t)}$ with size $|E^{(t)}| = m^{(t)}$ links or edges. During the construction process, all the affected cells present in the spanning window t are mapped as nodes. The linking process considers the neighbor nodes inside the perimeter of influence of each cell/node, making a directed connection with its k nearest cells by the widely-used k-NN algorithm [29, 36, 37]. For illustration, we show the cells that have been affected by events during $t = \text{year1}$ in our network representations as k equal to 1 and 3 in the top panels in Figure 1. As this figure illustrates, the number of connections can vary by the k parameter, which defines the number of out-connections for any node but not the number of in-connections. Additionally, considering its geodesic distance from the other cells, cell e remains isolated no matter what value the number of possible linkages k assumes. In this case, the defined radius $d_{\max}$ is insufficient for e to become connected to any cell. In contrast, during the next year ($t = \text{year2}$), the value of our k parameter has no impact on the structure of the spatio-temporal network considering geodesic distance across cells/nodes, as illustrated in the bottom panel of Figure 1. In this example, even though all the cells except b experienced events during $t =$ year2, the constructed networks is the same for both cases of k, and e is isolated in both $t =$ year1 and year2.

We notice that the time dimension is treated implicitly in the snapshot networks. For example, two geographically proximate conflict events that occur in different years can belong to different civil war communities due to underlying temporal differences. Additionally, neighbor civil war affected cells are connected if they are recorded as co-experiencing a civil war event in the same year and with spatial relatedness. We are not considering self-loops, i.e., repeated events in the same location, as count-based operationalization. For example, if we have one civil war-affected cell, it is represented as a single node with no connections even if it experiences multiple events during a given year (although it is important to stress that almost all cell years that experienced civil war register between one and five events). On the other hand, events occurring in three different proximate cells are treated as three connected nodes. These cases illustrate the problem of *not* accounting for dependencies *over time*, in addition to spatial dependencies. Accordingly, we account for this issue, incorporating a temporal component into our community-based dynamic (temporal) network representation (CoDNet) technique.

C. Overview of CoDNet

We construct our temporal snapshot network [38] using local conflict event data over time to identify spatio-temporal civil war dimensions in several steps. As mentioned above, in line with past research [11, 19], we focus on state-based events, i.e., civil wars involving a rebel group and a recognized state, although the method is applicable to other types of sparse events as well. To ensure comparability across units of analysis, we use the year as our temporal resolution unit (that is, for year t), while reporting a model in the SI file (Table A2) a model that uses monthly quantities for robustness; and the $0.5° \times 0.5°$ cell (i.e., about 55km x 55km at the equator) as our geospatial resolution unit, in line with past research [20]. The 0.5 degree grid cell (hereon cell i) physically defines our nodes in the temporal networks. Each of the temporal snapshots in the CoDNet indicates a period for which conflict data were available (1989-2014). Nevertheless, we constrain our civil war indicator samples to the 1992-2012 period to account for independent variable data availability, covering the entire terrestrial globe (excluding Syria, as mentioned above). We then analyze the topological properties of these civil war communities over time, which we detect using annual snapshots over the entire period of interest. We detail the applied methodology for discovering these spatio-temporal civil war communities across the entire terrestrial globe in Figure 2, and summarize it as follows.

First, let the weighted temporal network of civil war events $\mathcal{G}$ be a set of $T = 21$ (1989-2014) year-snapshots $\mathcal{G} : \{G^{(0)}, G^{(1)}, ..., G^{(T)}\}$, where each undirected $G^{(t)}$ snapshot is represented by the set $V^{(t)}$ of nodes and the set $E^{(t)}$ of edges. The geospatial network for each year is constructed using the corresponding civil war events, in addition to the k and $d_{\max}$ parameters, which discussed in Section III-B. As part of the construction process, we set the maximum geodesic distance for defining each node's perimeter as $d_{\max} = 500$ km, to ensure that distinct civil war communities existing over the same period (e.g., the civil war in Algeria and terrorist attacks in Spain)

