Inferring Social Network Structure from Incident Size Distribution in Iraq

Timothy R. Gulden
Policy Researcher
RAND Corporation, Santa Monica, CA
tgulden@rand.org

Abstract

This paper reviews conflict micro-data from several civil conflicts including those in Kenya following the 2007 elections, Guatemala between 1977 and 1886, and Iraq between 2003 and 2009. It presents evidence for a consistent power-law regularity in incident sizes across these conflicts, including variations reflecting particular aspects of the conflicts in Guatemala and Iraq. In particular, it finds that the distribution for Iraq is strongly truncated with respect to large events, though otherwise appearing much like the other conflicts. It proceeds to develop a simple model derived from the distribution of spans of control in hierarchal social networks that can explain the consistent power-law exponent across conflicts as well as accounting for the truncated distribution given the stylized facts of the war in Iraq, particularly in its earlier years.

Introduction

Over the past fifteen years, a body of evidence has begun to emerge that points toward a general regularity in the size distributions of violent incidents in civil wars and conflicts, namely that they tend to follow a power-law distribution that is characterized by an exponent that is slightly greater, in absolute value, than negative one [Richardson 1948; Gulden, 2002; Cederman, 2003; Clauset & Wiegel, 2009; Bohorquez et al., 2009; Bhavnani et al., 2011; Kalyvas, 2012]. Several theories exist as to the mechanism driving this regularity and a proper understanding of these wide-spread, if not universal, conflict dynamics is essential to the development of useful conflict models. This paper seeks to further the discussion of the power-law regularity in conflict incident data by examining a data set from the Iraq war during the first decade of the 21st century, identifying a variation on this pattern in those data, and proposing a simple model to explain both the regularity and the observed variation.

The core insight from this model is that a self-organized critical system that is organized hierarchically should be expected to produce precisely this distribution of violent incidents and that such a model makes it simple to square the departures from the regularity that we observe in Iraq with the stylized facts of that conflict.

Detailed and reasonably complete information on violent incidents in civil wars is notoriously hard to come by for a host of reasons. One or several parties to a given conflict my have in interest in concealing aspects of the conflict, particularly if (as is often the case) the conflict involves aspects that are contrary to international or local norms and laws. Even when societies have decided to attempt a full accounting of violence that took place during a war – as part of a post-conflict truth and reconciliation program, for example – such data are often extremely difficult to compile because of the chaos and information loss that is inherent in times of war: witnesses have been killed or displaced, records have been lost, rumors and accusations have taken on the stature of facts, etc.

Efforts have, however, been made to collect such data and we will briefly examine two other cases, namely Kenya following its elections in 2007 and Guatemala between 1977 and 1986, to establish the pattern before moving on to examine the modified version of the pattern that emerges in Iraq between 2003 and 2009.

Methods

The basic power-law regularity can be seen in two related datasets on violent incidents related to widespread violence following Kenya's 2007 national elections. Teams working under the direction of Dr. Kevin Jones developed these two datasets as part of the ARGUS project at Georgetown University in 2008 [Jones, 2008]. They are summarized in figure 1.

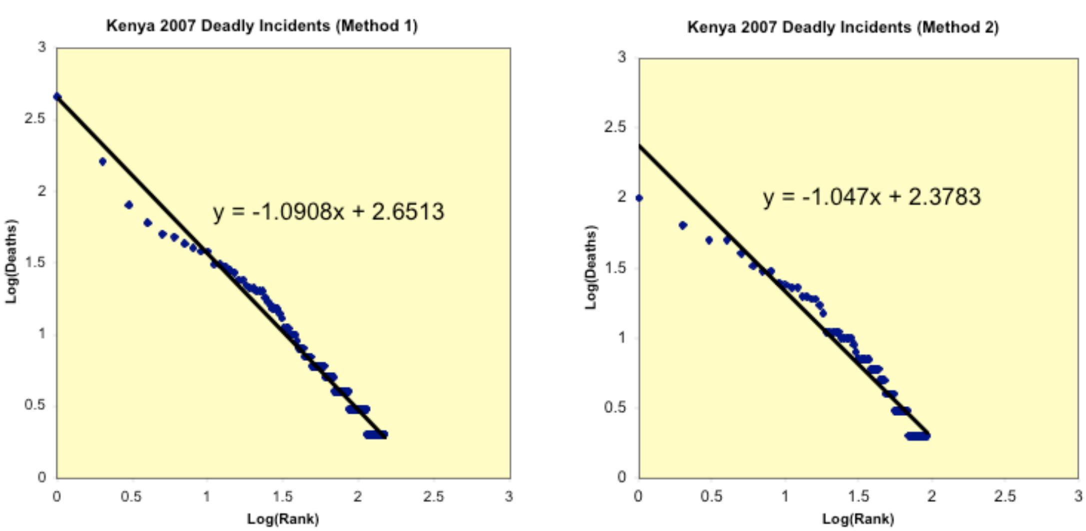

Figure 1. Rank-Size plots of violent incidents following the 2007 elections in Kenya. Both X and Y axes are in logarithmic coordinates. Incidents on the left plot (method 1) were compiled via expert review of heterogeneous sources. Incidents in right plot (method 2) were compiled via machine assisted non-expert review of local and international press sources.

