

Structure of ethnic violence in Sudan: a semi-automated network analysis of online news (2003–2010)

Tracy Van Holt · Jeffrey C. Johnson ·
James D. Brinkley · Kathleen M. Carley ·
Janna Caspersen

© Springer Science+Business Media, LLC 2012

Abstract Mining textual sources of data can be used to design studies and test theories at temporal and spatial scales unheard of in the past. This opens up new opportunities for conflict studies and ethnographic research. We conducted a semi-automated network analysis of the 2003–2010 Sudan Tribune online news articles and modeled ethnic-group conflict in Sudan. We tested whether an ethnic group's connections to the environment (livestock, biomes, and other resources) and other ethnic groups was associated with severe conflict and peace terms and whether ethnic-group richness at a given geospatial location was associated with severe conflict. Ethnic groups with more connections to livestock, environmental resources, and those ethnic groups with denser environmental–resource networks in their sphere of influence had more co-occurrences with severe conflict terms. Ethnic groups that had more connections to biomes in their sphere of influence were co-associated with peace terms. Locations that had high ethnic-group richness were associated with a higher frequency of severe conflict terms.

Keywords Conflict · Content analysis · Social networks · Ethnic mapping

T. Van Holt (✉) · J.C. Johnson · J.D. Brinkley · J. Caspersen
Institute for Coastal Science and Policy, East Carolina University, Greenville, NC 27858, USA
e-mail: vanholt@ecu.edu

T. Van Holt · J. Caspersen
Geography Department, East Carolina University, Greenville, NC 27858, USA

J.C. Johnson
Sociology Department, East Carolina University, Greenville, NC 27858, USA

K.M. Carley
Institute for Software Research, School of Computer Science, Carnegie Mellon University,
Pittsburgh, PA 15213, USA

1 Introduction

Online texts give researchers new opportunities to understand the drivers of conflicts since news reports provide rich ethnographies of events. Never before has there been such an occasion for scientists to conduct comparative studies at scales and units of analyses that include every city or ethnic group in a country, region, or globally. Historically conflict researchers have used aggregated, country-level data to explain inter-state conflicts. However, today many conflicts are geospatially restricted civil wars or terrorist attacks that are led by groups that are not tied to a nation state. The Correlates of War (COW) Project (Singer and Small 1972) and the UDCP/PRIO Armed Conflict Dataset are the leading datasets to understand armed conflicts (Gleditsch et al. 2002). The COW requires 1,000 battle deaths and the UDCP/PRIO dataset requires 25; our approach doesn't operationalize conflict by the number of causalities, instead we run content analyses for words associated with conflict and extract word co-occurrences to understand possible drivers of these conflicts.

According to the literature, geographic, environmental and demographic factors appear to play a role in the state of conflict and peace. In particular, historic lines of ethnic group politics and resources are often pointed to as factors effecting the presence or absence of conflict. The conflict literature has debated whether scarce resources (Homer-Dixon 1994, 1999; Ross 2004a; Theisen 2008), abundant resources (Gilmore et al. 2005) or resource distribution (Ross 2004a, 2004b) is related to increased violence, conflict severity, and/or conflict persistence, with empirical research showing mixed results. Buhaug and Gates (2002) analyzed the geography of civil war and found that the presence of natural resources increased the scope of conflict, but they did not find that land cover (forest and mountainous areas) had the same effect. Horowitz' (1985) work outlines ethnic-conflict theories and considers the role of competition over resources and incompatible values; however, Horowitz argues that these theories do not explain why non-elites follow elites. Many ethnic conflict studies are at country-level scale such as Fearon and Laitin (2003), who found that ethnic diversity did not explain civil war outbreaks. In contrast, Sambanis (2001) concluded that ethnic wars had distinct drivers than non-ethnic wars, which included more ethnic heterogeneity; these wars were also politically (not environmentally) motivated wars. Raleigh (2010) argued that the risk of conflict depended on the size and importance of ethnic groups, with small, politically insignificant ethnic groups experiencing more conflicts related to environmental pressures. We analyze data at a finer resolution. Our analysis, which addresses conflict within a single country, may offer the ability to collect data at a smaller spatial resolution as Buhaug and Lujala (2005) and set up a comparative research design where smaller regions (not countries) can be compared.

Our concern is with the factors relating to conflict in Sudan prior to the secession of South Sudan. We model how the structure of the social-ecological system influences conflict and peace and how the geospatial overlap of ethnic groups at a given location is related to severe conflict using automated text and network analysis as described by Carley et al. (2012). We analyze texts drawn from the Sudan Tribune online-newspaper from 2003–2010 using an ethnic group's sphere of influence to test the following hypotheses:

H₁: *The more complex the environmental and socio-political network in which an ethnic group is embedded, the more likely they are involved in severe conflict.*

H₂: *The more associations an ethnic group has with livestock, the more likely they are involved in severe conflict.*

H₃: *The more associations an ethnic group has with biomes, the less likely they are involved in severe conflict and the more likely they are associated with peace.*

We used a whole network, the entire network of severe-conflict by geospatial location, to test the following hypothesis:

H₄: *The more ethnic groups overlap geo-spatially, the more likely there will be severe conflict.*

We focus on ethnic relationships to livestock since many livelihoods in Sudan depend on livestock and the importance of livestock in an individual's livelihood was shown to vary with conflict (Biong Deng 2010). We also consider how an ethnic group's ties to biomes, which are defined as mountains, rivers, and grassland, etc., relate to conflict. Since the resilience literature links increased resilience to higher diversity and redundancy in the system (Folke et al. 2004), we argue that ethnic groups that are tied to multiple biomes or those that dominate across one biome have less conflict (more ability to resist conflict). We also draw on social network concepts and test whether the density of the resource and socio-political networks or the composition of the textual networks (i.e. number of environmental resources or ethnic groups in a network) influence conflict or peace. We predict that a complex environmental and socio-political network may foster more conflict, because there is the possibility that there are more ties to elements within the network associated with conflict.