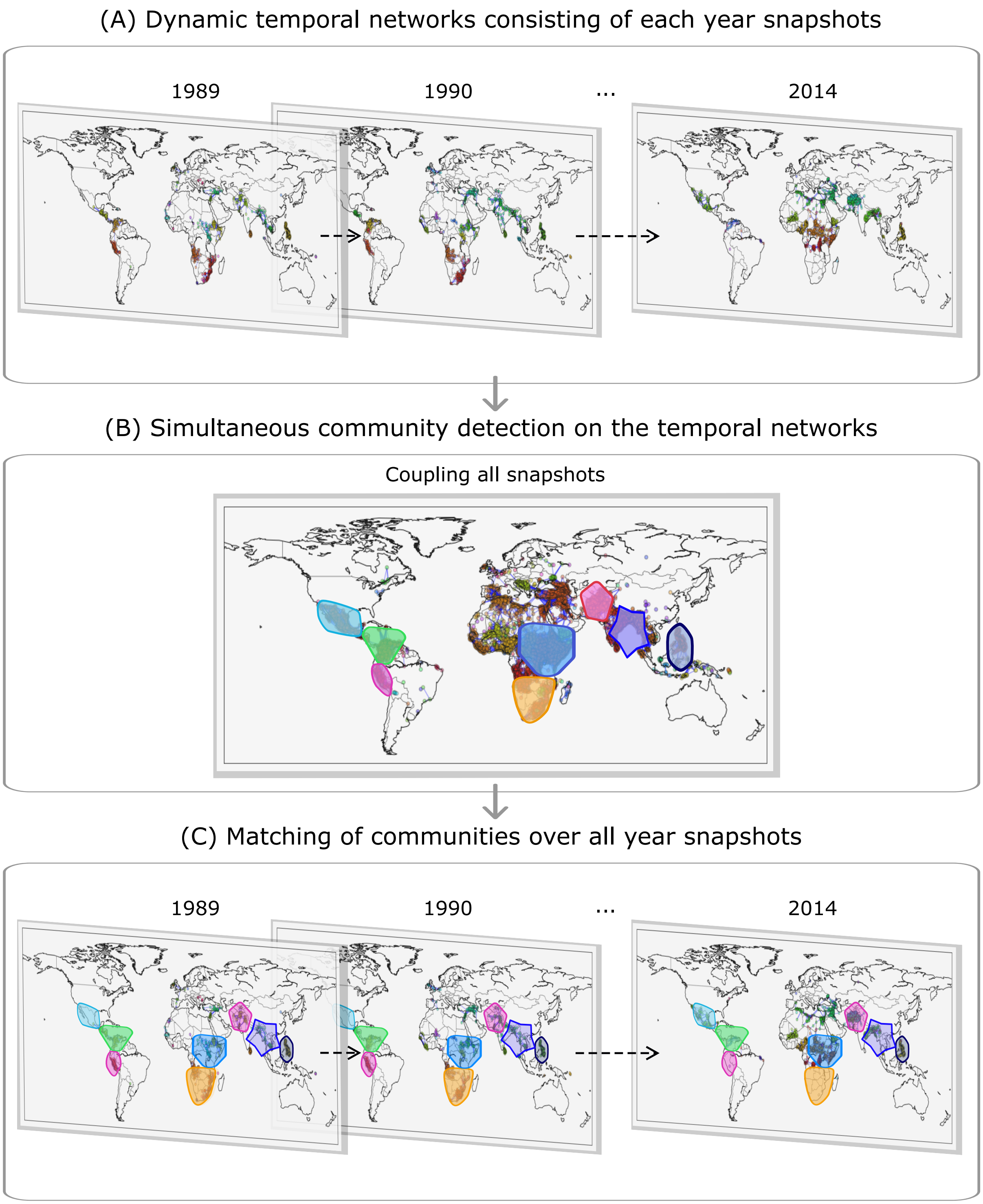


Fig. 2. Methodology of spatiotemporal community detection and matching approach.

are not being attached as the same community. In other words, two events are connected in a $G^{(t)}$ only if any of them are a nearest neighbor of the other and their geographical close, with distance lower than 500 km. In terms of the nearest neighbors parameter, we explored the values of $K = \{3, 7, 11\}$. For illustration, Figure 2(A) plots snapshots that represent the spatial network of civil war events for several randomly-selected years, where $k = 3$ and $d_{\max} = 500$ km. In the SI file (Tables A3–A5), we report relevant maps and robustness model results for different values of k as well as for a sample where there are no distance limits.

We deal with the dynamical aspects of these communities using a temporal snapshot network approach [39, 40]. Note that different spatial communities are reflected as different shapes in the snapshots, which can split, merge, expand, or reduce their area over the years [39]. To correctly match these communities, we apply a simultaneous community detection technique [40] by spatially aggregating all temporal snapshot networks. Thereby, we find the geographical and temporal communities considering all the snapshot networks simultaneously, in a generated single network $G^{(\text{all})} = \bigcup^{\mathcal{G}} G^{(t)}$ with $G^{(t)} \in \mathcal{G}$, as shown in Figure 2(B). Formally, $G^{(\text{all})} = (V^{(\text{all})}, E^{(\text{all})})$, where

$$V^{(\text{all})} = \{v_i \,|\, i \in [1, 2, \ldots, n]\,,\, v_i^{(y)} \mapsto v_i^{(x)}\, \forall x, y \in [0, \ldots, T]\,,\, n = \max(n^{(t)})\} \tag{1}$$

$$E^{(\text{all})} = \{w_{i,j} * e_{i,j} \,|\, w_{i,j} = \sum_{t=0}^{T} e_{i,j}^{(t)}\,,\, e_{i,j}^{(t)} \in E^{(t)}\,,\, e_{i,j} \in \bigcup_{t=0}^{T} E^{(t)}\,,\, e_{i,j}^{(t)}, e_{i,j} \in [0, 1]\} \tag{2}$$

in which each cell i is represented by the same node v_i across time t, i.e., node $v_i^{(t)}$ is the same node $v_i^{(t+x)}$ that describes cell i. In terms of $E^{(\text{all})}$, we join together all the links and the weight is the number of occurrence of each particular link through time. The single network represents the historical footprint of the community over all $G^{(t)}$ years of civil wars. The temporal properties of each community emerge in the network topology on each time-slice snapshot, such that even highly-proximate cells that experienced civil war in different periods are treated as distinct nodes and can therefore belong to different communities.

We then employ the multiplex Leiden community detection algorithm [41] in $G^{(\text{all})}$, which produces the set $C^{(\text{all})} = c_0^{(\text{all})}, ..., c_l^{(\text{all})}$, where l is the number of detected communities. Each $c_z^{(\text{all})}$ is the set of nodes related to the community z from the different snapshots that belongs to the specific region, i.e.,

$$c_z^{(\text{all})} = \{v_i \in V^{(\text{all})} \,|\, v_i \in \text{ community } z \text{ from } G^{(\text{all})}\} \tag{3}$$

We selected this algorithm because it scales well and can be run on networks of millions of nodes. The Leiden is an optimized extension of the Louvain algorithm [42], which ensures convergence across optimal communities locally connected in terms of modularity [41]. Notes that we are using a non-overlapping clustering method, which means nodes can only belong to one community.

At this point, we have created a unique map binding together all the annual civil wars in our data into a single network, in which we detected a set of coherent community structures corresponding to spatial civil war-afflicted regions over all our years of interest (Figure 2(B)). To identify the evolution of each community over time, we then map the historical civil war communities $C^{(\text{all})}$ into the corresponding regions for each year t snapshots, i.e.,

$$c_z^{(t)} = (V_z^{(t)}, E_z^{(t)}) \tag{4}$$

$$V_z^{(t)} = \{v_i \in V^{(t)} \mid v_i \in c_z^{(\text{all})}\}, \text{ in which } n_z^{(t)} = |V_z^{(t)}| = |c_z^{(t)}| \tag{5}$$

$$E_z^{(t)} = \{e_{i,j} \in E^{(t)} \mid v_i, v_j \in c_z^{(\text{all})}\}, \text{ in which } m_z^{(t)} = |E_z^{(t)}| \tag{6}$$

where a isolated sub-network or community $c_z^{(t)}$ from the $G^{(t)}$ snapshot is composed by the nodes in $V^{(t)}$ that belong to the community region $c_z^{(\text{all})}$ (Figure 2(C)). This results with a temporally discrete network, where the temporal unit corresponds to each year observed in the data, and the cross-sectional unit to nodes and communities observed during each year.

D. Three spatio-temporal civil war-network dimensions

Using the method outlined above, we operationalize three dependent variables corresponding to the distinct traits of our civil war event community-based dynamic (temporal) network representation (CoDNet). Each of these variables hence captures a different spatio-temporal feature of civil war activity.