These data were compiled as part of a validation effort for a machine-assisted news searching system designed to monitor the development of civil conflict situations using an approach that was originally developed in a public health context for monitoring the emergence of new diseases. Incidents on the left plot of figure 1 (method 1) were complied via traditional expert reviewing various local sources. Incidents in right plot (method 2) were compiled via the machine assisted non-expert review of on-line press sources.

Though they were conducted independently, both approaches produced a similar pattern of events. The rank-size plots display incidents in rank order, with the largest incident on the left (with between 100 and 500 people killed) and incidents arranged in order of decreasing number of deaths moving rightward to incidents of size 1. When plotted on log-log axes, both distributions can be reasonably approximated with a straight line. The slope of this line y=mx+b is equivalent to the exponent on a power-law regression of the form $S = \alpha R^m$ where size is proportional to rank raised to an exponent of the same value [cite]. In both datasets, the best-fit power law exponent has a magnitude that is just greater than unity, -1.09 in the case of the first method, -1.05 in the case of the second.

We note a similar pattern in data from the Guatemalan civil war as observed in incidents spanning between 1977 and 1986. This pattern is analyzed more thoroughly in Gulden [2002], but it will be useful to reiterate the main argument here.

This analysis uses the AAAS/CIIDH database of human rights violations [Ball, 1999] that was assembled by a coalition of popular movements in Guatemala as part of that nations peace and reconciliation efforts in the late 1990s with assistance and direction from Dr. Patrick Ball at the Human Rights Data Analysis Group (HRDAG) at the American Association for the Advancement of Science (AAAS). The records used for this analysis document over 24,000 killings and disappearances that occurred between 1977 and 1986 and for which the date is known within the nearest month.

These data can be partitioned into two subsets: on the one hand the majority of the conflict, which had the character of an insurgency and counterinsurgency, and on the other hand a period of two years in the north western highlands that included a documented campaign of genocide coordinated by the government against a largely indigenous population. In figure 2, the data are shown in a pair of rank-size plots in log-log coordinates.

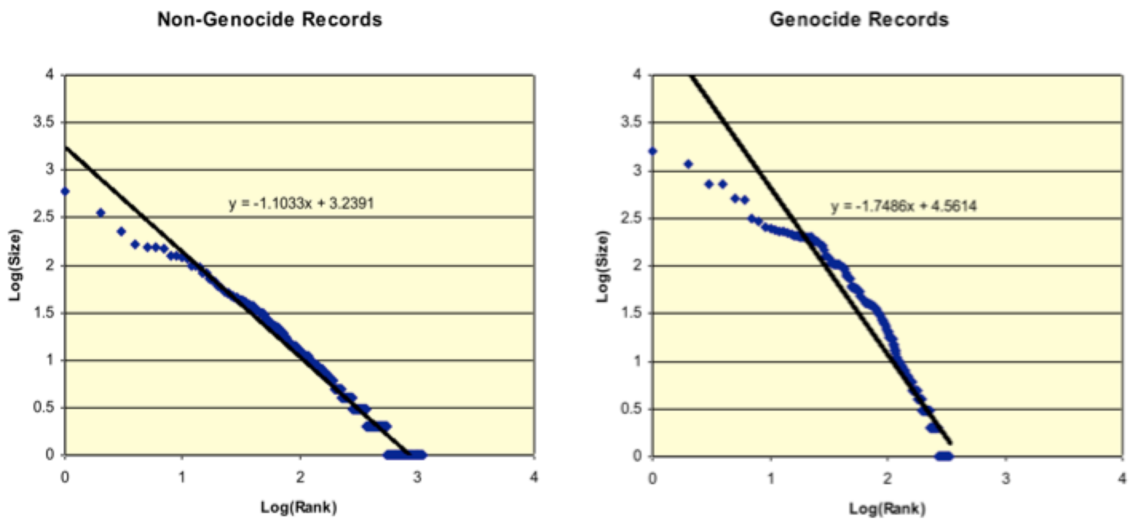

Figure 2. Rank-size plots of violence for municipality/months in Guatemala. Left panel shows records from the mostly non-genocidal, counterinsurgency part of the conflict. Right panel shows records from the spatio-temporal sub-part of the conflict in which genocide was documented to have taken place.

