2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Spatio-temporal networks of social conflicts: analysis and modeling

Gunjan Sehgal
TCS Research
New Delhi, India
sehgal.gunjan@tcs.com

Kiran Sharma
Jawaharlal Nehru University
New Delhi, India
kiran34_sit@jnu.ac.in

Arnab Chatterjee
TCS Research
New Delhi, India
arnab.chatterjee4@tcs.com

Anirban Chakraborti
Jawaharlal Nehru University
New Delhi, India
anirban@jnu.ac.in

***Abstract*—Social interactions can be both positive and negative, and at various spatial and temporal scales. Negative interactions such as conflicts are often influenced by political, economic and social pre-conditions. The signatures of conflicts can be mapped and studied in the form of complex social networks. Using publicly available large digital databases of media records, we construct networks of actors involved in conflicts by aggregating the events over time. We then study the spatio-temporal dynamics and network topology of conflicts, which can provide important insights on the engaging individuals, groups, establishments and sometimes nations, pointing at their long range effect over space and time. Network analyses of the empirical data reveal certain statistical regularities, which can be reproduced using agent based models. The fat tails of actor mentions and network degree distributions indicate dominant roles of the influential actors and groups, which over time, form a part of a giant connected component. Targeted removal of actors may help preventing unruly events of conflicts. Inspired by the empirical findings, we also propose a model for interacting actors that can reproduce the most important features of our datasets.**

***Index Terms*—social networks, conflicts, data analysis, modeling and simulations**

I. Introduction

The quantitative analyses of massive data related to human social conditions have gained much attention in the recent years, due to the important insights gained from multidisciplinary approaches. Merging complex network analysis [1]–[3] with traditional approaches in social sciences, as well as addition of tools and paradigms from several disciplines [4]. Digital data has drawn researchers collaborated across disciplines to scientifically understand complex social phenomena [5], [6] in the recent years, not known previously to the scale of detail as in the present.

Social behavior, social conditions, events and organizations have contributed to shaping the history of human civilization, in a constructive way by building societies and shaping cultures, and sometimes destructive as well – giving rise to conflicts and wars, leading to restructuring countries or nations. Conflicts take place across the world due to a variety of reasons, ranging from socio-economic disparity to political power competitions. As of today, ethnic wars continue to be the most common form of armed conflicts across the globe, but the forces at play that lead a society to ethnic conflicts are not yet well understood. There seems to be a connection between democratization and the occurrence of ethnic conflict. While stable democracies rarely go to war with other democracies, countries that are socio-politically unstable report frequent conflicts between groups with opposing interests [7]. Ethnic conflicts can escalate to human rights violations [8]. Hence, the spatio-temporal studies of conflict formations and the statistical studies of the associated variables are important.

Temporal data for human to human communications and physical contacts/proximity has been studied extensively (see e.g., Holme and Saramäki [9]). We have studied the scale and topology of conflicts, using data from publicly accessible databases, which keep account of events from news available in media. We particularly focus on conflicts in general, as well as *armed conflicts* from two separate databases. The availability of high precision data along with precise spatio-temporal information makes it possible to look for correlations between events, involved actors (individuals, groups, organizations or states) and the geographical pattern of spreading of conflicts, among many other things. Although there have been studies on speculation of ethnic conflicts using census data for segregated population [10], a comprehensive and comparative study of the involved actors in conflicts has been lacking. We present an analysis of the activity of the actors, frequently engaging actor pairs, and the network of actors in general, and provide insights into the static and dynamical aspects of the actor networks in different scenarios of conflicts. The results also guided us to propose a simple model for interacting actors, that can reproduce the main features as reported in our data analyses.

II. Data Description

The Integrated Crisis Early Warning System (ICEWS) Dataverse [11] contains detailed information about various news *events* happening around the world. By filtering out the related subcategories – *protest, assault, fight* and *violence*, using suitable queries, similar to the GDELT Event Database [12], we extracted data related to conflicts. Each event data contains information about the pair of actors involved, a unique time stamp, names of individuals, organizations or groups, the location information of the event, as well as latitude, longitude data of actors and the event. The registered news and hence the data serves as a proxy for the actual event and its intensity (in terms of number of reports). We analyzed data for $1,048,575$ events

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

in the period 1995-2015. Although this database presents each event data with a *source* actor and a *target* actor, we ignore the directionality of the information in our analysis. We also analyzed data from the Armed Conflict Location & Event Data Project (ACLED) [13] for Africa during 1997-2016, reporting a total of $105,459$ events. Each event entry in the latter data set contains information about the number of *fatalities*, because of the nature of the database (armed conflicts).