1 **Average intensity** – operationalized as the average degree between civil war cell-years within a given community z during a given year t ($c_z^{(t)}$). This variable captures the average activity level of a given civil war node (cell-year), or the average number of links, within a given community annually – average node degree within a community $\langle k_z^{(t)} \rangle = 2m_z^{(t)}/n_z^{(t)}$.

This indicator captures, in effect, how likely a given civil war is to spread over space and time. A different way of thinking about this measure is as follows: if one were dropped off randomly in a given cell within a given civil war community during year t, what is the average number of cells with active ongoing warfare that exist in one's immediate vicinity. In other words, the average intensity measure tells us how 'volatile' a given cell is within a given community, even if it did not experience active civil war in year t or not.

2 **Dispersion** – operationalized as the number of active cell-month (i.e., that experienced civil war events) during each year within a community as identified by CoDNet $n_z^{(t)} = |c_z^{(t)}|$.

In measuring dispersion this way, we capture the immediate geographic spread of civil war events over space and time within a new or an ongoing civil war community. The dispersion indicator therefore allows us to measure how far a given civil war (community) reaches *based on the underlying structure as detected by CoDNet*. This separates it from other, commonly used indicators of spatial civil war, which measure and distinguish violence using qualitative information included in the underlying reports used construct the dataset, and which accordingly do not account for hierarchical relations between conflict and potential violence 'spillovers.' Instead, researchers often 'allocate' a particular incident to a broader civil war based on specific qualitative guidelines and reports, an approach that can introduce coder and other types of bias. Our dispersion indicator, in contrast, does not rely on qualitative definitions; instead it leverages the underlying topological structure and the geographic characteristics of each event to identify related civil war over time and space.

3 **Intraconnectivity** – operationalized as the number of links between civil war cell-years in each community $c_z^{(t)}$. i.e., the number of edges in the community $m_z^{(t)}$.

Substantively, this indicator allows us to measure how 'active' a civil war-afflicted area is, i.e., how many nearby cell-years to i experience civil war simultaneously. Higher intraconnectivity corresponds to civil wars (communities) that experience more overall fighting *over time and space*, and not only in terms of the number of causalities or civil war events within particular cell-years. Within such communities, civil war persists over time. Importantly, a civil war can involve a small number of events and casualties, and still have a high level of intraconnectivity – the more cells that experience civil war during a given year, and the more connected these cells are, which allows fighting to move and spread across locations more easily, the higher the intraconnectivity of a given civil war community is.

Measured as such, these different spatio-temporal civil war-network indicators allow us to consider multiple time-varying dimensions of geospatial civil war. Accordingly, in the next section we estimate a set of regression models on each of these dependent variables, accounting for a host of potential confounders, and compare these effects to the same confounders' effects on an unmodified (conflict event-count) civil war indicator.

IV. Internal Validation

We report a map where all civil war communities were collapsed over the 1989-2014 period for $K = 3$ therein in Figure 3. We report separate figures for different values of K and d_{max} in the SI file. As Figure 3 illustrates, our CoDNET algorithm indeed captures distinct geospatial clusters of civil war that are in line with historical record, including – for instance – three distinct historical civil war trends in India (Kashmir, Naxalite rebellion, the northeastern secessionist civil wars), the civil wars in Colombia and (in a distinct cluster) Peru, violence in the Horn of Africa, and the civil war in Angola [10, 11, 19, 43].

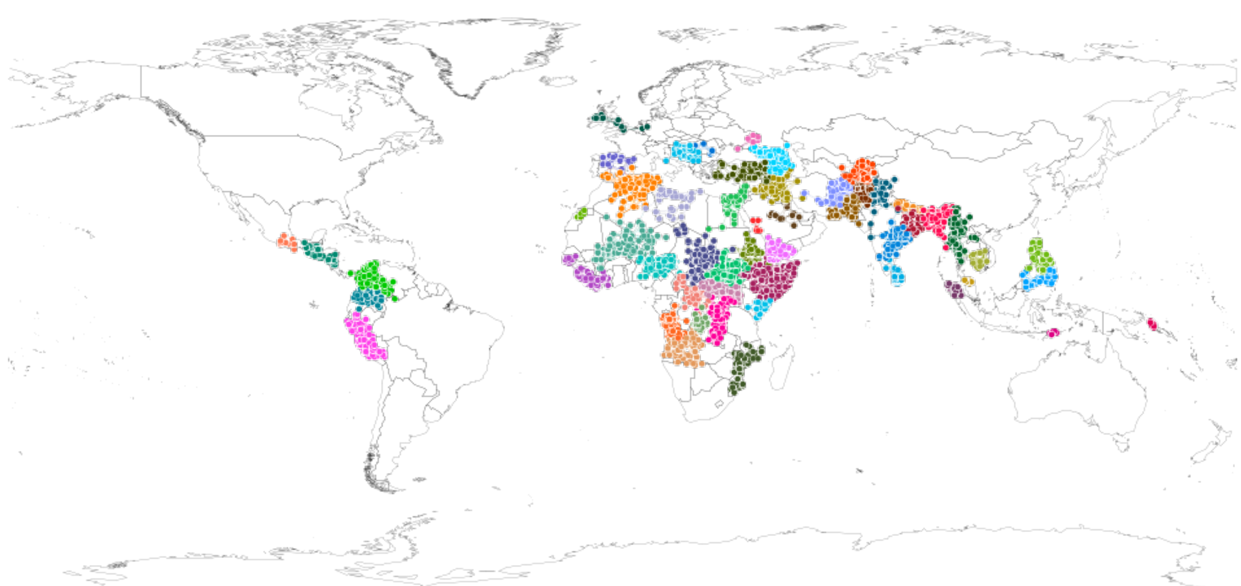


Fig. 3. Global temporal communities of civil wars encompassing 26 years of data between 1989 and 2014 across the entire globe

To verify that the patterns found within our detected communities are valid, we next focus on a particular world region and examine whether our network-based characterization of civil war communities corresponds to – or at least significantly overlaps with – qualitative understandings of different civil war clusters. To this end, we focus on South Asia – specifically India (excluding the northeastern territories) and Sri Lanka – considering that these two countries are host to several ongoing civil wars. We plot the maps corresponding to the evolution of different civil war communities in these regions between 1989 and 2014 in Figure 4(a).