1.1 Data

We used the data-to-model (D2M) approach as described by Carley et al. (2011a, 2011b) that utilizes Network Text Analysis (Carley 1997; Popping 2000) to extract meta-network information (Carley et al. 2011b; Diesner and Carley 2005). Meta-network analysis is the analysis of individual networks (organization, location, resources, and knowledge) and the integration of these networks. We downloaded all of the electronically available articles from the Sudan Tribune (www.sudantribune.com) from 2003 to 2010. Then we de-duplicated the data and resolved anaphors (Carley et al. 2012). We then coded the text with our thesaurus (content dictionary) that was developed via both an apriori and inductive approach. Words identified apriori included words such as agricultural commodities listed on the Food and Agriculture Organization website and other relevant glossaries. Words identified inductively included high frequency words (appearing > 20 times across all years) within the text that were not classified according to our apriori thesauri. We classified all coded words according to socio-political (knowledge) and environmental (resource) characteristics following the Outline of Cultural Materials categories (Murdock 1983) as a guide. Locations were identified using the National Geospatial Intelligence Agency dataset for Sudan. This dataset provides geospatial coordinates for over 25,000 locations in Sudan. Ethnic-group organizations were classified in terms of ethnic and

linguistic groupings by Richard Lobban, a subject matter expert who has worked in Sudan for over thirty years (Lobban 2010). We then converted the coded texts to meta-networks using AutoMap (Carley et al. 2011a). AutoMap first codes the text according to the thesauri. It then generates a relational matrix (network) according to word co-occurrences. If coded words appeared within seven words of another coded word, then a tie was created between those words. Both one mode (knowledge by knowledge; resource by resource; organization by organization; location by location) and two mode (organization by knowledge; organization by resource; and organization by location) networks were generated. We used only two mode networks in our analyses. ORA (Carley et al. 2011b) offers the flexibility to change and merge meta-network categories and thesauri entries because in one analysis an organization can be considered a resource, while in another analysis, an organization can be a resource. We used ORA software (Carley et al. 2011b) to run node-level and network-level statistics. We unioned the networks for eight years of Sudan Tribune articles for this study.

1.2 Ethnic group level ($N = 42$) peace and conflict analysis

We created and visualized a two-mode (organization by a merged environmental resource and socio-political) meta-network (see Fig. 1). The environmental resources coded for words related to biomes and land cover, agricultural crops, and land use, etc. Socio-political terms were all concepts that were not specifically environmental and included terms such as severe conflict, economy, infrastructure, ideology, etc. The nodes were sized by degree, which is the number of adjacent nodes to a given node, which, in this case, means the number of times that concept was tied to another concept in the Sudan Tribune from 2003–2010. We did not control for a multiple report of the same conflict in one article. This means if the Nuer ethnic group, for example, was associated with the word peace two times in one news article, then the data would show that this link appeared two times, not one time. However, given the

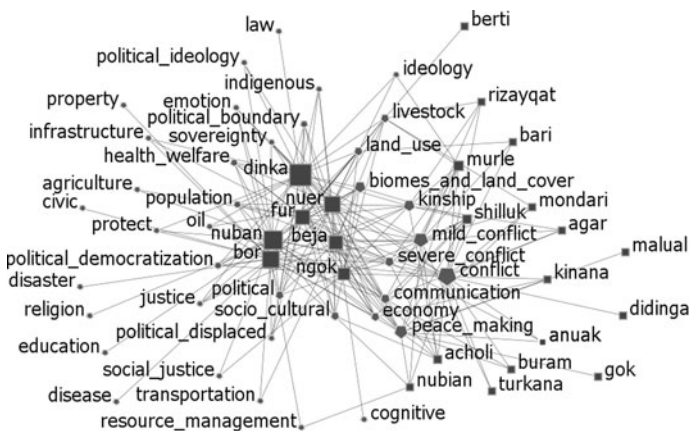


Fig. 1 Meta-network of ethnic groups (square) by merged environmental and socio-political terms (pentagons). The socio-political terms include concepts such as severe conflict, as well as concepts related to politics, ideology, and the economy. Environmental resource terms include land use, livestock, and biomes and land cover. Nodes are sized by degree centrality (the number of adjacent nodes to a given node)