Note that the counterinsurgency part of the conflict is well approximated by a straight line in log-log coordinates while the genocidal part of the conflict is much more curved – indicating that it is not well described by a power-law. Possible reasons for this difference will be discussed below.

The power-law exponent of -1.1 observed in the counterinsurgency part of the conflict is a bit larger than that observed in Kenya, above, but the real exponent is probably closer to -1. This is because the data being plotted are aggregated at the municipality/month level (the highest spatio-temporal resolution that the data will support). See Gulden [2002] for details of how this estimation was carried out.

We can conduct a similar investigation of the war in Iraq using the Iraq Body Count (IBC) database [Iraq Body Count, 2011]. This database used ongoing surveillance and analysis of press sources to document killings in Iraq, requiring at least two reports before an incident is considered well enough documented to include in the database. There is considerable debate as to the completeness of this survey, with its coverage being estimated at anywhere between 10% (Burnham et al., 2006) and upward of 50% (Iraq Body Count, 2011). IBC's limitation in coverage stems primarily from the absence of reporters from many of the most violent parts of the country at precisely the times that the violence in these areas was at its peak. It seems likely, however, that IBC's coverage is reasonably good at least for areas from which reporting was possible.

We will analyze a subset of the IBC data spanning from 2002, when the war began, through the end of 2009.

It is worth noting in this section that the definition of a 'violent incident' is not an entirely straightforward matter. A short engagement between two armed parties might span the midnight hour and thus take place in two different days. A large battle might include many smaller skirmishes as part of it – creating murkiness as to whether the battle or the skirmish is the relevant incident. In most cases, including these two, most reasonable people would have more trouble defining an incident in the abstract than they would have identifying an actual case. By using the IBC database, we leverage this human judgment. News stories are, for the most part, concerned with *something* that happened. This approach relies on reporters to break the conflict down to reportable incidents. A piece of reporting, and its resulting story thus provide a plausible operational definition of an incident. The resulting data, when properly interpreted, agrees well with our previous findings and is not likely to be improved upon within the context of the Iraq conflict.

Results

An initial look at all of the incidents reported in the IBC database between 2003 and 2009 (figure 3) seems to align reasonably well with the pattern that we have observed in Kenya and Guatemala.

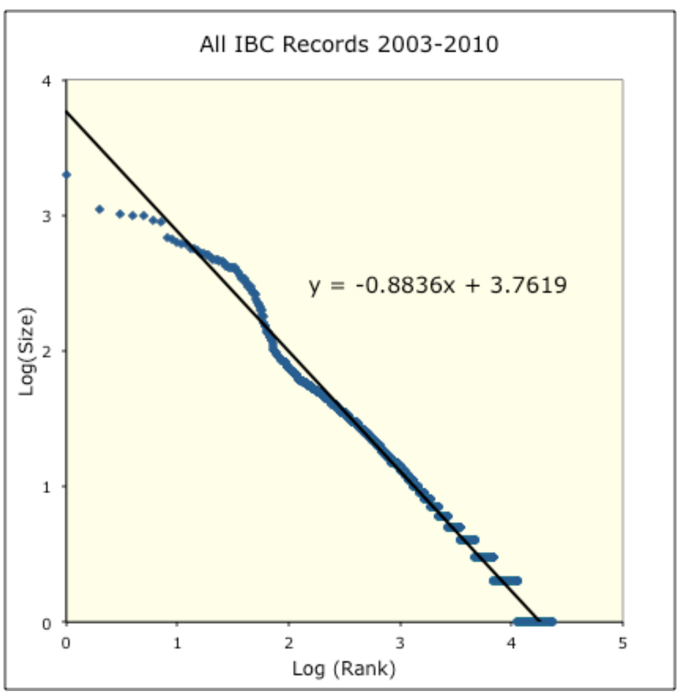

Figure 3. Rank-size plot of all records from the Iraq Body Count (IBC) database dated between 2003 and 2009 on log-log axes.

A closer inspection of this plot, however, reveals a set of serious deviations from the power-law approximation. Close inspection of the data themselves reveal that the vast majority of the high casualty records are not reports of incidents at all, having titles like "Baghdad Morgue Report, April 2005". These records can be readily identified because they are tagged not with a single date, but with a date span that is most often either a week or a month long. While these do represent real information about the conflict, they do not fit our notion of an incident. Also, if the rest of the reporting were complete, morgue reports would constitute double counting as a killing would be reported both as an incident on the street, and then again when the body was taken to the morgue.