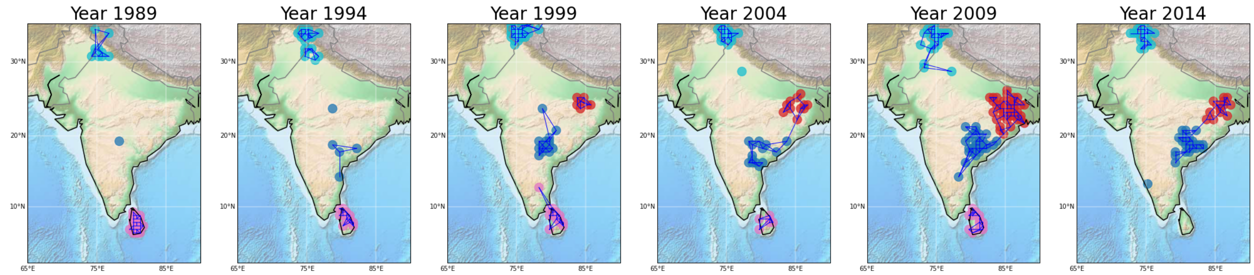
As the different maps in Figure 4(a) illustrate, CoDNeT identifies three distinct civil war clusters in mainland India. The northern cluster (in teal) corresponds to the ongoing civil war in Kashmir, which – as illustrated in these plots – has remained relatively persistent over the entire 1989-2014 period [43]. In the center, our spatio-temporal network algorithm identified two distinct clusters related to the Maoist Naxalite rebellion, which has began in 1980 and has intensified over our period of analysis, particularly during the early 2000s. As Figure A3 in the SI file illustrates, these two spatio-temporal civil war clusters correspond to the two observed clusters of Naxalite activity in the 'Red Corrider,' as well as the intensification of civil war therein during the last few years of the first decade of the 2000s [44]. Finally, it is also important to highlight that CoDNet does not conflate these civil wars – or any ongoing civil wars in the Indian mainland – with nearby distinct civil wars. To this end, the bottom cluster (pink) reflects the Sri Lankan civil war, which CoDNet detects as a completely separate community. The civil war clusters/communities detected by CoDNet therefore directly related to qualitative expectations regarding these wars, illustrating the internal validity of this spatio-temporal network approach.

Having examined the geographic clustering of our CoDNet based communities, we turn to evaluate variations in degree symbol of our three indicators of interest – average intensity, dispersion, and intraconnectivity – over time, and compare them to observed event counts from GED in Figure 4(b). Interestingly, these figures illustrate just how different operationalizations of spatio-temporal civil war dynamics can be from observed counts. For instance, compared with event counts, the average intensity (left column) of Naxalite violence in both clusters of the Red Corridor picks up during the late 1990s, when the Naxalites begin their major offensives [44], even though the number of these attacks does not peak until several years later. In contrast, dispersion and intraconnectivity values (the second and third columns from the left) approximate observed conflict-event counts more closely in this case. Examining the civil war in Sri Lanka (third row), all three spatio-temporal indicators exhibit markedly different trends from the count based measure, which is illustrative of the advantates of dynamic geospatial network modeling, especially considering the relatively small size of the island. Similarly, all three implicit measures illustrate a saw-blade-like pattern in violence in Kashmir (bottom row), whereas the event-count measure record only two peaks, one in 1998-1999 and another one in 2003-2007. Figure 4(b) therefore illustrates the potential advantages of using CoDNet to operationalize spatio-temporal civil war dimensions. We analyze systematically whether this could produce new inferences with respect to civil war determinants in the next section.

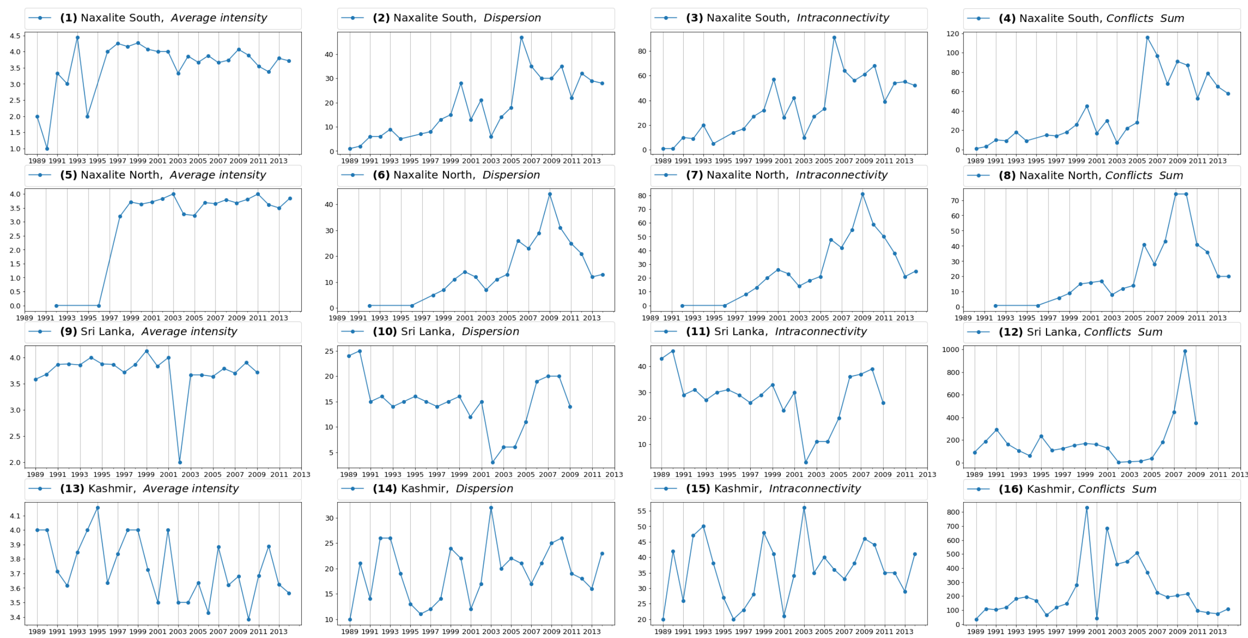
V. External Validation

A. Data, Variables, Methods

This section illustrates the potentially inaccurate research narratives and over/underestimated inferences that not operationalizing the spatially-dependent nature of civil war may induce in the extant research. Here, we specifically compare how the effect of some key determinants of civil war varies across our three CoDNet-based indicators,



(a)



(b)

Fig. 4. Network-based spatio-temporal characterization of civil war clusters in India and Sri Lanka.

and between these variables and the unprocessed event-count variable. To do so, we employ a global grid-cell framework encompassing 21 years of data between 1992 and 2012 across the entire globe, incorporating our three CoDNet-based indicators as well as the preprocessed count indicator therein. Note that some of our models evaluate effects over a shorter period due to the lack of availability of relevant information on some controls.