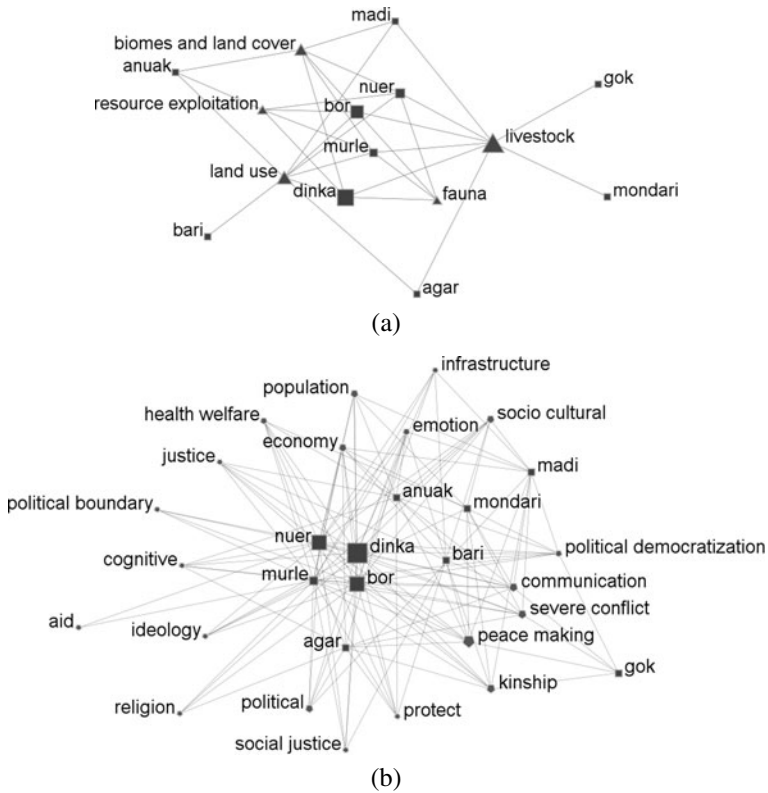


Fig. 2 Environmental resource (a) and socio-political knowledge (b) spheres of influence for the Murle ethnic group. For another ethnic group to appear in the sphere of influence, they must be connected via an environmental resource or socio-political knowledge node

sheer volume of the texts analyzed for a concept to frequently appear, it most likely is picking up on multiple reports of the concept rather than one news article report.

From this meta-network, we created two different spheres of influence for each ethnic group ($N = 46$), which was the unit of analysis. The sphere of influence for an ethnic group, which is conceptually similar to an ego network for an ethnic group, is all of the nodes associated with that ethnic group and the relationship among those nodes. The first sphere of influence was the ethnic-group organization by environmental-resources network (Fig. 2a) and a second sphere of influence was the ethnic-group organization by socio-political knowledge network (Fig. 2b). For an additional ethnic group to appear in the sphere of influence, they had to be tied to the main ethnic group via socio-political or environmental resource terms. For example, the Murle ethnic group environmental-resource sphere of influence network (Fig. 2a) shows that the Murle were tied to the Dinka, Bor, and Nuer ethnic groups via livestock. The socio-political network (Fig. 2b) shows that the Murle had socio-political ties to the Dinka, Neur, and Bor, and, in addition, to the Bari and Madi ethnic groups, and others. All ethnic groups were connected because they co-occurred with severe conflict terms in the text. The ethnic group sphere of influence networks were

used to generate variables for a regression analysis. The dependent variables in the regression analysis, CONFLICT and PEACE, include degree centrality¹ of the (1) severe conflict node and (2) the peace making node (Table 1). Degree centrality for these variables was used because it measures connectivity in ego networks (McCarty 2002), which are similar to spheres of influence networks. A higher degree centrality of severe conflict would indicate that the ethnic group analyzed in that profile had many connections to severe conflict. Independent variables in the regression model include network-level metrics that create a profile of the structure (density) of the environmental (ENV_DENSITY) and socio-political (SOCP_DENSITY) networks of the ethnic group. We also measured the composition of the networks by calculating the number of environmental (ENVNODES) and ethnic group nodes (ETHNODES) in each network to create a richness value. Two node-level metrics include BIOMES and LIVESTOCK. The degree centrality of biomes gives an estimate of the number of distinct habitat types (diversity) that can be accessed by an ethnic group or the dominance (redundancy) that one ethnic group has over a biome, while the degree centrality of the livestock node, estimates how dominant livestock is to an ethnic group's livelihood.

We checked for multi-collinearity (see correlation matrix Table 2), transformed the variables as needed, and ran the regressions with all variables. We eliminated five cases due to lack of sufficient information about these ethnic groups for analysis. We also eliminated the ORGNODE since it was highly inter-correlated with another independent variable.

1.3 Ethnic group richness & conflict frequency by location analysis

To model ethnic-group overlap and severe conflict at a given geographic location we used the geospatial extension of ORA to export a unioned 2003–2010 network of (1) ethnic group by location and (2) severe conflict (a node of the socio-political network) by location. These data were imported into ArcMap (ESRI 2011) and we joined these data by their geographic location and visually displayed ethnic-group richness (number of ethnic groups) and frequency of severe conflict terms associated with each geographic location from 2003–2010. We then tested whether more ethnic groups at a given location was associated with frequency of severe conflict via correlation analysis. We then examined unique cases, that is, those with many overlapping ethnic groups and a lower frequency of severe conflict and those with few ethnic groups overlapping and high frequency of severe conflict co-occurring at the same location.

