Report on Social Network Analysis of Human Rights Violation Network

Settur Anantha Sai Jithamanyu
Kiran Sharma, Anirban Chakrobarthy, Sayan Gupta

Contents

1 Introduction

Social network analysis has become valuable for understanding complex phenomena, including human rights violations. Human rights violations occur in various contexts and impact individuals and communities globally. Researchers gain insights into these violations' underlying structures, patterns, and dynamics by applying social network analysis to human rights violation data. This approach enhances understanding of the actors involved and contributes to efforts in promoting and protecting human rights.

Advancements in digital technologies and the availability of extensive human rights violation data have created opportunities to apply computational methods and network analysis. This interdisciplinary approach integrates social science, computer science, statistics, and data analytics to analyse the interconnected networks of human rights violations.

In this study, we analyse the degree of distribution of social networks related to human rights violations. By examining the distribution of node degrees, we aim to uncover the underlying connectivity patterns and identify the most central and influential actors within the network. The degree distribution is a key measure of the network's structure and provides valuable information about the prevalence and distribution of connections among actors involved in human rights violations.

Moreover, we employ the backbone concept to extract the network's most significant and robust connections. The backbone represents the core structure of the network, filtering out noise and weaker ties to highlight the most important relationships. By extracting the spine, we aim to identify the key actors and connections that play a critical role in the network's organisation and dynamics regarding human rights violations.

2 Literature review

In recent years, social network analysis has attracted much attention for studying human rights breaches. This multidisciplinary approach examines the intricate dynamics, structures, and actors involved in human rights breaches by combining social science ideas, computer approaches, and network analytic tools. The main conclusions, approaches, and contributions of social network analysis to the investigation of human rights breaches are summarised in this literature survey.

Eck and Madensen (2009)[3] discuss the application of social network analysis to understand criminal networks involved in various illegal activities, including human rights violations. They provide insights into network properties, roles, and relationships among individuals and organisations engaged in criminal behaviour.

Antonopoulos, Papanicolaou, and von Lampe (2018)[2] explore illegal markets' structural and dynamic aspects, including those related to human rights violations. Although they do not explicitly apply social network analysis to human

rights violation data, their book briefly talks about applications of social network analysis on this form of data.
Power (2017)[8] examines the role of social networks in facilitating human rights abuses in North Africa. The study uses social network analysis to investigate the connections between individuals involved in human rights violations and their influence within the broader network. The findings highlight the importance of network structures in perpetrating and sustaining human rights abuses.
Young and Everitt (2018)[10] conduct a systematic review of studies that apply social network analysis to study human rights. They review various methodologies and frameworks that exist currently. Their work reviews and helps identify the gaps and suggests the directions for research in the social network analysis on human rights.
Iyer and Tucker (2016)[4] analyse the structure of networks formed through international human rights co-litigation. By performing social network analysis on the international human rights co-litigation data, they examined the patterns of collaboration and the flow of information among organisations involved in human rights. They highlight the connections and partnerships among various organisations in advancing human rights initiatives.
Sharma et al. (2017) conducted a comprehensive complex network analysis of ethnic conflicts and human rights violations. Their study analysed the distributions of actor mentions, co-actor mentions, and degrees and the dominance of influential actors and groups. They found that the targeted removal of specific actors could effectively mitigate the spread of unruly events. Additionally, they examined the cause-effect relationship between different types of events. The present report closely aligns with Sharma et al.'s study of analysing degree distributions. However, this work diverges by emphasising the extraction of the network's backbone using filtering techniques and identifying the robust subgraph within the overall network.

3 Data Description

Our dataset is a subset of the GDELT database[9]. GDELT Event Database contains a database of news articles from around the world in several languages hosted through Google Cloud. It is possible to extract data for each event, having a unique time stamp, and providing the data about news about worldwide human rights violations spanning a large time scale. Finally, the extracted dataset contains valuable information regarding the actors involved, event locations, and the latitude and longitude data of both actors and events. For our network construction task, we will concentrate only on two columns of the many. We will construct the network based on the pairs of actors involved in the event. The subsequent sections will shed more light on network construction and analysis. This analysis was performed on data acquired in the year 2010. One could easily apply the same methodology for any timeline.