To evaluate how our three network-based projections compare to a non-preprocessed indicator and whether this can yield biased theoretical inferences regarding civil war determinants, we deploy key explanatory indicators highlighted by past research [e.g., 11, 18, 19, 45]. These indicators operationalize: (i) the local strength of the state (AKA 'state capacity'), often measured using nighttime light emissions [11, 46]; (ii) population density (measured at five-year intervals and interpolated using a last-value-carried-over approach); (iii) precipitation; (iv) temperature; (v) drought severity; and (vi) political exclusion. These variables were all obtained from the PRIO-Grid dataset [20], and are measured at the same cell-year resolution as our civil war indicators. We also include the lag of each respective dependent variable to account for civil war dependencies over time. A detailed description of how each of these independent variables was constructed and summary statistics are reported in the SI file.

To account for any additional confounders at the grid, year, country, and country-year levels, we estimate each model three times. All specifications include grid-cell and year fixed effects to ensure we capture only within- as opposed to across-grid cell variations, as recommended in econometric research. We also report a model that adds fixed effects by country and country $\times$ year to all country-level confounders. Accordingly, our models are specified in a manner that ensures the cell-year level effects of our key variables of interest are effectively identified [47].

Considering the wide range on the count (i.e., non-modified) civil war indicators we treat it as continuous, while also operationalizing each of our CoDNet-based indicators continuously as the range of each of the three relevant dimensions within a given cell-year node. Accordingly, we rely on ordinary least squares (OLS) estimators for analysis, which also ensures observations with zero-only values are omitted. For identification, we rely on the following three specifications (where subscripts i, t, and s refer to grid cell, year, and country, respectively), which are directly in-line with those deployed by past research at conducted similar orders of spatio-temporal resolution [e.g., 48]:

$$\mathbf{y_{it}} = \beta_1 \mathbf{h_{it}} + \beta_2 \mathbf{y_{it-1}} + \omega_\mathbf{i} + \phi_\mathbf{t} + \epsilon_t \tag{7}$$

$$\begin{aligned} \mathbf{y_{it}} = {} & \beta_1 \mathbf{h_{it}} + \beta_2 \mathbf{y_{it-1}} + \beta_3 \ln \mathbf{p_{it}} + \beta_4 \ln \mathbf{r_{it}} + \beta_5 \mathbf{d_{it}} \\ & + \beta_6 \mathbf{c_{it}} + \beta_7 \mathbf{e_{it}} + \omega_\mathbf{i} + \phi_\mathbf{t} + \epsilon_t \end{aligned} \tag{8}$$

$$\begin{aligned} \mathbf{y_{it}} = {} & \beta_1 \mathbf{h_{it}} + \beta_2 \mathbf{y_{it-1}} + \beta_3 \ln \mathbf{p_{it}} + \beta_4 \ln \mathbf{r_{it}} + \beta_5 \mathbf{d_{it}} \\ & + \beta_6 \mathbf{c_{it}} + \beta_7 \mathbf{e_{it}} + \omega_\mathbf{i} + \phi_\mathbf{t} + \psi_\mathbf{s} + \psi_\mathbf{s} \times \phi_\mathbf{t} + \epsilon_t \end{aligned} \tag{9}$$

In these equations, $\mathbf{y_{it}}$ is a vector of denoting the values of each dependent variable in grid cell i for each year. $\mathbf{h_{it}}$ denotes calibrated nighttime light values, $\ln \mathbf{p_{it}}$ denotes the (log) size of the population, $\ln \mathbf{r_{it}}$ is (log) precipitation levels, $\mathbf{d_{it}}$ is drought severity, $\mathbf{c_{it}}$ is temperature, and $\mathbf{e_{it}}$ is ethnic exclusion, each measured in grid cell i during year t. $\omega_\mathbf{i}$, $\phi_\mathbf{t}$, and $\psi_\mathbf{s}$ denote fixed effects by grid cell, year, and country, respectively, while $\psi_\mathbf{s} \times \phi_\mathbf{t}$ is the interaction of country and year fixed effects. ϵ_{it} denotes standard errors clustered by grid cell.

B. Results

Ordinary least squares (OLS) regressions estimated on each of our four dependent variables are reported in Table I. Building on equations 7–9, for each dependent variable, we report three different specifications: (i) a baseline model with only state capacity and grid cell and year fixed effects (Baseline in Table I); (ii) a medium specification, which adds to the baseline model all other cell-year variables mentioned above (Controls in in Table I); and (iii) a country-year specification that adds fixed effects by each country and – separately – country $\times$ year interactions, to account for all country-level features that are either constant or vary by year (e.g., democratization, GDP, population) (CFEs$\times$YFEs in Table I).

Table I illustrates the potential inferential inaccuracies that can arise in this context when the underlying spatio-temporal dimensions of civil war are ignored. Specifically, the coefficient estimates for *Nighttime light*$_{it}$ are negative and statistically-significant (to the $p < .01$ level) in the count-based (i.e., not preprocessed models). This is not a surprising finding, and indeed, it supports what several highly-cited studies [e.g., 18, 19] that find that areas within countries where the state is *weaker* and government reach is lower experience more civil war. A substantive interpretation of *Nighttime light*$_{it}$'s coefficient suggests that a one unit increase in the (log) level of nighttime light emissions is associated with an average reduction of about 0.56-0.83 incidents within a particular cell during a given year.

Moving to the CoDNet-based indicators, however, we find the opposite relationship, namely that the coefficient estimates for *Nighttime light*$_{it}$ are now *positive* and statistically significant (to at least the $p < .1$ level) across all but two of our preprocessed spatio-temporal civil war dimension measures. While some studies do suggest that, at the local level, civil war arises within areas where the state is *stronger* [11], this is nevertheless a counterintuitive and theoretically-relevant finding. Substantively, *Nighttime light*$_{it}$ is now associated with an average *increase* of about 0.02-0.04, 0.83-0.97, and 1.5-1.7 in the degrees of conflict average intensity, dispersion, and intraconnectivity, respectively.