Borhoquez et al (2009) analyze incident data from a number of conflicts including IBC data from Iraq, finding comparable power-laws for insurgent conflicts (though specified differently). They use these data to develop a regrouping model adapted from physics that is shown to fit the data well. However, both rank-size plots and a small sample of their data that were published as supplementary material to the main article demonstrate that this group failed to drop grouped records such as monthly morgue reports. This raises serious questions about the meaningfulness of their published model – a model that fits all data (even data that are inappropriately defined) has limited explanatory power. There are interesting parallels to be drawn between the model of Borhoquez et al. and the on presented here, but that comparison lies beyond the scope of this paper.

With the grouped records dropped, the rank-size plot comes to look like a much worse fit for a power-law, with its upper tail now displaying a definite bending in log-log coordinates as seen in figure 4.

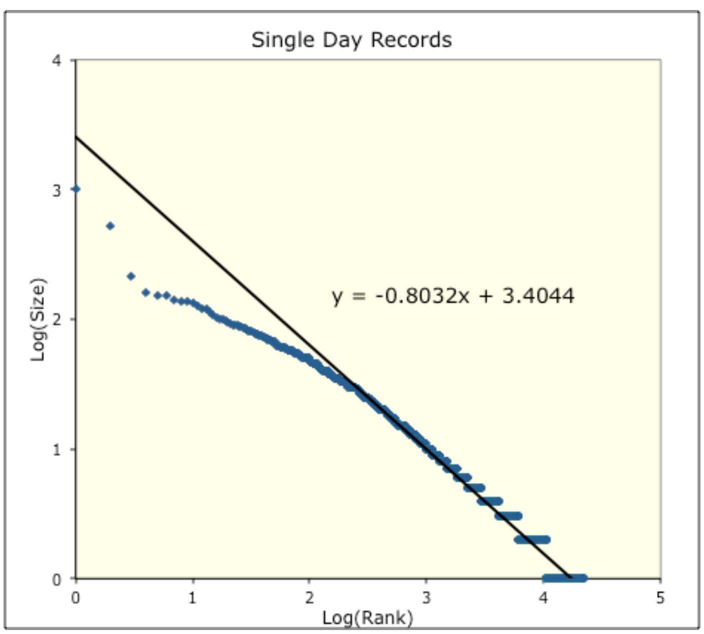

Figure 4. Rank-size plot of single-day records from IBC database.

We can gain insight into the nature of this deviation from the power-law distribution by moving from a rank-size plot to a probability density function (PDF) plot as shown in figure 5.

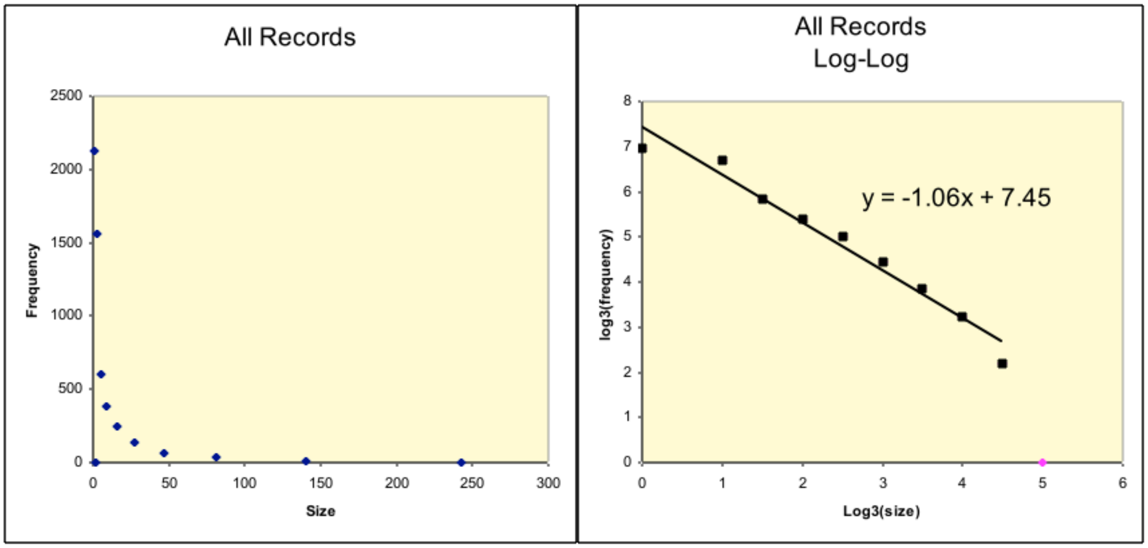

Figure 5. Probability density function (PDF) plot of violent incidents in Iraq between 2003 and 2006. Left plot in linear coordinates, right plot in log-log coordinates.

