Final Project [INST 633 Summer 2016]

Analysis of the Iraq and Coalition forces social networks during the Iraq War
Gowtham Ashok

Introduction and Background

The Iraq war was an armed conflict that was fought between Iraq and the United States along with its coalition allies. It was conducted amid allegations that Saddam Hussein, the leader of Iraq at that time, had chemical weapons, and was planning to use them on his people. It started in 2003 and ended in 2011.

Wikileaks[1] is an international media organization that was founded in 2006 by Julian Assange. They specialize in analysis and publication of restricted materials usually referred to as "leaks". They released the Iraq war logs [2] (US Army field reports) in 2010. It details the US war activities between the years 2004 to 2009, contains about 60% of the recorded civilian deaths and is considered the biggest military leak in the history of the United States.

This collection of data is interesting to me, because it provides an insight on how modern wars are conducted and recorded. I chose to do this as it seemed to be challenging to extract information from a set of logs and construct a social network out of it. The network here is between US-led coalition forces and people in Iraq (Civilians and enemies).

There is a lot of data on the Iraq war, and I narrowed to using the Iraq war logs as a main source. I also went to areas at a safe distance from Syria during the summer to get an idea on the ground, on the war works.

Related work

This topic is a gold mine for sociology research and hence has a lot of material. I chose to focus on papers and articles that detail how the dynamic worked between Iraqi civilians and the US-led military coalition worked in reality.

Hicks MHR, Dardagan H [3], analyzed civilian deaths caused by weapons used, their demographics and the region in which it was done. They used the Iraq Body Count database for it, and came to a conclusion that most deaths were caused by "unknown" perpetrators, by attacks directed at them by terrorists and by the coalition forces. They also came up with an index called a "Dirty War Index" to measure the proportion of women and children killed in the war. A NY Times article [4] analyzed the logs and concluded that the public reporting of deaths was lower than the actual number. They also found the extremely high usage of private contractors, who sometimes were greater in number than the actual soldiers themselves. They found that the abuse of prisoners by soldiers was mostly not scrutinized by the American military. The Guardian [5] published a map that gives a visualization of the places in Iraq where the civilian deaths occurred. This shows that the deaths were concentrated in certain regions, mostly cities. NY Times [6] published an interactive demo showing a glimpse of the issues faced in the war logs. It also summarizes them, and shows them in context. It organizes them based on categories such as Iran, prisoner abuse, kidnapped hikers, et cetera. An article from the Guardian [7] provides the weaknesses that are present in the data. They pointed to the massive underreporting of deaths and the reasons for it. They also pointed out the changing standards for reporting events. They also tell about a possible bias in their reviews, which may be due to their way of reporting.

Research questions and setup

Research Questions:
What information can we get from the Iraqi War logs? What made the war difficult? How can we use that information to avoid the same mistakes to avoid in future wars?

Data collection process

The primary data source for the data leak was down, so I used a compressed version of the log from the Guardian stored in Google Fusion tables. It contained only entries that recorded deaths. This log was then processed using Python, and R.

Other sources for this analysis include reddit posts, blogs, news articles, papers, a trip to Turkey. I went to areas which were at a safe distance to Syria in Turkey. I talked to some people, who knew English, and got an idea of context. I got a general idea of the civilians in the region.

Analysis methods

I did text parsing, data cleaning for the data. I used Stanford's Named Entity Recognizer, integrated it with Python Natural Language Processing Toolkit to get important persons and organizations from the dataset. I tried some NLP algorithms present in the NLTK toolkit. I also loaded the dataset on Gephi and performed various actions to get meaningful data.

I read through the logs on the wardiaries website maintained by wikileaks, and also read many articles on the internet that referenced the leak.