[11] suggest that such a relationship is the result of rebel groups more likely to emerge in areas of state power (where local political elites reside) or because groups seek to stage their first attack in locations where it will especially hurt the government, namely areas with high economic and political importance, which are also more likely to emit nighttime light [46]. By effectively operationalizing these spatial dependencies, we are able to not only verify the later body of research's hypothesis, but also reconcile these divergent results. We show that one reason research failed to reach a consensus on this relationship may be because of variations in the spatial dynamics of civil wars over time. Importantly, the fact that these divergent results are consistent across our three preprocessed dependent variables (although they are not as strong in the Average Intensity models) and across sensitivity analyses accounting for temporal and parameter (k and d_{max}) operationalization choices (Tables A2–A5 in the SI file), suggests that this finding is not the result of specification or control choices, but is rather driven by the choice of accounting-vs-not-accounting for geospatial civil war dependencies over time.

Interestingly, *Nighttime light*$_{it}$ is not the only variable whose effect's magnitude and reliability change depending on whether one relies on count vs. spatio-temporal geospatial civil war indicators. The coefficient estimates for *Drought*$_{it}$, a widely-used indicator in environmental conflict research, vary not only across the preprocessed and non-preprocessed dependent variables, but also across our *Average intensity*$_{it}$ indicator on the one hand, and *Dispersion*$_{it}$ and *Intraconnectivity*$_{it}$ on the other, although this effect is far less statistically robust than *Nighttime light*$_{it}$'s, and hence should be interpreted with caution. Substantively, the effects of drought are most observable in the spatio-temporal variables, where the onset of drought raises the degree of *Average intensity*$_{it}$ and by about 0.02, and reduces the degree of *Dispersion*$_{it}$ and *Intraconnectivity*$_{it}$ by 0.03–0.03 and 0.03-0.4, respectively.

This finding does suggest that operationalizing civil war along different spatio-temporal network dimensions can help in identifying particular contextualized effects of some determinants on civil war. Overall, Table I thus provides strong evidence that taking spatio-temporal civil war trends can improve our understanding of civil war, and help preventing biased inferences. The rest of the coefficient estimates are not robust in terms of sign or signficance, and are hence not discussed.

Additional illustration of the substantive contribution preprocessing geospatial longitudinal event data can make in civil war analysis research, and social science research broadly, are reported in Figure A4, SI file. As illustrated in this figure, once the data is preprocessed, network degrees along our different implicit civil war variables generally exhibit positive covariation (i.e., intensification) with *more* rainfall and *colder* temperatures. Again, this can help in reconciling some expectations in mainstream research [49] with the findings of other studies that highlights the contextualized affect of weather shocks or links research abundance with violence [50, 51].

To ensure the robustness of our findings from Table I, we also report several robustness models in the SI file. Therein, we reestimate the more robust *CFEs×YFEs* specification from Table I and assess the significance and direction of our coefficient of interest therein. These tests, reported in Table A2–A5, include: (i) using cell-months instead of cell-years to created our CoDNet-based indicators (Table A2), and (ii) using different distances to operationalize $d_{\max}$ and (iii) using different values to operationalize k (Tables A3–A5).

TABLE I

DETERMINANTS OF GEOSPATIAL CIVIL WAR, 1992-2012

	Count			Average intensity			Dispersion			Intraconnectivity		
	Baseline	Controls	CFEs×YFEs	Baseline	Controls	CFEs×YFEs	Baseline	Controls	CFEs×YFEs	Baseline	Controls	CFEs×YFEs
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Nighttime light$_{it}$	−0.560**	−0.704**	−0.836**	0.035*	0.023	0.018	0.847**	0.974**	0.826**	1.503**	1.708**	1.496**
	(0.186)	(0.231)	(0.235)	(0.014)	(0.017)	(0.015)	(0.161)	(0.201)	(0.186)	(0.308)	(0.384)	(0.354)
Population$_{it}$[1]	0.078**	0.097**	0.034**	0.007**	0.007**	0.008**	0.266**	0.281**	0.149**	0.489**	0.520**	0.288**
	(0.009)	(0.012)	(0.009)	(0.002)	(0.002)	(0.001)	(0.017)	(0.019)	(0.016)	(0.032)	(0.037)	(0.031)
Precipitation$_{it}$[1]		0.023**	0.018**		0.008**	0.013**		−0.022†	0.090**		−0.028	0.189**
		(0.005)	(0.004)		(0.001)	(0.001)		(0.012)	(0.012)		(0.022)	(0.022)
Temperature$_{it}$		0.003†	0.005**		0.002**	0.001**		−0.012**	−0.004		−0.024**	−0.006
		(0.002)	(0.002)		(0.0003)	(0.0003)		(0.003)	(0.003)		(0.005)	(0.005)
Drought$_{it}$		−0.007	0.007		0.017**	0.015**		−0.294**	−0.031		−0.425**	−0.030
		(0.016)	(0.014)		(0.004)	(0.004)		(0.043)	(0.036)		(0.079)	(0.067)
Ethnic exclusion$_{it}$		−0.021	0.006		0.009**	0.006†		−0.200**	−0.017		−0.366**	−0.038
		(0.020)	(0.023)		(0.003)	(0.003)		(0.027)	(0.030)		(0.052)	(0.057)
DV$_{t-1}$	0.511**	0.498**	0.480**	0.589**	0.592**	0.563**	0.761**	0.754**	0.709**	0.752**	0.744**	0.699**
	(0.040)	(0.040)	(0.042)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Observations	1,306,473	1,040,428	1,040,428	1,306,473	1,040,428	1,040,428	1,306,473	1,040,428	1,040,428	1,306,473	1,040,428	1,040,428
R^2	0.484	0.476	0.519	0.918	0.920	0.942	0.870	0.871	0.917	0.863	0.864	0.911
Adjusted R^2	0.458	0.449	0.492	0.914	0.916	0.939	0.863	0.864	0.912	0.856	0.857	0.906

† indicates $p < .1$; * indicates $p < .05$; ** indicates $p < .01$. [1] Natural log

Variable coefficients are reported with standard errors clustered by grid cell in parentheses. We do not report fixed effects due to space constraints.

VI. CONCLUSION

Network-based methods yielded import insights a range of social phenomena [1, 52], including in political science [53, 54], and more specifically, in civil war analysis [22–25]. Yet, especially in the latter, scant attention has been given to dynamic geospatial operationalizations. Using a new spatio-temporal network-based tool, we introduced a different approach to theorizing and empirically operationalizing civil war, directly incorporating some of its geospatial dynamics. Briefly, our results suggest that (i) effectively operationalizing spatio-temporal dimensions of civil war using networks, and particularly accounting for their variability over time, is important in improving our ability to generate better understandings of civil war; (ii) that a relationship between some determinants of civil war, most explicitly state power, and civil war exists at the local level, but (iii) that this effect is arguably different than expected in the processed spatio-temporal network-based civil war indicators, which have not been analyzed previously; (iv) that hence the reliance on standard casualty- and event-based measures – while useful – is unlikely to provide a complete picture of how different local-level features impact armed conflict; and finally (v) that taking into account several of civil war's implicit spatio-temporal network dimensions, rather than focusing on one specific feature, can reveal different impacts – and hence different understandings – of the same determinants. Methodologically, the results also highlight interesting future paths of research into dynamic and overlapping community behaviors and prediction of links and incoming nodes according to diffusion dynamics [55], not only in conflict analysis, but in other social-scientific areas.