The PDF displays size (number killed) on the x-axis and the frequency of events of that size on the y-axis. For perfectly power-law distributed data, a rank-size plot and a PDF plot can be expected to have the same slope (exponent) on these two types of plot. Data are binned by whole number intervals on a log base 3 scale. This scale is used because it avoids introducing artifacts due to whole number cutoffs among the low frequency events. When plotted on $\log_3$-$\log_3$ coordinates, the data fall once again along a straight line over a limited range and then fall off dramatically.

This indicates that we are looking at a truncated power-law, where the smaller (more frequent) incidents are described by a power-law with exponent slightly greater than one, but there is a definite cut off above which the frequency of events is greatly reduced. The larger events, in other words, appear to be missing as compared to the power-law distributed data that we might expect.

We can see this truncation more intuitively by constructing a shifted rank-size plot for the incidents between 2003 and 2006 where we assume that the largest 1% of incidents are missing. IBC documents 7,098 incidents during this time. Rather than ranking the largest of these as #1, we renumber it as #71 – assuming that the 70 largest incidents are somehow "missing" relative to what we would expect. The result, shown in figure 6, conforms well to the pattern we have seen in Kenya and Guatemala, with the data falling on a reasonably straight line in log-log coodinates and a power-law exponent slightly bigger than -1.

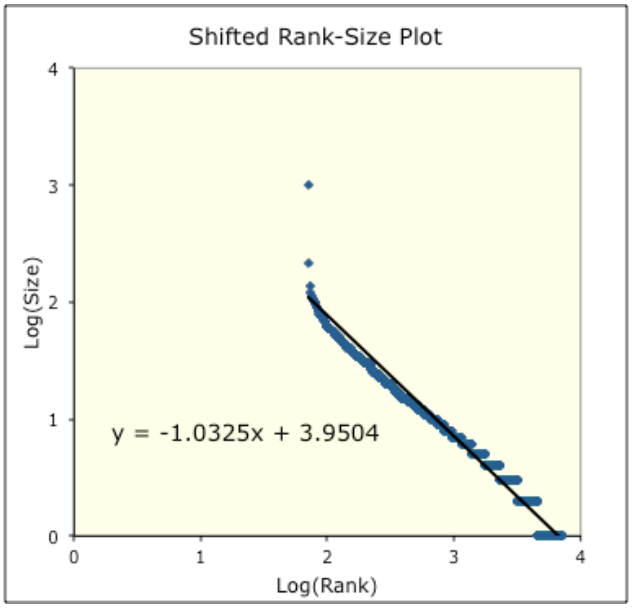

Figure 6. Shifted rank-size plot showing single-day records from IBC with ranking shifted rightward by 1% (7,098 records with the largest ranked #71).

This pattern holds up if we break the data out by year with particularly strong conformance in the 2003-2006 period and meaningful, but less precise conformance in the years 2007-2009. This can be seen in figure 7.