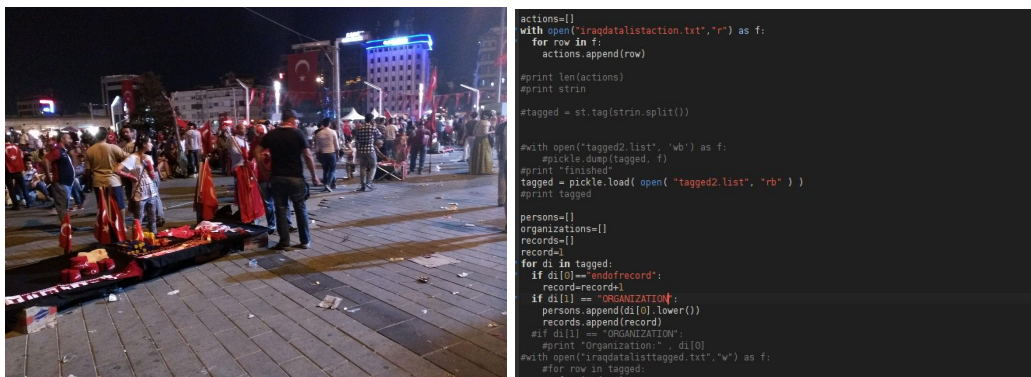I also analyzed reddit posts about modern wars (from a US perspective) to get the soldier's perspective on wars.

I interacted with civilians in the region of Turkey close to Syria, which had similar characteristics to the conditions described in the logs. I also interacted with demonstrators in the main town square in Istanbul. I only used them to get a better idea of the data present in the logs. I could not use them in writing results as I did not record the data. I did not do so, because there was a threat of getting kidnapped.

Environment

Computer: i7-2670QM, 16GB RAM
Tools used: Python 2.7, wordcloud-python, python-nltk, httrack, wget, Gephi, RStudio, Stanford Named Entity Recognizer
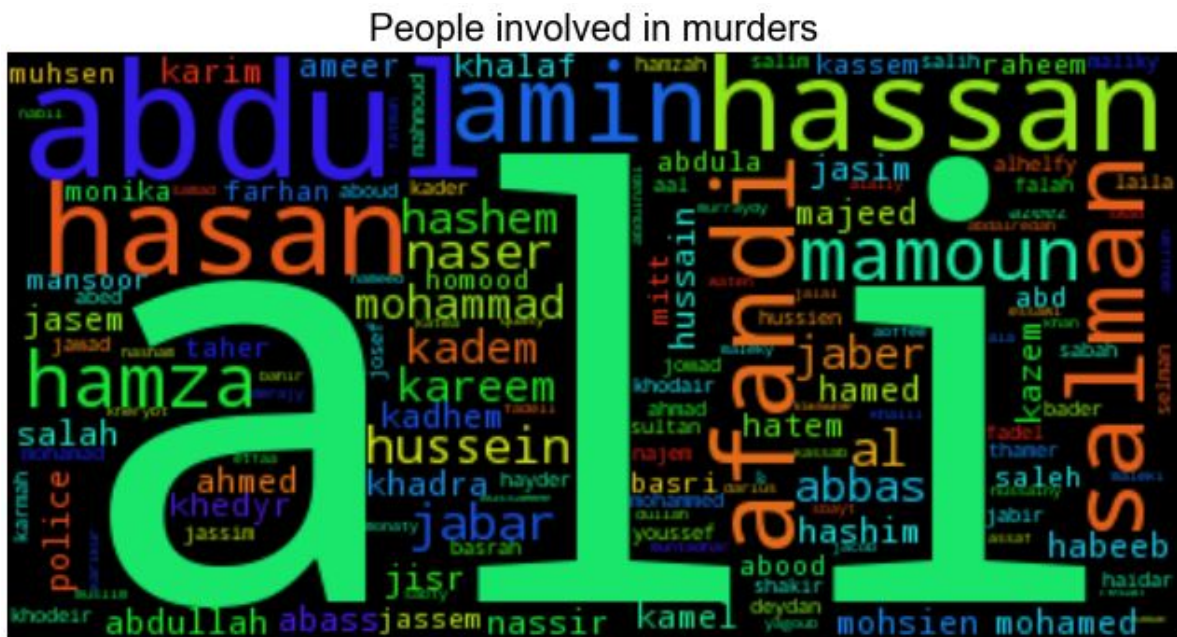Sources used: Wikileaks, The Guardian, NYTimes, Reddit, Blogs, News Articles, Turkey


```python
actions=[]
with open("iraqdatalistaction.txt","r") as f:
    for row in f:
        actions.append(row)

#print len(actions)
#print strin

#tagged = st.tag(strin.split())

#with open("tagged2.list", 'wb') as f:
    #pickle.dump(tagged, f)
#print "finished"
tagged = pickle.load( open( "tagged2.list", "rb" ) )
#print tagged

persons=[]
organizations=[]
records=[]
record=1
for di in tagged:
    if di[0]=="endofrecord":
        record=record+1
    if di[1] == "ORGANIZATION":
        persons.append(di[0].lower())
        records.append(record)
    #if di[1] = "ORGANIZATION":
        #print "Organization:" , di[0]
#with open("iraqdatalisttagged.txt","w") as f:
    #for row in tagged:
        #f.write(row)
```