III. Results

A. The network picture

For both datasets, we construct the aggregate networks over the entire time periods (1995-2015 for ICEWS and 1997-2016 for ACLED). Each actor is visualized as a node and an event is visualized as a link between the actors involved.

Over a period of time T, we construct the network of 'connected' actors in the following way: any pair of actors A_1 and A_2 mentioned together in an event E_1 reported at time $t \in [t_0 : t_0 + T]$ are 'connected' by a link of unit weight, t_0 being the beginning of an interval of time span T. If another event E_2 within the same time window involves actors A_2 and A_3, then A_3 is connected to A_2 with a link of unit weight. Thus A_1 and A_3 are both connected to A_2. Aggregating all such events over the time window T, connected components are formed. The *link weights* w increase if the same pair of actors linking them repeat in multiple events (actor pair mentions). These connected components form a complex network of nodes (actors) and links (actor pair mentions). The number of distinct co-actors is hence the 'degree' k and the total number of actor mentions is the node weight m. A particular actor may be co-mentioned with many other actors and hence have a larger 'degree' k, defined as the number of distinct co-actors it has. A network aggregated over a time period can have several disconnected components or 'clusters', the largest of which is known as the *giant component*.

The ACLED actor network for the entire period consisted of $3,813$ nodes, $105,459$ mentions and $8,621$ unique edges, with the giant component of $3,197$ nodes and $104,909$ edges, while the ICEWS network contains $231,96$ nodes, $1,048,575$ mentions and $101,596$ unique edges, with the giant component having $22,874$ nodes and $1,048,363$ edges. The giant component encompass the majority of actors, more than 83% in case of ACLED while 98% in case of ICEWS data.

The Complementary Cumulative Distribution Function (CCDF) for degree $Q(k)$, node weight $Q(m)$ and link weight $Q(w)$ show broad distributions, resemble either lognormal or stretched exponential distributions (Figure 1), compared to power laws reported earlier for GDELT data [8]. Nevertheless, the above results quantitatively characterize the heterogeneity in the activity of actors, while most actors are relatively less active. The broad distributions for actor mentions indicate that there are a significant few who constantly engage in conflicts, that for actor pair mentions indicate similar characteristic for pairs of actors. The broad degree distributions indicate that the number of actors engaging with very large number of actors are also quite significant.

B. Growth of degrees and nodes

We computed the growth pattern for top 10 actors ranked according to mentions and degrees and found that the cumulative growth rates for mentions and degrees are superlinear functions of the respective arguments (see Figure 2). This implies that the asymptotic growth rates are rather weakly dependent on the arguments in turn. This suggests that in the long run, the giant components will possibly engulf all of the nodes, and what we are observing is possibly a transient, intermediate state of the entire evolution.

C. Network resilience

We also demonstrate how the network breaks down under attack, to investigate the possibility of preventing the spread of unruly events [14]. We perform targeted removal of nodes from the giant component by removing nodes according to descending degree sequence. This leads to rapid destruction or fragmentation of the network. The fraction of nodes G present in the largest surviving cluster decrease very quickly (Figure 3(a),(b)). We observe that this network can be destroyed by targeted attack just by removing 11-15% of the nodes. These numbers are drastically less compared to the case when nodes are removed randomly (random failure) one after another (Figure 3(c),(d)) and requires removing more than 80% of the nodes to dismantle the network. This study indicates that targeted intervention may help stop spreading of conflicts by disconnecting the actors very rapidly.

D. Statistics of fatalities

The ACLED armed conflicts data contained information about the number of fatalities for each event. Here, the probability density function (PDF) of the number of fatalities in a single event (Figure 4) has a broad distribution, and in fact, most of distribution fits to a power law decay with exponent close to 2, higher than the range $1.54 - 1.64$ reported in earlier extensive studies for the statistics of deaths in conflicts and disasters [15]. We also note that the empirical data contains higher frequency of rounded figures approximations – numbers like $10, 20, 50, 100, 150, \ldots$ and 1000 comes with very large frequencies compared to other numbers in their proximity, which is a consequence of rounding off numbers in social data, when exact numbers are not available for reporting.