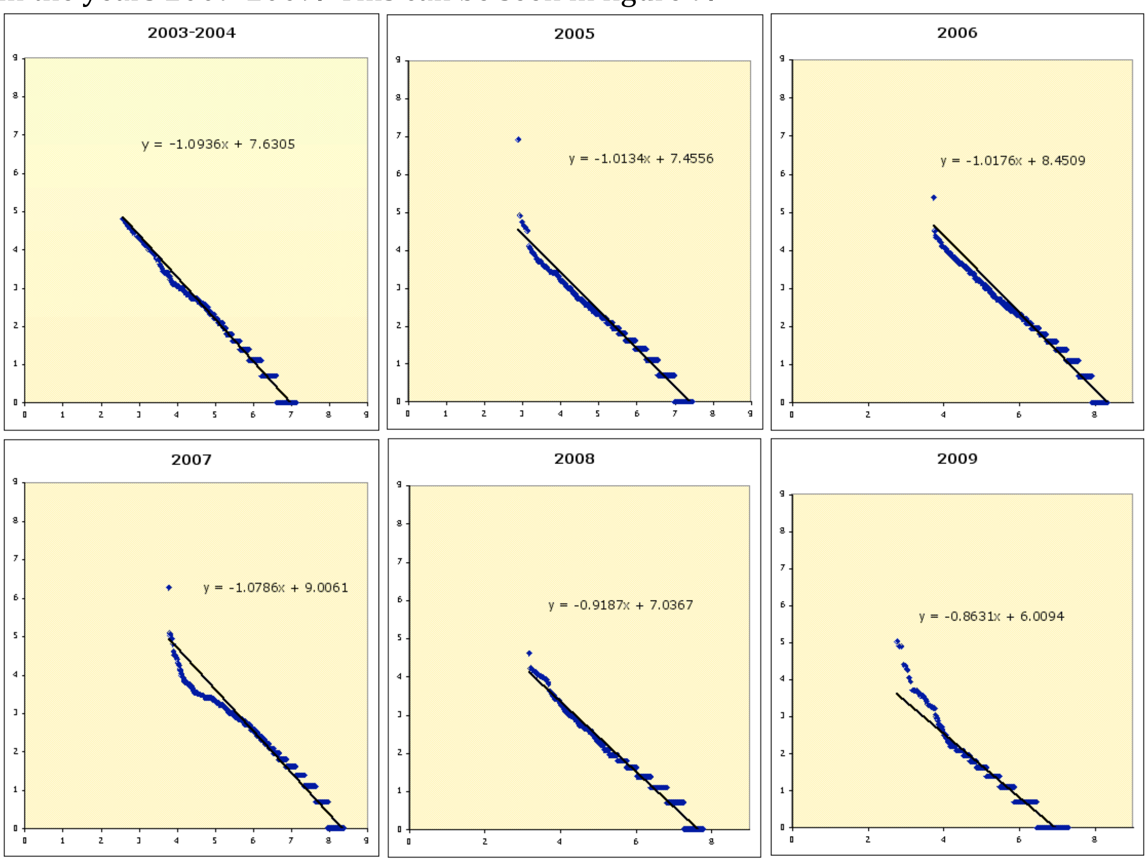

Figure 7. Shifted rank-size plots (1% rank shift) for single-day incidents for the years 2003-2009. 2003 and 2004 are combined because of smaller incidents counts.

Analysis

In examining the IBC data from Iraq for the years 2003 through 2009, we have found evidence of a truncated power-law pattern that, aside from this truncation, has an exponent much like the incident data observed in Kenya and Iraq. This section proposes a simple model that accounts for this truncation using the stylized facts of the war in Iraq. We will use social network structure to analyze the situation and propose that the unusual, truncated, nature of the violent incident distribution in Iraq stems from the unusual social organization of the country during this period.

We will base this analysis on Herbert Simon's analysis of the span of control in organizations. [Simon, 1957] Figure 8 presents a stylized hierarchical control network that can be used to visualize this idea.

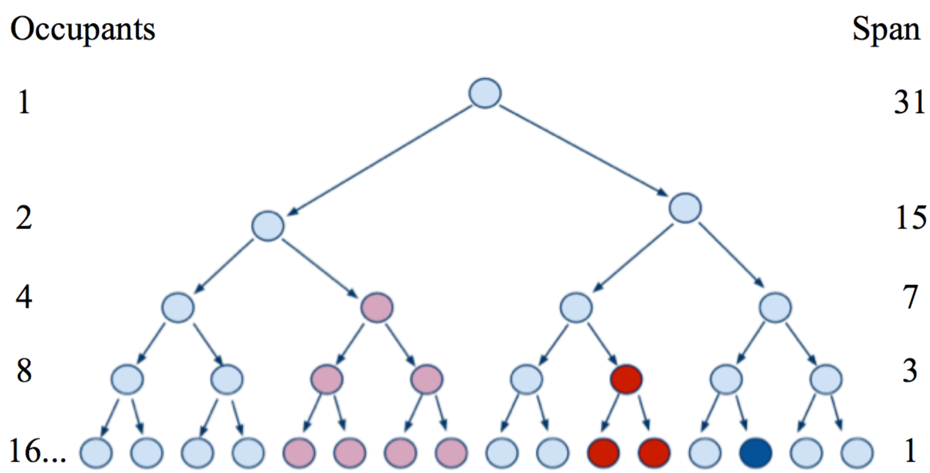

Figure 8. An example of a hierarchical network illustrating the number of occupants at each level of an organization and the span of control of those occupants.

Simon observes that the distribution of spans of control for such a network approximates a power-law distribution with a rank-size exponent slightly larger than -1. This exponent is, perhaps surprisingly, quite insensitive to the branching structure of the hierarchical network, falling between -1 and -1.2 under a very broad range of hierarchical network configurations.

This could account for the power-law regularity in conflict data if decisions were being made and actions undertaken at all levels of the network. The idea here is that each person in the network can be responsible for a number of deaths that is roughly on the order of his (usually "his" is the appropriate pronoun in such conflicts) span of control. A person at the bottom of the hierarchy can be killed only once and can kill only a small number of people acting as an individual. A person who controls a dozen people can cause (or take) on the order of a dozen causalities, and so on up the hierarchy.