Results

Figure 1: This visualization shows the most common names of people found in murder reports.

Excerpts from the log:

"MND-N G2 reported a kidnapping of 3 Americans who were being taken to the Iranian border. The Americans were hiking near the Iranian border when taken. A fourth tourist did not go hiking with them and reported that a kidnapped female called him saying that they were being surrounded by armed men." [8]

"ON 29MAY05 AN MNF-I APPROVED JOINT INTERROGATION PROJECT WITH THE IRAQI POLICE IN MOSUL REVEALED EVIDENCE OF PRIOR TORTURE/ABUSE ON 3X LOCAL NATIONALS THAT HAD BEEN CAPTURED APPROX. 15MAY. EXAMINATION OF THE DETAINEES SHOWED LACERATIONS ON WRISTS FROM HANDCUFFS, BRUISING ON THE BACK AND THIGHS, BRUISING ON FACE. DETAINEES ALLEGE THAT THEY WERE BEATEN BY POLICE WITH CABLE ON THE BACK, CHEST AND FACE; HUNG BY THE WRISTS AND FORCED TO CONFESS TO TERRORIST ACTS. " [9]

Excerpt from the perspective of a coalition soldier:

"Let me tell you a story...when I was in Iraq, the bad guys would give realistic toy guns to the Iraqi children in hopes that we would start shooting these kids. Well, one day, I was up in the turret of my HMMWV and we were driving down a city street. To stay alive, I would scan people's hands, not faces, because threats come from hands. As I scanned the alleyway that we were driving past, my stomach nearly came out of my throat because I realized that I was looking down the barrel of an AK-47. It was pointed right at me and it was only 30 meters away. I switched off the safety on my gun, moved my finger to the trigger and then looked through my ACOG (3.5x magnification). My ACOG gave me a clear view of my target, a boy, that looked to be about 10-12 years old. I still couldn't tell that the gun was fake, but I figured that I have more armor and that I am a much better shot than this kid, so I am going to wait until he shoots