2 Results

2.1 Ethnic group level ($N = 42$) peace and conflict analysis

An ethnic group that had more ties to CONFLICT within their sphere of influence also had more ties to LIVESTOCK (Table 3; Fig. 3). The Erenga and the Berti ethnic

¹Degree centrality measures were normalized by the maximum possible number of nodes they could be connected to ($N - 1$).

Table 1 Descriptive statistics of the variables used in the analysis. All metrics are calculated on the sphere of influence for each ethnic group

Variables	Description	Level of measurement	Mean	Standard deviation	Range
<i>Dependent variables</i>					
Log of Severe Conflict (CONFLICT)	The log of the normalized degree centrality ^a of words associated with severe conflict in the socio-political ethnic group sphere of influence network	Node	0.174	0.203	0–1
Peace Making (PEACE)	Normalized degree centrality for words associated with the establishment, maintenance or attempts toward peace in the socio-political sphere of influence network	Node	0.335	0.293	0–1
<i>Independent variables</i>					
Biomes and Land Cover (BIOMES)	Normalized degree centrality for words associated with biomes, that is, geographical areas defined by certain ecological patterns, e.g. forests, rivers, desert in the environmental sphere of influence network	Node	0.153	0.268	0–1
Log of Livestock (LIVESTOCK)	The log of the normalized degree centrality of livestock within each ethnic group's sphere of influence. Livestock represents words associated with cattle, goats, herding, etc.	Node	0.143	0.214	0–0.7
Environmental Network Density (ENV_DENSITY)	The average strength of ties across all possible ties for each ethnic group's environmental sphere of influence network	Network	0.633	0.406	0–0.97
Socio-political Network Density (SOCP_DENSITY)	The average strength of ties across all possible ties for each ethnic group's socio-political sphere of influence network	Network	0.892	0.148	0.03–1
Ethnic Nodes (ETHNODES)	Number of ethnic groups present in the ethnic group's socio-political sphere of influence	Network	3.804	3.862	1–18
Environmental Nodes (ENVNODES)	Number of environmental nodes present in the ethnic group's environmental sphere of influence	Network	3.304	3.546	1–12

^aDegree centrality measures were normalized by the maximum possible number of nodes they could be connected to ($N - 1$)

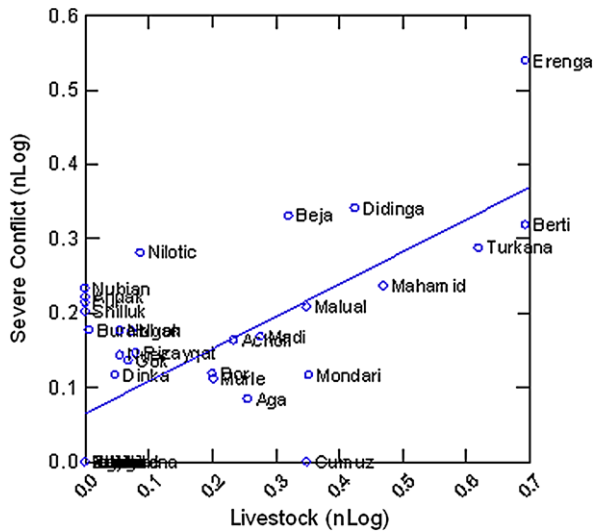
Table 2 Pearson correlation matrix for variables used in regression analysis. Number of observations: 41

	CONFLICT	PEACE	BIOMES	LIVESTOCK	ENV_DENSITY	SOCP_DENSITY	ETHNODES	ENVNODES
CONFLICT	1.000							
PEACE	0.058	1.000						
BIOMES	0.266	0.526	1.000					
LIVESTOCK	0.688	0.126	0.077	1.000				
ENV_DENSITY	0.558	0.200	0.417	0.334	1.000			
SOCP_DENSITY	-0.181	0.100	0.125	-0.047	-0.135	1.000		
ETHNODES	0.130	-0.134	-0.072	-0.072	0.142	-0.876	1.000	
ENVNODES	0.460	0.130	0.296	0.125	0.314	-0.728	0.739	1.000

Table 3 Regression coefficients and *p*-values for the CONFLICT and PEACE regression analyses

Level of metric	Variables	Coefficient for CONFLICT model	<i>p</i> -value	Coefficient for PEACE model	<i>p</i> -value
–	CONSTANT	−0.247	0.070	0.226	0.628
node	BIOMES	−0.045	0.475	0.639	0.006
node	LIVESTOCK	0.345	0.000	0.148	0.498
network	ENV_DENSITY	0.091	0.015	−0.039	0.752
network	SOCP_DENSITY	0.240	0.082	0.044	0.926
network	RESNODES	0.019	0.003	−0.001	0.969

Fig. 3 Scatterplot of CONFLICT [the log of the normalized degree centrality of severe conflict for the sphere of influence for the ethnic group] by LIVESTOCK [the log of the normalized degree centrality of livestock terms as determined by the sphere of influence for each ethnic group (see Fig. 2a for example sphere of influence)]. Ethnic groups labeled



groups, which had many connections to conflict and livestock, are located in known areas of conflict in western Sudan. Another group scoring high on both axes is the Turkana, who are located near the border of Kenya where cattle is critical to people’s livelihoods. The regression model with CONFLICT as the dependent variable was highly significant (adjusted $r^2 = -0.662$; $p = 0.000$). Ethnic groups that had more connections with severe conflict in their sphere of influence (CONFLICT) had more ties to livestock (LIVESTOCK), their sphere of influence had more environmental terms (ENVNODES), and they had more ties to these environmental terms (ENV_DENSITY).