E. A model for interacting actors

A huge amount of literature deals with modeling complex networks, taking cues from empirical observations. Most of these models [1], [2] try to capture the basic essence of the structure and dynamics of networks, while there exist the specific ones which try to replicate each and every minute aspect of a given dataset. Our empirical findings indicate a strong growth component at the level of actor degree or node weights (mentions) and we can model our systems as growing networks, as has been previously well studied in complex network literature [1], [2], [16]. We exploit the basic *rich gets richer* phenomena in our microscopic model, since it is logical to assume that actors who are already active and involved in

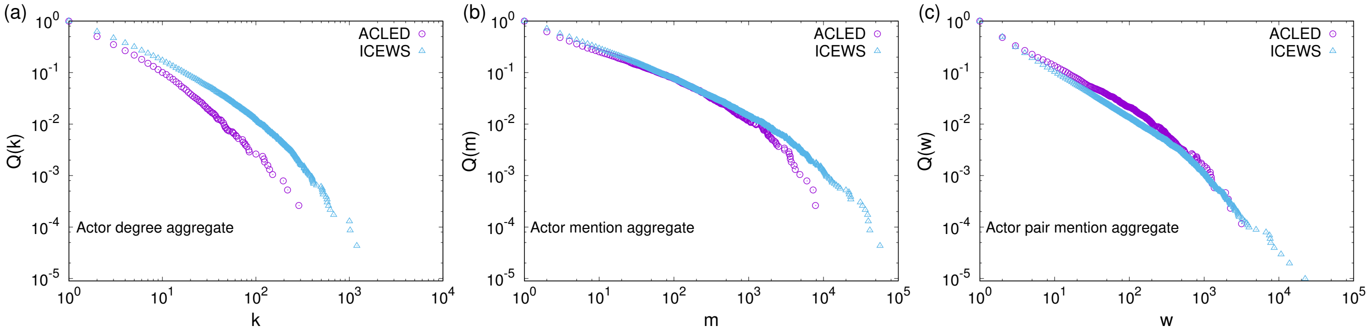

Fig. 1. Statistics for aggregate networks: cumulative probability (CCDF) for (a) $Q(k)$ that an actor is connected to k others or more; (b) $Q(m)$ that an actor is mentioned m times or more; (c) $Q(w)$ that an actor pair is mentioned w times or more. The data is shown for ACLED and ICEWS datasets.

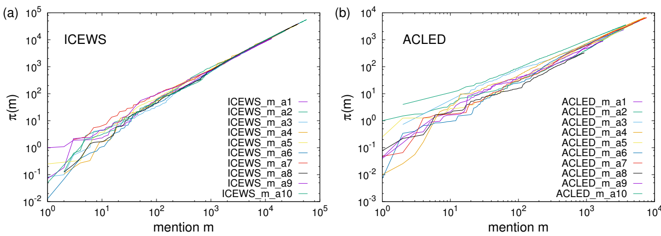

Fig. 2. Cumulative growth rates $\pi(m)$ and for mentions m for ICEWS and ACLED datasets. The curves asymptotically fit to $\pi(m) \sim m^a$ with $a > 1$. The data is shown for the top 10 actors ranked according to degrees and mentions. Similar results were obtained for degrees.

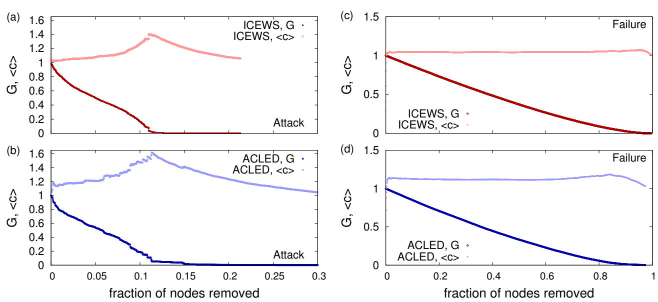

Fig. 3. The structure of the network under attack: The plots show the behavior of the giant component G (fraction of nodes in the largest connected component) and the average number of nodes in the clusters other than the giant component $\langle c \rangle$, with increasing fraction of removed nodes, for (a) ICEWS and (b) ACLED. The structure of the network under random failure: for (c) ICEWS and (d) ACLED networks.