first.He dropped the gun and ran away. thank God. We later determined the gun to be fake. All of this happened in a split second." [10]
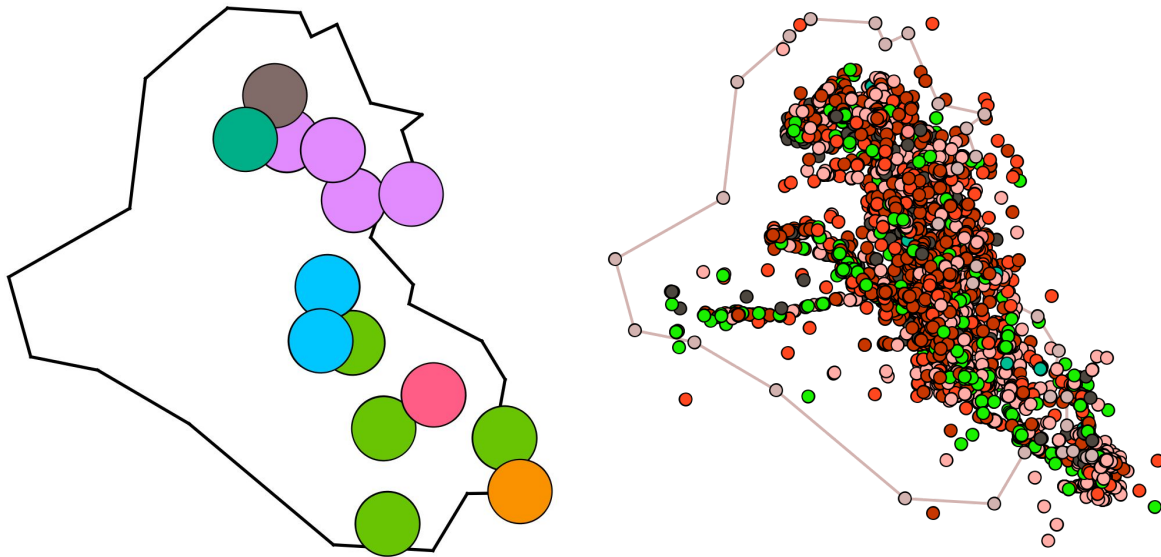

Figure 2: These graphs visualizes the ethnic composition(left) the network of deaths(right) in Iraq. They are color-coded. On the ethnic composition graph, the green ones represent majority shias, the blue ones sunnis, and the purple ones, Kurds. On the death network graph, the red deaths are caused by the enemy, while the greener ones are caused by coalition forces (intentionally or unintentionally).

Discussion

Figure 1

Figure 1 shows that most of the people involved in murders or are murdered, have common arabic names. I could not find any case of European-sounding names in the logs. I found that the reason for this was due to redaction by the Wikileaks authorities [14]. Even with their redaction, some of the names leaked through. This invades the "privacy" of the individual. A concern from the government was that such information would be harmful in the wrong hands. Indeed, with the exact location and names, one can easily identify a single person and put out a hit on him/her. Hence it is imperative to have personal information stored in a separate place, or even destroyed.

Especially in times of war, it is dangerous to store personally-identifiable information that can be used maliciously in the wrong hands. Hence to combat this, we can have bug bounty programs that incentivize making the system secure, and having User Experience Researchers perform a thorough analysis, so that security issues such as these, do not exist in the system in the first place.

Log 1

From the first log, we can see how the kidnapping of tourists near the iraqi border was recorded. Similar kidnappings have occurred in this area, where people are often held for ransom. This issue is complex, as there a lot of things at play in the background. In this example, the Iranian government used this as a warning for its own people, while the kidnappers wanted money. Both parties, which are enemies most of the time, come together to achieve their own goals. The people being kidnapped are the losers in this case, as, the non-payment of ransom made

them subject to execution. One interesting observation I noted was the importance of citizenship in these cases. There have been logs where the kidnapped person's citizenship is questioned.

The current system of making the respective citizen's government pay the ransom, should be banned. There should be an efficient international system that takes away any incentive for people to be kidnapped. This will ensure that the people who are citizens of certain countries are not taken advantage of unfairly. For instance, in my readings, I have found that France pays the ransom most of the time, and hence the French citizens are quite sought after, for getting kidnapped [11].

Log 2

This log shows an instance where iraqi nationals were tortured by the coalition government. This made the war difficult for both the sides. Through instances such as these, the local people in Iraq lost trust in the coalition government. News like this propagates very quickly through the Iraqi network, as this is sensational. This made them suspicious of the occupation and hence reluctant to cooperate with them. The local informants who did assist the coalition forces in arresting criminals sometimes did it for their own personal benefit , and hence innocent people got tortured [12].