ACKNOWLEDGMENT

Research carried out using the computational resources of the Center for Mathematical Sciences Applied to Industry (CeMEAI) funded by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (grant 2013/07375-0). DAVO acknowledges FAPESP grants 19/26283-5, 18/24260-5, 17/12646-3, and 16/23698-1.

REFERENCES

[1] D. Lazer, A. S. Pentland, L. Adamic, S. Aral, A. L. Barabasi, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann *et al.*, "Life in the network: the coming age of computational social science," *Science*, vol. 323, no. 5915, p. 721, 2009.

[2] L. Collingwood and J. Wilkerson, "Tradeoffs in accuracy and efficiency in supervised learning methods," *Journal of Information Technology & Politics*, vol. 9, no. 3, pp. 298–318, 2012.

[3] J. Gao, Y.-C. Zhang, and T. Zhou, "Computational socioeconomics," *Physics Reports*, vol. 817, pp. 1–104, 2019.

[4] C. A. Hidalgo and R. Hausmann, "The building blocks of economic complexity," *Proceedings of the national academy of sciences*, vol. 106, no. 26, pp. 10 570–10 575, 2009.

[5] C. Raleigh, A. Linke, H. Hegre, and J. Karlsen, "Introducing acled: An armed conflict location and event dataset," *Journal of Peace Research*, vol. 47, no. 5, pp. 651–660, 2010.

[6] R. Sundberg and E. Melander, "Introducing the ucdp georeferenced event dataset," *Journal of Peace Research*, vol. 50, no. 4, pp. 523–532, 2013.

[7] N. P. Gleditsch, P. Wallensteen, M. Eriksson, M. Sollenberg, and H. Strand, "Armed conflict 1946-2001: A new dataset," *Journal of peace research*, vol. 39, no. 5, pp. 615–637, 2002.

[8] S. C. Nemeth, J. A. Mauslein, and C. Stapley, "The primacy of the local: Identifying terrorist hot spots using geographic information systems," *The Journal of Politics*, vol. 76, no. 2, pp. 304–317, 2014.

[9] O. J. Walther, S. M. Radil, D. Russell, and M. Trémolières, "Introducing the spatial conflict dynamics indicator of political violence," *Terrorism and Political Violence*, 2021.

[10] J. O'Loughlin, F. D. Witmer, A. M. Linke, A. Laing, A. Gettelman, and J. Dudhia, "Climate variability and conflict risk in east africa, 1990–2009," *Proceedings of the National Academy of Sciences*, vol. 109, no. 45, pp. 18 344–18 349, 2012.

[11] O. Koren and A. K. Sarbahi, "State capacity, insurgency, and civil war: A disaggregated analysis," *International Studies Quarterly*, vol. 62, no. 2, pp. 274–288, 2018.

[12] K. Petrova, "Natural hazards, internal migration and protests in bangladesh," *Journal of Peace Research*, p. 0022343320973741, 2021.

[13] H. Hegre, C. Bell, M. Colaresi, M. Croicu, F. Hoyles, R. Jansen, M. R. Leis, A. Lindqvist-McGowan, D. Randahl, E. G. Rød *et al.*, "Views2020: Revising and evaluating the views political violence early-warning system," *Journal of peace research*, vol. 58, no. 3, pp. 599–611, 2021.

[14] B. E. Bagozzi and O. Koren, "Using machine learning methods to identify atrocity perpetrators," in *2017 IEEE International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 3042–3051.

[15] L. N. Ferreira, D. A. Vega-Oliveros, M. Cotacallapa, M. F. Cardoso, M. G. Quiles, L. Zhao, and E. E. Macau, "Spatiotemporal data analysis with chronological networks," *Nature Communications*, vol. 11, no. 1, pp. 1–11, 2020.

[16] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-temporal data mining: A survey of problems and methods," *ACM Comput. Surv.*, vol. 51, no. 4, pp. 83:1–83:41, Aug. 2018.

[17] A. Hamdi, K. Shaban, A. Erradi, A. Mohamed, S. K. Rumi, and F. D. Salim, "Spatiotemporal data mining: a survey on challenges and open problems," *Artificial Intelligence Review*, pp. 1–48, 2021.

[18] H. Hegre, T. Ellingsen, S. Gates, and N. P. Gleditsch, "Toward a democratic civil peace? democracy, political change, and civil war, 1816-1992," *American political science review*, pp. 33–48, 2001.

[19] H. Buhaug, S. Gates, and P. Lujala, "Geography, rebel capability, and the duration of civil conflict," *Journal of Conflict Resolution*, vol. 53, no. 4, pp. 544–569, 2009.

[20] Tollefsen, A. Forø, H. Strand, and H. Buhaug, "Prio-grid: A unified spatial data structure," *Journal of Peace Research*, vol. 49, no. 2, pp. 363–374, 2012.

[21] J. Schon and O. Koren, "Introducing afrogrid, a unified framework for environmental conflict research in africa," *Scientific data*, vol. 9, no. 1, pp. 1–11, 2022.

[22] U. Brandes, J. Lerner, and T. A. Snijders, "Networks evolving step by step: Statistical analysis of dyadic event data," in *2009 international conference on advances in Social network analysis and mining*. IEEE, 2009, pp. 200–205.

[23] M. D. König, D. Rohner, M. Thoenig, and F. Zilibotti, "Networks in conflict: Theory and evidence from the great war of africa," *Econometrica*, vol. 85, no. 4, pp. 1093–1132, 2017.

[24] C. Dorff, M. Gallop, and S. Minhas, "Networks of violence: Predicting conflict in nigeria," *The Journal of Politics*, vol. 82, no. 2, pp. 476–493, 2020.

[25] E. K. Gade, M. Gabbay, M. M. Hafez, and Z. Kelly, "Networks of cooperation: Rebel alliances in fragmented civil wars," *Journal of Conflict Resolution*, vol. 63, no. 9, pp. 2071–2097, 2019.