4 Methodology

4.1 Network Construction

As mentioned in the above section, we will use the pair of actors involved in the event to construct the network. We will start by creating the edge list for the network. The actors form the nodes of the network, The edges between the nodes come from the interaction between the actors. The pairs of actors involved in the event taken directly from the data will form the edge list that we want. Closely examining the data tells us that multiple edges repeat, meaning those pairs occur repeatedly. The count of the number of times these pairs occur will become the weight of the network's edges. With this edge list data, we construct a weighted network. We use the NetworkX[7] package to build the network. From the above image, both degree and weighted degree are

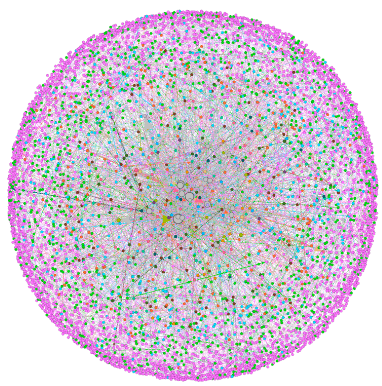
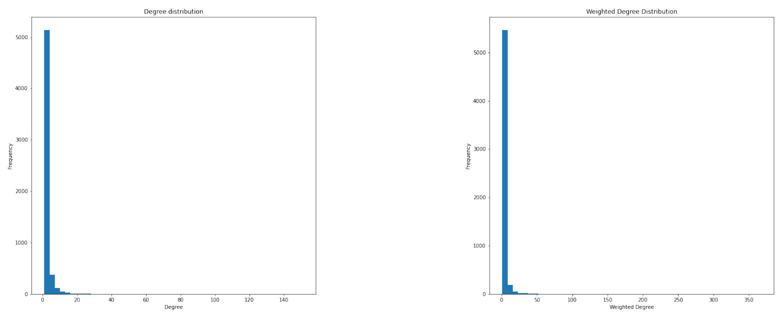


Figure 1: The constructed network with each node coloured according to their degree and sized according to their weighted degree

concentrated at some values; hence, we see many nodes in similar colour and size. A more detailed analysis of their degree distribution will be done in the latter subsections. But for now, the degree and weighted degree values are concentrated at some values.

4.2 Degree and Weighted Degree distributions

The degree of a node is the number of connections it has to other nodes in the network. It measures how connected it is to the different parts of the network. The weighted degree of the network is the weighted sum of edges connected to the node. This measures how strongly it is related to the network. Since we have already constructed a network using the networkx package, we can use their inbuilt commands to obtain both their weighted degree and the degree of each node. After receiving these to get the distribution, we take the count of these and plot a histogram. A better understanding of the distribution can be



(a) Plot of Degree Distribution (b) Plot of Weighted Degree Distribution

Figure 2: Histograms with linear bins

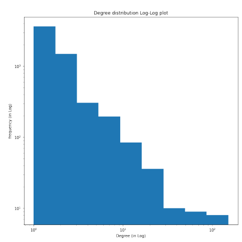
obtained when we plot the log-log plot of the count and degree of the network. A log-log plot will help us uncover the power law characteristics of the data.

4.3 Power Law characteristics

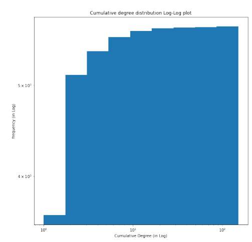
As mentioned, we do a log-log plot of the degree frequency and degree to understand the characteristics of the power law. If this is linear, we can conclude that the degree distribution follows a power law distribution. Now the question is, how do we make a log-log plot? We convert the bins into log scale and the frequency into log scale to create a log-log plot. This can be done directly with the help of the numpy package. Once we inspect the log-log plots, we will understand the power law characteristics of degree. We do the same to the weighted degree distribution too.