But how plausible is it that decisions and actions are being taken more-or-less equally at all levels of the hierarchy? In the course of a conflict, this seems quite plausible. Wars can be thought of as self-organized critical systems [Cederman, 2003] and such systems are characterized by fractal structure and activity at all levels of organization [Bak, 1996]. During a heated conflict, every combatant is required to do his or her absolute utmost. If a sentry is sleeping, it puts his whole unit in grave danger and an inattentive General can put his whole army at risk. It seems plausible that the activity level of people at all levels of the hierarchy are operating at a pace determined by their (approximately equal) human capabilities while the impact of these actions corresponds roughly to their spans of control.

We turn now to the truncation of the power-law in the case of Iraq.
Before the US invasion, under Saddam Hussein, the society was somewhat divided along sectarian lines between the majority Shia sect and the ruling (but less populous) Sunni sect. Saddam's state apparatus, embodied largely in the Baath party, prevented significant organization among the Shia. Anyone who came to control more than a mosque worth of people (perhaps a couple of hundred) was seen as a grave danger to the state and was in danger of imprisonment or worse. The Shia were therefor organized only at low levels – from a couple of hundred downward.

The Baath party, in contrast, was a strongly hierarchical system, with Saddam Hussein at its top. It was not purely Sunni, but was Sunni dominated and also included more secular aspects. The Baath party was first and foremost a political rather than religious organization.

One of the primary actions that the US military undertook upon taking control of the country was the disbanding of the Iraqi military and the "de-Baathification" of the Iraqi state. This had the effect of removing the top several layers from the Baath party hierarchy, reducing the organizational level of Sunni society to something qualitatively similar to that which the Baath party had imposed on the Shia.

Figure 9 shows a hypothetical set of 7,654 simulated incidents based on an nine-level hierarchal network with degrees of branching *3, 3, 3, 3, 3, 3, 3, 2, 0*, assuming that each person in the network is responsible for one incident on average. This is an arbitrary branching structure – a full exploration of the relationship between power-law exponent and network branching structure is beyond the scope of this paper.

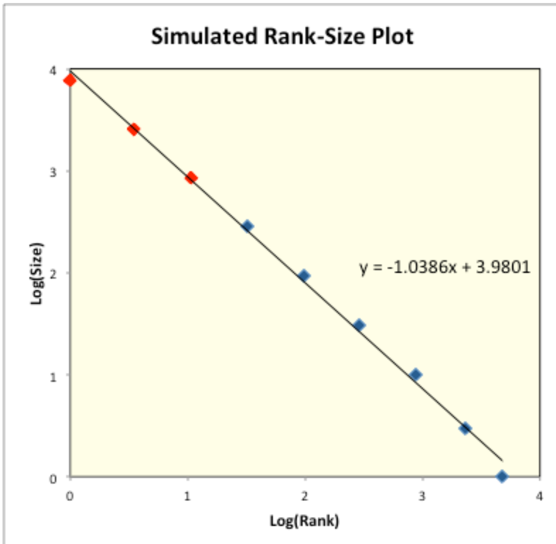

Figure 9. Hypothetical rank-size plot of 7,654 incidents generated by a nine level hierarchal network with branching structure 3, 3, 3, 3, 3, 3, 2, 0. Incidents in red are generated by the top three levels of the hierarchy and are "missing" in the Iraq data.

Figure 9 can be compared to figure 6, which shows the rank/size plot for all single day incidents between 2003 and 2006. They display a similar number of events, arranged with a similar rank-size power-law exponent. If we hypothesize that the top three levels of hierarchal organization were removed (on the Shia size, by Saddam Hussein, on the Sunni side by US de-Baathification) we see a comparable truncation of the resulting power-law.

It is important to note that while the power-law exponent is largely insensitive to the branching structure of the network, the number of incidents and the truncation point of the power-law are sensitive to this structure. The structure shown here was chosen to correspond roughly to the data from Iraq. Further investigation is required to relate this model to empirical social structure.

Discussion

We have presented evidence of a power-law structure for the sizes of violent incidents in various civil conflicts. We have then proceeded to identify a truncated form of this power-law in incident data from Iraq and have presented a simple model based on the span of control in a hierarchal network that is capable of explaining both the power-law distribution, along with its characteristic exponent and the truncation in a way that corresponds with the stylized facts of the Iraq conflict.

This model allows us to speculate on two other phenomena that are visible in the data presented.