An ethnic group that had more ties to PEACE within their sphere of influence also had more ties to BIOMES (Table 3; Fig. 4). The Sere and the Turkana scored highest on BIOMES and PEACE. The PEACE regression model was also significant (adjusted $r^2 = -0.186$; $p = 0.030$). Ethnic groups that had more connections to peace terms (PEACE) within their sphere of influence had more ties to biome terms (BIOMES) suggesting ecological diversity may have mitigated the potential for conflict. No network-level metrics were significant.

Fig. 4 Scatterplot of PEACE [the log of the normalized degree centrality of peace terms in the sphere of influence for the ethnic group (labeled) by BIOMES [the log of the normalized degree centrality of biome terms as determined by the sphere of influence for each ethnic group (see Fig. 2a for example sphere of influence)]. Ethnic groups labeled

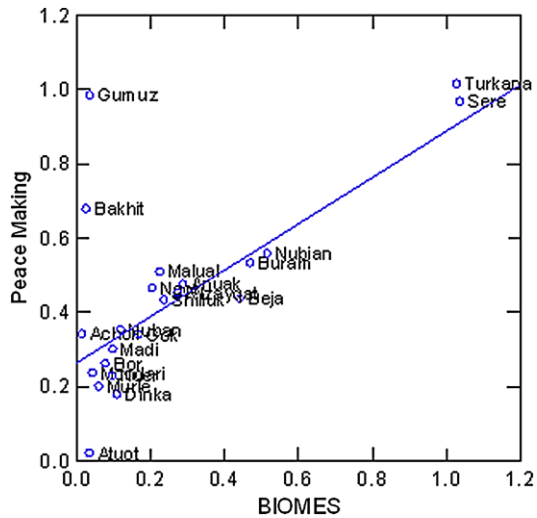
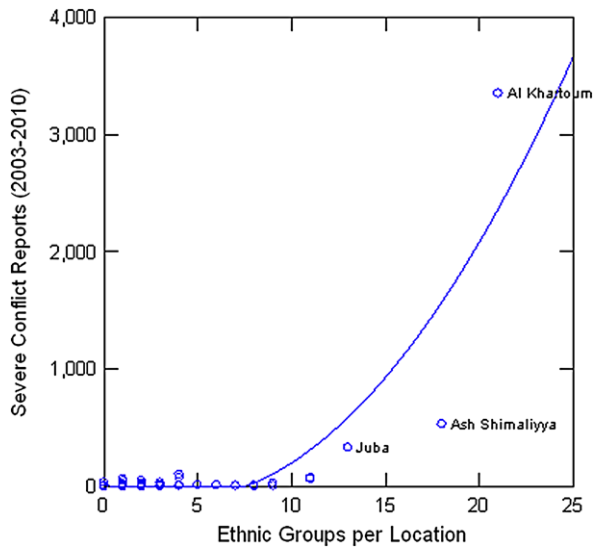


Fig. 5 Scatterplot of the frequency of severe conflict reports by number of ethnic groups at the same geographic location from 2003–2010 as reported in the Sudan Tribune online newspaper. Select locations are labelled



2.2 Ethnic group richness & severe conflict frequency by location

Ethnic-group richness at an x, y location and the co-occurrence of severe conflict at the same location were highly correlated ($r = 0.619$; $p = 0.000$) and show similar findings when they are geospatially visualized (Figs. 5 & 6). The northern Sudan capital of Khartoum was the location with the highest co-occurrence with severe conflict and Khartoum had ties with the most ethnic groups in the text. Other cities that had high conflict and ethnic richness included Ash Shimaliyya, reportedly the most populous city in the north, Jonaqoley, the most populous city in the south, and Juba, the capital of the Southern Sudan after the secession. Some locations, such as Ajok, had high ethnic richness and fewer reports of severe conflict. Figure 6 shows the pos-