[26] N. B. Weidmann, "A closer look at reporting bias in conflict event data," *American Journal of Political Science*, vol. 60, no. 1, pp. 206–218, 2016.

[27] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, and A.-L. Barabási, "The fundamental advantages of temporal networks." *Science (New York, N.Y.)*, vol. 358, no. 6366, pp. 1042–1046, nov 2017.

[28] R. V. Donner, M. Wiedermann, and J. F. Donges, *Complex Network Techniques for Climatological Data Analysis*. Cambridge University Press, 2017, pp. 159–183.

[29] Y. Zou, R. V. Donner, N. Marwan, J. F. Donges, and J. Kurths, "Complex network approaches to nonlinear time series analysis," *Physics Reports*, vol. 787, pp. 1 – 97, 2019, complex network approaches to nonlinear time series analysis.

[30] M. A. D. C. Santos, D. A. Vega-Oliveros, L. Zhao, and L. Berton, "Classifying el niño-southern oscillation combining network science and machine learning," *IEEE Access*, vol. 8, pp. 55 711–55 723, 2020.

[31] A. Barberán, S. Bates, E. Casamayor, and N. Fierer, "Using network analysis to explore co-occurrence patterns in soil

microbial communities," *ISME J*, vol. 6, pp. 343–351, 04 2014.

[32] R. Wang, W. Liu, and S. Gao, "Hashtags and information virality in networked social movement: Examining hashtag co-occurrence patterns," *Online Information Review*, vol. 40, no. 7, pp. 850–866, 2016.

[33] Y. Charette and A. V. Papachristos, "The network dynamics of co-offending careers," *Social Networks*, vol. 51, pp. 3–13, 2017.

[34] D. Vega-Oliveros, M. Cotacallapa, L. Ferreira, M. Quiles, Z. Liang, E. Macau, and M. F. Cardoso, "From spatio-temporal data to chronological networks: An application to wildfire analysis," in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, ser. SAC'19, New York, NY, USA, 04 2019.

[35] M. Cotacallapa, L. Berton, L. N. Ferreira, M. G. Quiles, L. Zhao, E. E. N. Macau, and D. A. Vega-Oliveros, "Measuring the engagement level in encrypted group conversations by using temporal networks," in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–8.

[36] A. K. Charakopoulos, T. E. Karakasidis, P. N. Papanicolaou, and L. A., "The application of complex network time series analysis in turbulent heated jets," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 24, no. 2, p. 024408, 2014.

[37] L. Berton, A. de Andrade Lopes, and D. A. Vega-Oliveros, "A comparison of graph construction methods for semi-supervised learning," in *2018 International Joint Conference on Neural Networks (IJCNN)*, ser. IJCNN'18. IEEE, July 2018, pp. 1–8.

[38] P. Holme, "Modern temporal network theory: a colloquium," *The European Physical Journal B*, vol. 88, no. 9, p. 234, 2015.

[39] Y. Sun, J. Tang, L. Pan, and J. Li, "Matrix based community evolution events detection in online social networks," in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*. IEEE, 2015, pp. 465–470.

[40] N. Dakiche, F. B.-S. Tayeb, Y. Slimani, and K. Benatchba, "Tracking community evolution in social networks: A survey," *Information Processing & Management*, vol. 56, no. 3, pp. 1084–1102, 2019.

[41] V. A. Traag, L. Waltman, and N. J. van Eck, "From louvain to leiden: guaranteeing well-connected communities," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[42] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P10008, oct 2008.

[43] V. Schofield, *Kashmir in conflict: India, Pakistan and the unending war*. London, UK: Bloomsbury Publishing, 2021.

[44] T. F. Lynch, *India's Naxalite Insurgency: History, Trajectory, and Implications for US-India Security Cooperation on Domestic Counterinsurgency*. National Defense University Press, 2016.

[45] H. Fjelde and L. Hultman, "Weakening the enemy: A disaggregated study of violence against civilians in africa," *Journal of Conflict Resolution*, vol. 58, no. 7, pp. 1230–1257, 2014.

[46] X. Chen and W. D. Nordhaus, "Using luminosity data as a proxy for economic statistics," *Proceedings of the National Academy of Science*, vol. 108, no. 21, pp. 8589–8594, 2011.

[47] J. D. Angrist and J. S. Pischke, *Mostly Harmless Econometrics*. Princeton, NJ: Princeton University Press, 2009.

[48] O. Koren, "Food abundance and violent conflict in africa," *American Journal of Agricultural Economics*, vol. 100, no. 4, pp. 981–1006, 2018.

[49] M. Burke, E. Miguel, S. Satyanath, J. Dykema, and D. Lobell, "Warming increases the risk of war in africa," *PNAS*, vol. 106, pp. 20 670–20 674, 2009.

[50] T. Ide, A. Kristensen, and H. Bartusevičius, "First comes the river, then comes the conflict? a qualitative comparative analysis of flood-related political unrest," *Journal of Peace Research*, vol. 58, no. 1, pp. 83–97, 2021.

[51] O. Koren, B. E. Bagozzi, and T. S. Benson, "Food and water insecurity as causes of social unrest: Evidence from geolocated twitter data," *Journal of Peace Research*, vol. 58, no. 1, pp. 67–82, 2021.

[52] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *science*, vol. 311, no. 5757, pp. 88–90,

2006.

[53] J. H. Fowler and N. A. Christakis, "Cooperative behavior cascades in human social networks," *Proceedings of the National Academy of Sciences*, vol. 107, no. 12, pp. 5334–5338, 2010.

[54] S. J. Cranmer, E. J. Menninga, and P. J. Mucha, "Kantian fractionalization predicts the conflict propensity of the international system," *Proceedings of the National Academy of Sciences*, vol. 112, no. 38, pp. 11 812–11 816, 2015.

[55] D. A. Vega-Oliveros, L. Zhao, A. Rocha, and L. Berton, "Link prediction based on stochastic information diffusion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 8, pp. 3522–3532, 2022.

VII. Biography Section

Didier Vega-Oliveros Artificial Intelligence Postdoctoral Researcher at Recod.ai, Institute of Computing, University of Campinas, Brazil, and research collaborator with the Center for Complex Networks and Systems Research, Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, USA; research interests: complex networks analysis and graph-based machine learning methods.

Ore Koren Assistant Professor of Political Science, Indiana University Bloomington (2018–); research interests: civil war analysis, international security, environmental politics, statistical methodology.