4.4 Extracting the Power Law Distribution

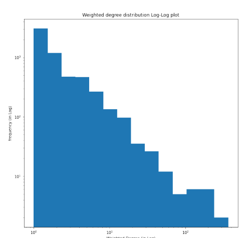
Variables that follow power law distribution are assumed to obey the law: $p(x) \propto x^{-\alpha}$. From the log-log plots3, we can see that both degree and weighted degree assume a linear log-log plot and imply that they follow a power law behaviour. We need to extract the values of α and a constant to convert this into equality. To estimate the parameters, we will first write them in log terms. This will give us $log(p(x)) = -\alpha log(x) + k$. One obvious way to estimate these parameters is through the least squares method. But this method can produce substantially inaccurate estimates of parameters[1]. Hence, we move to maximum likelihood estimates to estimate these parameters. The computation of such estimates allows us to get better estimates of the power law distribution and was found to be more accurate. We can do that in Python directly with the help of the Powerlaw package[5]. The results of this estimation will be discussed further in the results section.



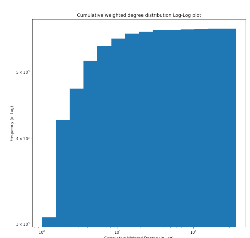
(a) Plot of Degree Distribution in log-log scale



(b) Plot of cumulative Degree Distribution in log-log scale



(c) Plot of Weighted Degree Distribution in log-log scale



(d) Plot of Cumulative Weighted Degree Distribution in log-log scale

Figure 3: Log-Log plots of Distributions

4.5 Extracting Backbone of the Network

A network backbone refers to a sub-graph or a simplified representation of a complex network that captures the most significant and robust features across multiple scales. The backbone extraction involved retaining the most relevant edges from the network. This extraction considers the degree of nodes the edges are attached to and the normalised weights of the edge. This report employs the disparity filter [6] to assign significance values to each edge. This computation of the significance value is based on preserving the heterogeneity of the structure and identifying the robust weights or edges that will hold the maximal strength of the node and also preserve the heterogeneity. After computing the values for each edge, we apply a threshold value alpha depending on our observations and remove all those edges below the threshold.

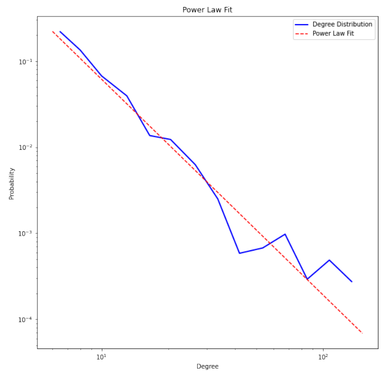
5 Results

5.1 Power Law distribution

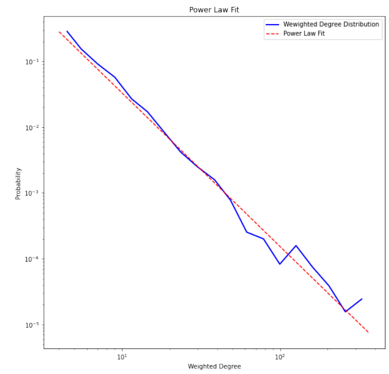
From the figure3, it is clear that degree and weighted degree both follow a power law distribution. And now, using the power law package mentioned in the above methodology section, we extract the values of the α and x_{min}, which is a constant factor. The figure4a gives us the plot of degree distribution and the fitted power law distribution in log-log scale. From the maximum likelihood estimation process, estimated power law exponent: $\alpha = 2.5043079084479034$, estimated lower bound $x_{min} = 6.0$ for degree distribution. The figure4b gives us the plot of weighted degree distribution and the fitted power law distribution in log-log scale. From the maximum likelihood estimation process, the estimated power law exponent is $\alpha = 2.33631326505899$, and the estimated lower bound $x_{min} = 4.0$ for weighted degree distribution. The exponent values of both these do not differ much, but the lower bound values show a difference of 2. From this analysis, it can be seen that the weighted degree obeys power lay distribution more rigidly than the degree of a node.