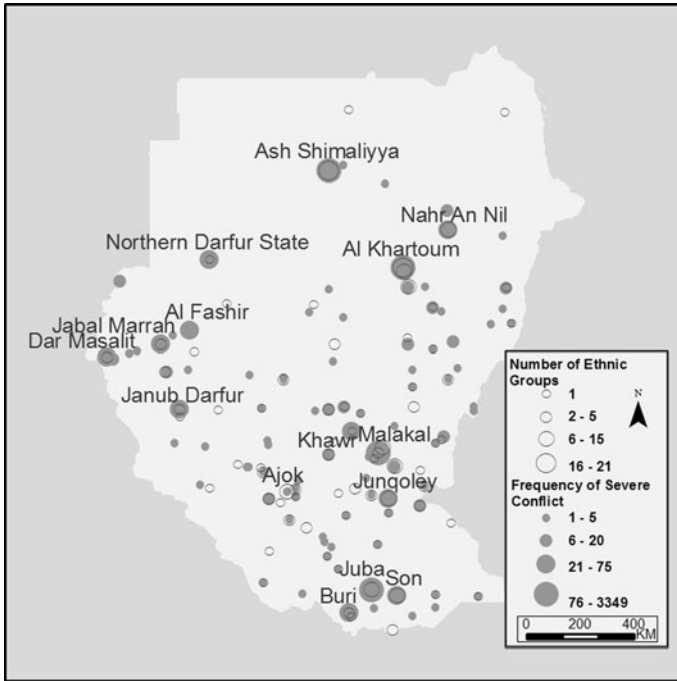


Fig. 6 Geospatial visualization of the ethnic group richness and the number of reports of severe conflicts at a given location as per the Sudan Tribune from 2003–2010. A *sole grey circle* (no outlined circle associated) represents reports severe conflict in a location but no ethnic groups. A *sole outlined circle* (without a grey circle) shows multiple ethnic groups associated with that location but no severe conflict

itive association between severe conflict (grey circles) and number of ethnic groups (outlined circles) by location.

3 Discussion

3.1 Conflict

According to the regression analysis, within an ethnic group’s sphere of influence, greater ties with livestock activities, associations with environmental resources, and denser environmental-resource networks resulted in more ties to severe conflict. This means that it is not just associations with environmental resources or livestock that is important, but the structure of how these resources were tied to other ethnic groups and resources. At this time, we don’t have the data to show whether it is scarcity or abundance of livestock or other environmental resources that determine the effect on space so we cannot determine whether our data support Gilmore’s research on abundance (2005) or whether it is scarcity of the resource itself that is connected to conflict as outlined by Homer-Dixon (1994, 1999). Based on the articles, however, there appears to be more support to the scarce resource hypotheses if we view land or space as the scarce resource. Our work may have the closest link to Ross’ (2004a,

2004b) work on distribution of resources because it may not matter whether resources are abundant or scarce, but rather the relationships among people with each other and with the environment that best predicts conflict.

Textual information confirms some of our findings on livestock and severe conflict. According to the Sudan Tribune (2007), on March 25, a group of Erenga rebels reportedly attacked a group of Arabs and stole 300–400 sheep and, as a consequence, five people died. The Turkana ethnic group shows how overlapping interests with other ethnic groups over space, in addition to livestock livelihoods, may play a role in conflict severity. Some of the Turkana live in the Nadapal Belt region in northern Kenya that borders southern Sudan. The Turkana have been in constant conflict with the Toposa ethnic group of southern Sudan over cattle (McEvoy and Murray 2008). The Toposa people desire the Kenyan lands occupied by the Turkana pastoralists because they are attempting to separate themselves from the Dinka and Nuer ethnic groups, which are more dominant groups (API 2009). According to the Sudan Tribune (Osinde 2005), the Toposa people raided 500 cattle from the Turkana April 29, 2005, leaving at least 14 people dead in the conflict. The Dongiro have raided cattle from the Turkana on April 14th, 2005, stealing 2,600 goats and leaving seven people dead (Osinde 2005). More cattle were reportedly stolen on August 24/25 and 29th, 2008 from the Turkanas and, again, people were killed and abducted (Sudan Tribune 2008). Interestingly, these reports would not make it into standard databases of conflict where larger death thresholds are required for the conflict to be included.

3.2 Peace

Ethnic groups that had salient connections with biomes in their sphere of influence had more reports of peace. However, no other structural features of the sphere of influence network were relevant. Buhaug and Gates (2002) did not find a relationship with conflict and forested areas and mountainous regions and we found biomes, which would include forest and mountains, to be associated with peace in our model. Our metric, however, did not test forests or mountains specifically, but rather either the diversity or redundancy of those resources. If resources are rare, as we suspect, having access to multiple biomes or having a strong connection with one biome, may mitigate conflict. This analysis also possibly supports Folke et al.'s (2004) work that relates resilience to diversity. In this study if we consider peace as a simplistic operationalization of resilience (Cumming et al. 2005), more peace was related to more biomes, which may represent diversity. In our current analysis, however, we cannot tease apart whether the high centrality measures for biomes was because an ethnic group dominated across one biome or occupied multiple biomes. However, this may entail the inclusion of other data sources to more thoroughly answer the question.

The Turkana are also associated with peace and they may be accessing additional biomes to mitigate conflict. Biomes are likely showing up because Turkana are associated with Lake Turkana. According to Yongo et al. (2010), the pastoralist Turkana are taking up fishing in Lake Turkana because of conflicts with cattle and challenges to pastoralist livelihoods because of climate change. While this study focused on Turkana that live close to Ethiopia, it does show that Turkana are tapping new biomes as a means to escape or mitigate conflict.

3.3 Geospatial overlap of ethnic groups

Our findings lean toward supporting the theory that conflict among ethnic groups is tied to competition over resources, which, is a component of Horowitz' (1985) theory on ethnic conflict, though he doesn't find this as an all-encompassing theory to explain ethnic conflict. The salience of social structure and how it is related to natural resources in our analysis may explain a little bit of Horowitz' (1985) concern that ethnic conflict theories don't explain why ethnic groups follow elites. The hidden relationships that can be identified via network analysis can perhaps advance ethnic conflict theory in new ways. While we didn't test ties with specific elites, we could do this with a future analysis.