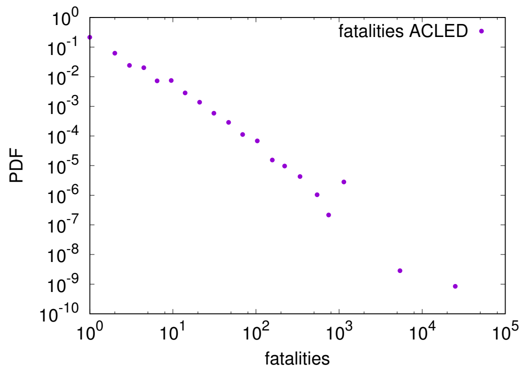

Fig. 4. Probability density function (PDF) of the number of fatalities reported in a single event of armed conflict, computed from the ACLED dataset for Africa (1997-2006).

events with others are likely to attract more events with the same or other actors.

We start with N (even) nodes (actors), who are initially not connected to each other. A connecting link of unit link weight ($m = 1$) corresponds to an interaction event. To ensure that the dynamics do not leave any single actor alone, we assume initially they are pairwise connected. This is achieved by constructing $N/2$ pairs of connected actors. This is done to keep it consistent with the fact that our given empirical dataset comes in the forms of events that involve a pair of actors, which, in the network structure cannot leave an isolated node. Thus, to begin with, each actor has a degree $k_i = 1$ and node weight/mention $m_i = 1$, with $N/2$ pairs having co-mentions/link weights $w_{ij} = 1$, while others zero. Then comes the growing phase of the network, when, at each time step, a pair of nodes $\{i, j\}$ is selected with probability f proportional to (i) k^b or (ii) mentions m^b, and the variables associated with them are updated. i.e., $m_i \to m_i + 1$ and $m_j \to m_j + 1$, $k_i \to k_i + 1$ and $k_j \to k_j + 1$ if $\{i, j\}$ were not connected previously, else k_i, k_j remain unchanged. Link weight also increases by 1, i.e., $w_{ij} \to w_{ij} + 1$. Distributions for degree $P(k)$, mentions $P(m)$ and co-mentions/link weights $P(w)$ were measured after T iterations. In our numerical simulations, we kept $T = 10N$ and system sizes N varied between $10^3 - 10^4$, keeping realistic with our empirical data. The results shown in the plots are produced for 10^2 configuration averages. For $f = f(m)$, power law distributions in the measured quantities m, k and w are observed roughly in the range $1.3 \leq b < 2$, while for $f = f(k)$, power law distributions are observed for $1.6 < b < 2$. Below the lower limit of the range, we find non-power law behavior of the probability distributions, resembling either lognormal, and becoming exponential dominated for even smaller values of b. Beyond $b = 2$, the power law distribution breaks down into two distinct parts – smaller clusters and a very large cluster (akin a *condensate*). In Figure 5, we show the typical cases where the connection probabilities are assumed to be dependent on node weights (mentions) as m^b, with $b = 1.5$. Thus, a minimal model can reproduce the basic features of the network that was constructed by analyzing the event news data involving several actors.

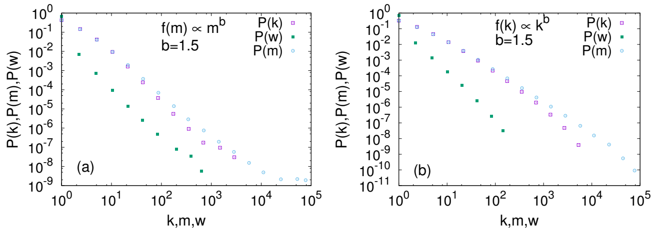

Fig. 5. Probability density functions (PDF) of degree k, node weights (mentions) m and link weights (co-mentions) w for a model networks generated for the case $b = 1.5$ with $N = 10^4$ nodes. (a) For the case of the rate of interaction proportional to node weights m, $f(m) \propto m^b$, the distributions have power law tails with exponents close to 2. (b) For the case of the rate of interaction proportional to node degree k, $f(k) \propto k^b$, the distributions have non-power law tails, resembling stretched exponential or lognormal functions.