First, the model may be helpful in understanding the difference between the counterinsurgency and genocidal warfare sides of figure 2, showing the different parts of the Guatemala conflict. The counterinsurgency part of the conflict follows the pattern of a self-organized critical system with hierarchal organization. Combatants on both sides are doing their utmost both to prevail and to protect themselves (though it should be noted that the Guatemalan government was much better organized and equipped and therefor accounted for a disproportionate portion of the killing). The genocidal part of the conflict does not follow this pattern. It contains fewer small incidents and smaller large incidents relative to what we would expect. It does not appear that soldiers are making decisions and taking action at all levels. Rather, the pattern that we see is consistent with mid-level commanders giving orders to somewhat consistent sized groups of soldiers to attack somewhat consistent sized groups (villages). The fact that this part of the conflict is not consistent with self-organized criticality may be independent evidence that the actions of the Guatemalan armed forces and related paramilitary groups were not legal acts under international laws of war.

Second, we note that the truncated power-law is a good fit for Iraq in the years 2003-2006 but becomes a worse fit as the conflict proceeds. 2009, in fact, may be better approximated by a non-truncated power-law. This is consistent with the "Sunni Awakening" and related strategies that started to be adopted by the US military in 2007 and beyond that involved both building up the Iraqi army and turning over some areas to local security forces who could maintain local order. This may have lead to more organized (and less) fighting, producing a power-law distribution of incidents that looks more like the ones we have observed elsewhere.

The model suggests an interesting trade-off in a conflict like the one in Iraq. On the one hand, the disbanding of the Iraqi army, de-Baathification, and the suppression of influential, militant Shia like Muqtada el Sadr prevented many large battles and violent incidents that would have produced significant loss of life. On the other hand, it is impossible to negotiate a peace agreement or a constitution when no one is in a position to speak for more than a few hundred people. The violence resulting from this situation was perhaps of a somewhat lower intensity than it would have been had the factions been fully organized, but it was very likely more intractable and extended. The model suggests that efforts beginning in 2007 to facilitate higher-level social organization, even among groups that were at least somewhat hostile to the US, were an unavoidable part of bringing the violence under control.

While this research is rather speculative at this point, we believe that it offers sufficiently plausible explanations of a number of observed phenomena in conflict micro-data including the persistent power-law distribution of violent incidents across very different conflicts, the different distribution of incidents in the genocidal part of the Guatemalan conflict, and the truncated power-law visible in data from Iraq. While further research in this area is required, we suspect that both hierarchal organization and self-organized criticality are necessary components for successful modeling and understanding of civil conflicts and that the kinds of power-law related incident distributions that we have observed here may be useful as basic validation of such models in the future.

References

Bak, Per. 1996. "How Nature Works: The Science of Self-Organized Criticality." *New York: Copurnicus*.

Ball, Patrick (1999): AAAS/CIIDH database of human rights violations in Guatemala (ATV20.1). http://hrdata.aaas.org/ciidh/data.html (July 1, 2000).

Bhavnani, Ravi, Dan Miodownik, and Hyun Jin Choi. 2011. "Violence and Control in Civil Conflict." *Comparative Politics*. http://graduateinstitute.ch/webdav/site/political_science/shared/political_science/Professors/Bhavnani/civilConflict.pdf.

Bohorquez, Juan Camilo, Sean Gourley, Alexander R. Dixon, Michael Spagat, and Neil F. Johnson. 2009. "Common Ecology Quantifies Human Insurgency." *Nature* 462 (7275): 911–14. doi:10.1038/nature08631.

Burnham, Gilbert, Riyadh Lafta, Shannon Doocy, and Les Roberts. 2006. "Mortality after the 2003 Invasion of Iraq: A Cross-Sectional Cluster Sample Survey." *The Lancet* 368 (9545): 1421–28.

Cederman, Lars-Erik. 2003. "Modeling the Size of Wars: From Billiard Balls to Sandpiles." *The American Political Science Review* 97 (1): 135–50.

Clauset, Aaron, and Frederik W Wiegel. 2009. "A Generalized Aggregation-Disintegration Model for the Frequency of Severe Terrorist Attacks." *Journal of Conflict Resolution*.

Gulden, Timothy R. 2002. "Spatial and Temporal Patterns in Civil Violence: Guatemala, 1977-1986." *Politics and the Life Sciences*, 26–36.

Iraq Body Count. 2011. "The Iraq Body Count Database." https://www.iraqbodycount.org/

Jones, Kevin DeWitt, 2008. Personal communication.

Kalyvas, Stathis N. 2012. "Micro-Level Studies of Violence in Civil War: Refining and Extending the Control-Collaboration Model." *Terrorism and Political Violence* 24 (4): 658–68. doi:10.1080/09546553.2012.701986.

Richardson, Lewis F. 1948. "Variation of the Frequency of Fatal Quarrels with Magnitude." *Journal of the American Statistical Association* 43 (244): 523–46. doi:10.1080/01621459.1948.10483278.

Simon, Herbert A. 1957. "The Compensation of Executives." *Sociometry* 20 (1): 32–35.