Ethnic-group richness was associated with more severe conflict. This provides support to Raleigh's (2010) and Sambanis' (2001) findings. Also there are cases where there is high ethnic group overlap and less cases of severe conflict, like in Ajok, Sudan. These locations need to be examined in more detail to understand why this case has high ethnic group richness and low levels of severe conflict. A recent news article by Hamilton (2011) on the conflict in Abyei and spillover effects featured Ajok and the potential for Ajok to support peace. The article (Hamilton 2011) reports that a UN official said "Agok was agriculturally productive. If you draw people here now they will be able to plant by next year. It will avoid getting them into a cycle of dependency on aid." This information supports the notion that associations with biomes mitigates conflict. Future research should compare characteristics of locations with many ethnic groups and less conflict to better understand how landscape features interact with social structure and conflict.

3.4 Limitations and future research

The data obtained in this research would be nearly impossible to obtain in the field. While there are limitations, and the data is only modeling what is reported on in an English language newspaper, we view the interrelationship of environmental issues with people as perhaps one of the reasons behind the debate over resource scarcity and abundance. Maybe it isn't scarcity or abundance, but structure of the relationship among people and the environment that can offer a new understanding of conflict. Likewise, the role of ethnic group relations may be broader and more complex than can be measured at site-specific locations. For example, it could be that in areas with little ethnic group overlap but severe conflict may face overspill effects of violence in nearby areas with higher levels of ethnic group overlap. Still we only have point data on ethnic group location and we are working to develop a more comprehensive map that describes ethnic group distribution over space.

Future research will involve examining networks over time and incorporating relationships with other organizations and actors from an expanded set of textual sources, in addition to refining our thesauri. We are also working to address the valence values of the terms. These tools can be used to select sites for future ethnographic research in the area to understand the patterns reported on in the Sudan Tribune and hopefully will serve to advance our understanding of conflict.

4 Final comments

Conflict in Sudan was examined using advanced text analysis techniques to assess how ethnic group strife is related to environmental and geographic factors using readily available textual sources. We found that resource linkages among ethnic groups were a determinant of severe conflict. Those groups that had either a strong connection to one biome or those that were tied to multiple biomes had more peace. This suggests that competition over the means of livelihood is critical. The finding that overlapping ethnic groups also is tied to conflict also lends supports that conflict may be over competition for resources. As such, interventions that exacerbate the competition are likely to be ineffective. As alluded to in the discussion above, future research should tease apart how abundance or scarcity of resources affects the distribution of environmental resources in a temporal context, possibly through changes and expansions in existing coding schemes.

Working with data at a finer resolution and unit of analysis than the COW and UDCP/PRIO datasets, we offer new perspectives in comparison to other research. What this study shows is that there are currently opportunities to mine textual sources of data to test existing theoretical models in the literature at temporal and spatial scales unheard of in the past. This can open a promising avenue for both model building and testing using only textual and geospatial data for inquiries into a whole range of theoretically interesting problems.

Acknowledgements This work is supported in part by the Office of Naval Research (ONR), United States Navy (ONR MURI N000140811186). We Additional support was provided by the center for Computational Analysis of Social and Organizational Systems (CASOS). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Office of Naval Research, or the U.S. government.

References

- API (African Press International) (2009) Kenya: never ending war between Turkana and Sudanese. African Press International. <http://africanpress.me> October 24, 2009
- Biong Deng L (2010) Livelihood diversification and civil war: Dinka communities in Sudan's civil war. *J East Afr Stud* 4(3):381–399
- Buhaug H, Gates S (2002) The geography of civil war. *J Peace Res* 39(4):417–433
- Buhaug H, Lujala P (2005) Accounting for scale: measuring geography in quantitative studies of civil war. *Polit Geogr* 24(4):399–418
- Carley KM (1997) Network text analysis: the network position of concepts. In: Roberts C (ed) *Text analysis for the social sciences: methods for drawing statistical inferences from texts and transcripts*. Lawrence Erlbaum Associates, Hillsdale, pp 79–100
- Carley KM, Bigrigg MW, Diallo B (2012) Data-to-model: a mixed initiative approach for rapid ethnographic assessment. *Comput Math Organ Theory* (this volume). doi:[10.1007/s10588-012-9125-y](https://doi.org/10.1007/s10588-012-9125-y)
- Carley KM, Columbus D, Bigrigg MW, Kunkel F (2011a) *AutoMap user's guide 2011*. Carnegie Mellon University, School of Computer Science, Institute for Software Research
- Carley KM, Reminga J, Storrick J, Columbus D (2011b) *ORA user's guide 2011*. Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-11-107R
- Cumming GS, Barnes Perz S, Schmink M, Sieving KE, Southworth J, Binford M, Holt RD, Stickler C, Van Holt T (2005) An exploratory framework for the empirical measurement of resilience. *Ecosystems* 8(8):975–987