IV. Conclusions and Discussions

Reports in news media serve as fair proxy for the importance and intensity of events, from the number of reports and lingering span of time through which the reports follow. Our study uses a time-bound sample of the same and focuses on events that pointed at various events of conflicts, including armed conflicts. The data of events aggregated over time enables us to construct a network of actors, and even finding disconnected groups. The frequencies of mentions of actors (node weight m) and co-mentions (link weight w) with others, as well as involvement with distinct actors (degree k) are good proxies for their importance and influence as well as involvement with other actors. Our study reveals that very large clusters of frequently engaging actors exist while there are also isolated clusters. Identification of influential groups of actors, in terms of their intensities of activities, is important for the purpose of possible intervention by concerned authorities (e.g., government agencies) that may prevent the spread of such unruly events. The probability distributions of degree, node weight and link weights have broad tails for the largest values, indicative of an underlying self-organizing principle behind the events. It would be interesting to address the sociological explanations, implications and policies in the future, although that is a challenging issue. The growth properties of the most influential actor nodes indicate that in the long run, very small fraction of disconnected clusters will be left, while most will get merged to a giant connected component. The data are not only found to be strikingly similar in terms of static and dynamic properties of the network, but also in terms of network stability against failure or targeted attack. This detailed study of network structure, dynamics, function and resilience of the event news data complex network may help policy makers, to point out influential actors and preventing the spread of conflicts. Similar analyses using data from other forms of online social (e.g., Twitter) can also be useful. In the spirit of minimal toy models to reproduce the basic characteristics observed from empirical data, we have also put forward a simple growing network model for interacting actors, where the probability of any actor to get involved is assumed to be proportional to the number of actors it has already interacted with. The numerical simulations reproduce well the various broad distributions found for the probability distributions of degree, node weights and link weights – both power law tails (as reported in GDELT data [8]) and lognormal or stretched exponential characteristics (as reported in this work for ACLED and ICEWS data).

While our study uses standard, well established statistical methods as well as tools of complex networks, the novelty of this work lies in the visualization of the given data in terms of networks and bringing out some important aspects of a network of actors, possibly in conflict. Our multidisciplinary approach to the problem also renders enough insight so as to enable us propose a simple toy mode that can reproduce the basic features of the network. Our work lays a foundation for possible future studies that can bring out details of the mechanisms of when and how the actors get involved, and this might have important social, economic and political implications.

References

[1] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, *The structure and dynamics of networks*. Princeton Univ. Press, Princeton, 2006.

[2] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, pp. 47–97, 2002.

[3] A.-L. Barabási, *Network Science*. Cambridge Univ. Press, Cambridge, 2016.

[4] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. Van Alstyne, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[5] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Rev. Mod. Phys.*, vol. 81, pp. 591–646, May 2009. [Online]. Available: http://link.aps.org/doi/10.1103/RevModPhys.81.591

[6] P. Sen and B. K. Chakrabarti, *Sociophysics: An Introduction*. Oxford University Press, Oxford, 2013.

[7] J. Vorrath, L. F. Krebs, and D. Senn, "Linking ethnic conflict & democratization: An assessment of four troubled regions," *National Center of Competence in Research, Challenges to Democracy in the 21st Century, Working Paper*, no. 6, 2007.

[8] K. Sharma, G. Sehgal, B. Gupta, G. Sharma, A. Chatterjee, A. Chakraborti, and G. Shroff, "A complex network analysis of ethnic conflicts and human rights violations," *Sci. Reports*, vol. 7, no. 1, p. 8283, 2017.

[9] P. Holme and J. Saramäki, Eds., *Temporal networks*. Springer, 2013.

[10] M. Lim, R. Metzler, and Y. Bar-Yam, "Global pattern formation and ethnic/cultural violence," *Science*, vol. 317, no. 5844, pp. 1540–1544, 2007.

[11] "Integrated Crisis Early Warning System (ICEWS) Dataverse, retrieved October, 2017," https://dataverse.harvard.edu/dataverse/icews/.

[12] "The GDELT Project, retrieved October, 2016," www.gdeltproject.org/.

[13] ACLED, *Armed Conflict Location & Event Data Project*, 2017.

[14] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.

[15] A. Chatterjee and B. K. Chakrabarti, "Fat tailed distributions for deaths in conflicts and disasters," *Rep. Adv. Phys. Sc.*, vol. 1, p. 1740007, 2017.

[16] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of networks," *Adv. Phys.*, vol. 51, no. 4, pp. 1079–1187, 2002.