5.2 Backbone of the network

As mentioned in the methodology part, we use the disparity filter to extract the network's backbone, which assigns significance to each edge of the network. Then thresholding will result in the removal of edges. This leaves us with the required sub-graph or the backbone of the network. Choosing the threshold value is answered with the help of a plot between the threshold and the ratio of edges, nodes and weights in the subgraph obtained to the number in the original network. We want to extract that part of a robust network, meaning the edges we extract should not be because of random fluctuations or noise. From the figure5, we can see that more than 20% of the weights and edges, 40% of nodes disappear for a threshold of 0.01. This implies that these weights, nodes and edges were not robust and occurred only due to some random fluctuations.



(a) Degree distribution (log-log)



(b) Weighted Degree distribution (log-log)

Figure 4: Power law Distribution plots

From 0.01 to 0.07, we observe no significant change in the values, and then from 0.07 to 0.08, we observe a small jump in value. Then from 0.08 to 0.1, the values don't change much at all. As we cross the value of 0.1, the values start decreasing faster. From this, we can conclude that a threshold value between 0.02 and 0.1 can be chosen as a threshold value without the loss of much information or nodes. To retain as much as possible and still remove the random fluctuations, we will choose a value between 0.01 and 0.07. For our case, we shall choose 0.04 midway between both values. In the figure6, we can see the network's backbone with nodes coloured as per the degree and sized as per their weighted degree. The backbone of the network has 3090 nodes and 4483 edges. The average clustering coefficient of the backbone of the graph is 0.05212718507607501. This value is almost twice that of the original network. The resultant backbone of the network is more interconnected or clustered than the original network we were given.

6 Conclusions

This report analysed the degree and weighted degree distribution of a complex network associated with human rights violations data. We have observed how both these follow a power law distribution. This could help us argue that the human rights violations network is scale-free. Removal of nodes at random will not guarantee the network's failure or collapse due to the network's high connectivity with some nodes. We also extracted the network's backbone using the disparity filter concept and computed its average clustering coefficient, almost twice the original network's value.

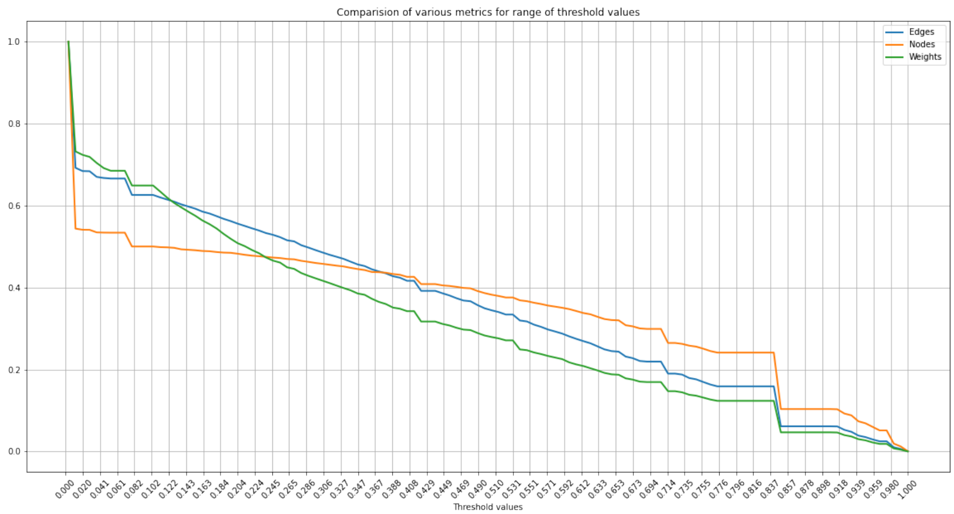


Figure 5: Ratio of Edges, Nodes, Sum of Weights in filtered network to original network plotted against threshold values

7 References

[1] M. E. J. Newman Aaron Clauset Cosma Rohilla Shalizi. "Power-law distributions in empirical data". In: *arXiv:0706.1062 [physics.data-an]* (2007).