- Diesner J, Carley KM (2005) Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis. In: Narayanan VK, Armstrong DJ (eds) Causal mapping for information systems and technology research: approaches, advances and illustrations, vol 4. Idea Group Publishing, Harrisburg, pp 81–108
- ESRI (2011) ArcGIS desktop: release 10. Environmental Systems Research Institute, Redlands
- Fearon JD, Laitin DD (2003) Ethnicity, insurgency, and civil war. *Am Polit Sci Rev* 97(1):75–90
- Folke C, Carpenter C, Walker B, Scheffer M, Elmqvist T, Gunderson L, Holling CS (2004) Regime shifts, resilience, and biodiversity in ecosystem management. *Annu Rev Ecol Evol Syst* 35:557–581
- Gilmore E, Gleditsch NP, Lujala P, Rod JK (2005) Conflict diamonds: a new dataset. *Confl Manage Peace Sci* 22(3):257–272
- Gleditsch NP, Wallensteen P, Eriksson M, Sollenberg M, Strand H (2002) Armed conflict 1946–2001: a new dataset. *J Peace Res* 39(5):615–637
- Hamilton R (2011) The politics of aid delivery. Pulitzer Center on Crisis Reporting. June 3. <http://pulitzercenter.org>. Accessed June 21, 2012
- Homer-Dixon T (1994) Environmental scarcities and violent conflict: evidence from cases. *Int Sec* 19(1):5–40
- Homer-Dixon T (1999) Environment, scarcity and violence. Princeton University Press, Princeton
- Horowitz DL (1985) Ethnic groups in conflict. University of California Press, Berkeley
- Lobban RA (2010) Global security Watch-Sudan. Praeger, Santa Barbara
- McCarty C (2002) Structure in personal networks. *J Soc Struct* 3:1
- McEvoy C, Murray R (2008) Gauging fear and insecurity: perspectives on armed violence in Eastern Equatoria and Turkana North. Small Arms Survey, Graduate Institute of International and Development Studies, Geneva
- Murdock GP (1983) Outline of world cultures, 6th edn. Human Relations Area Files, New Haven
- Osinde O (2005) 14 killed during raid by Sudanese tribesmen into northern Kenya. *Sudan Tribune*, April 29, 2005
- Popping R (2000) Computer-assisted text analysis. Thousand Oaks, Sage
- Raleigh C (2010) Political marginalization, climate change, and conflict in African Sahel States. *Int Stud Rev* 12(1):69–86
- Ross ML (2004a) How do natural resources influence civil war? Evidence from thirteen cases. *Int Organ* 58(1):35–67
- Ross ML (2004b) What do we know about natural resources and civil war? *J Peace Res* 41(3):337–356
- Sambanis N (2001) Do ethnic and nonethnic civil wars have the same causes? A theoretical and empirical inquiry (Part 1). *J Confl Resolut* 45(3):259–282
- Singer DJ, Small M (1972) The wages of war. Wiley, New York
- Sudan Tribune (2007) Three killed in intertribal fighting in West Darfur. *Sudan Tribune*, March 25, 2007
- Sudan Tribune (2008) South Sudan apologizes to Kenya over cattle attacks. *Sudan Tribune*, September 9, 2008
- Theisen OM (2008) Blood and soil? Resource scarcity and internal armed conflict revisited. *J Peace Res* 45(6):801–818
- Yongo EO, Abila RO, Lwenya C (2010) Emerging resource use conflicts between Kenyan fishermen, pastoralists and tribesmen of Lake Turkana. *Aquat Ecosyst Health Manag* 13(1):28–34

Tracy Van Holt is an assistant professor in Geography and the Institute for Coastal Science and Policy (ICSP) at East Carolina University (ECU). Her interests include human-environment interactions as they relate to natural resource use, landscape and environmental change, climate change, conflict, and sustainable development. Her research integrates geospatially explicit and remotely sensed data with social and environmental data to understand the social and ecological consequences of environmental change and identify how people adapt to be successful.

Jeffrey C. Johnson is a professor in the Department of Sociology and ICSP at ECU. Her interests include the influence of technological and environmental factors on the organization of work, leisure, and cognition, particularly in groups in extreme and isolated environments. He has focused a major portion of his teaching and research program around the use of social network theories and methods for understanding social structure and organization. His recent substantive interests have focused on the relationship between cognition and social structure. The bulk of his research has focused on these concerns among the maritime peoples of the Pacific basin, especially the insular Central Pacific, the Caribbean, and coastal North

America. Interdisciplinary in both training and orientation, he has had teaching experience in economics, anthropology, sociology, statistics, and Pacific studies.

James D. Brinkley is a Ph.D. student in ISCP at ECU. His interests include coastal hazards, cultural knowledge of maritime communities, coastal and environmental planning, and resource conflict issues. He applies social science and geospatial approaches to study cultural understanding of coastal natural hazards.

Kathleen M. Carley is a professor in the School of Computer Science in the department—Institute for Software Research at Carnegie Mellon University. Her interests include dynamic network analysis, computational, social and organization theory, adaptation and evolution, text mining, and the impact of telecommunication technologies and policy on communication, information diffusion, belief evolution, disease contagion and response within and among groups particularly in disaster or crisis situations. In her research she combines results and approaches from cognitive science, organization science, social networks and computer science to address complex social and organizational problems. She and her center for Computational Analysis of Social and Organizational Systems (CASOS) have developed advanced technologies for network analytics and visualization (ORA), Network extraction from texts (AutoMap), and network evolution and diffusion (Construct and BioWar).

Janna Caspersen is a M.S. student in the Geography Department at ECU. Her research interests include development, bio-geography, and racial inequality.