A possible solution to this problem is to have the local forces take over prisons as soon as it was possible, and to not disenfranchise the local population by undertaking actions that could be considered "war crimes".

From the perspective of a coalition soldier

This excerpt gives us an idea of what it is like to be soldier on the front line. It sometimes is difficult to determine if a person is a civilian, a friendly person or a member of the enemy. Sometimes, as in the case of the Iraq War, they can switch sides. Also, if they kill an innocent civilian, they can be court-martialed. This makes it very difficult for them to take correct decisions. A lot of the people return from service getting Post Traumatic Stress Disorder (PTSD) due to events like this.

One solution I can think of, is to treat them better, provide them better equipment and create better tools to identify the enemy. Another solution is to completely stop sending ground troops and send drones and intelligent robots instead.

Visualization map

This map gives us an idea of where deaths occur, and what type of person (friend or enemy) caused it. It also gives us details about the ethnic composition of the country. We can see that most of the deaths occurred inside Iraq. However, quite a lot of fighting occurred near the Iran border too. This map can be better understood by considering the demographics of the population. The Muslim population is divided into two sects Shias and Sunnis.

Shias side with Iran, and form the majority of the population in Iraq. On the other hand, Sunnis side with Saudi Arabia, and do not consider Shias to be "true muslims" On the northeastern side of Iraq, there are the "Kurds", some of whom want their freedom. This makes the region a very diverse and unstable place. Most insurgents are those who were formerly in the Iraqi army or had their lives destroyed during the war. They have cognition-based trust amongst their own people and hence engage in warfare against the coalition forces. This problem is not very prominent in the southwestern area of Iraq, as it is a desert. This type of complex demographics makes it difficult for outside forces to have a useful long-term effect. Hence a possible solution is to find a strong leader who can lead people and bring them together (ironically, someone like "Saddam Hussein").

Conclusion

I explored the Iraq War, mainly with the help of the Iraqi War Logs released by Wikileaks. With the help of papers and articles, I was able to get a good understanding. I wanted to explore further, so I collected data, did Language processing using python, read through the logs manually, analyzed reddit posts by members, and finally visited areas close to Syria/Iraq in Turkey. I found that there was personally identifiable information in the logs, hostage situations are complex, war crimes might have been committed; that it's hard to determine who is one's enemy and that there are complex demographics that made the war difficult.

Some of the recommendations were for any future military force to deal with personally identifiable information carefully, implement a system to discourage hostage-taking, give control to local forces as soon as possible, use advanced technologies, and to identify strong leaders from within the country.

Bibliography

[1] https://wikileaks.org/What-is-Wikileaks.html
[2] https://wardiaries.wikileaks.org/
[3] Violent Deaths of Iraqi Civilians, 2003–2008: Analysis by Perpetrator, Weapon, Time, and Location
[4] http://www.nytimes.com/2010/10/23/world/middleeast/23intro.html
[5]http://www.theguardian.com/world/datablog/interactive/2010/oct/23/wikileaks-iraq-deaths-map
[6] http://www.nytimes.com/interactive/world/iraq-war-logs.html
[7] https://www.theguardian.com/news/datablog/2010/oct/25/wikileaks-iraq-data
[8] https://wardiaries.wikileaks.org/id/D2430687-FCE7-C5BB-ED19F523FE225046/
[9] https://wardiaries.wikileaks.org/id/F9D5B1CE-2EBB-4178-878B-3E9ED6583F16/
[10]https://www.reddit.com/r/TalesFromTheMilitary/comments/2hgjrw/i_almost_killed_a_kid_in_iraq_combat/
[11]http://www.nytimes.com/2014/07/30/world/africa/ransoming-citizens-europe-becomes-al-qaedas-patron.html?_r=0
[12] http://www.al-monitor.com/pulse/originals/2015/10/iraq-arrests-confidential-informants.html
[13] http://nlp.stanford.edu/software/CRF-NER.shtml
[14] http://edition.cnn.com/2010/US/10/22/wikileaks.editing/