[2] Georgios A Antonopoulos, Georgios Papanicolaou, and Klaus von Lampe. *The architecture of illegal markets: Towards an economic sociology of illicit markets*. Vol. 22. Oxford University Press, 2018.

[3] John E Eck and Tamara D Madensen. "Analyzing criminal networks". In: *Handbook of qualitative research methods in criminology and criminal justice*. Routledge, 2009, pp. 303–326.

[4] Arpita Iyer and Joshua A Tucker. "The structure of international human rights co-litigation". In: *Journal of Legal Studies* 45.2 (2016), pp. 427–461.

[5] Dietmar Plenz Jeff Alstott Ed Bullmore. "Powerlaw: a Python package for analysis of heavy-tailed distributions". In: *arXiv:1305.0215 [physics.data-an]* (2013).

[6] Marian Bogun´a M. Angeles Serrano and Alessandro Vespignani. "Extracting the multiscale backbone of complex weighted networks". In: *arXiv:0904.2389 [physics.data-an]* (2009).

[7] "NetworkX package". In: (2023).

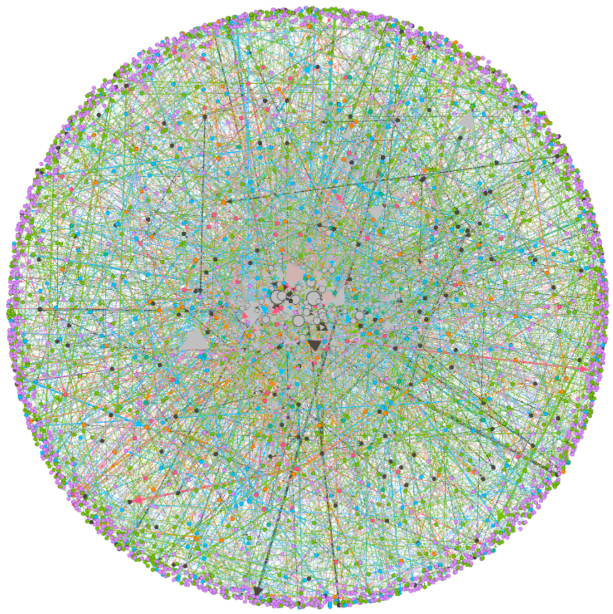


Figure 6: backbone of the network

[8] Eleanor A Power. "Social networks and human rights abuses: Evidence from North Africa". In: *American Journal of Political Science* 61.3 (2017), pp. 678–691.

[9] "The GDELT Project". In: (2010).

[10] Joseph K Young and Joanna C Everitt. "The social network approach to the study of human rights: A systematic review and future directions". In: *Journal of Human Rights Practice* 10.3 (2018), pp. 313–341.

Originality report

COURSE NAME
ID5090 Jan 2023

STUDENT NAME
Settur Anantha Sai Jithamanyu ed20b058

FILE NAME
Settur Anantha Sai Jithamanyu ed20b058 - End Term Project

REPORT CREATED
May 21, 2023

Summary

Flagged passages	3	1%
Cited/quoted passages	4	1%
Web matches		
nih.gov	2	1%
arxiv.org	2	0.7%
wikipedia.org	1	0.4%
lse.ac.uk	1	0.4%
swarma.org	1	0.3%

1 of 7 passages

Student passage CITED

…advancing human rights initiatives. Sharma et al. (2017) conducted **a** comprehensive **complex network analysis of ethnic conflicts and human rights violations**

Top web match

A complex network analysis of ethnic conflicts and human rights violations A complex network analysis of ethnic conflicts and human rights violations

A complex network analysis of ethnic conflicts and human rights
… https://pubmed.ncbi.nlm.nih.gov/28811628/

2 of 7 passages

Student passage FLAGGED

…conflicts and human rights violations. Their study analysed the **distributions of actor mentions, co-actor mentions, and degrees and** the **dominance of influential actors and groups**