- Social media, such as Twitter, have been exploited by trolls to spread disinformation during the 2016 US Presidential Election. Trolls are users with intentions to influence the public opinion by posting misleading or inflammatory information with malicious intentions.
- We focus on the content analysis of troll tweets to identify the major entities mentioned and the relationships among these entities, to understand the events and statements mentioned in Russian Troll tweets coming from the Internet Research Agency (IRA), a troll factory allegedly financed by the Russian government.

Table 1: Text preprocessing example.

Before Preprocessing	After Preprocessing
IGetDepressedWhen someone says he doesnÃƒÂ¢Ã‚Â€Ã‚Â™t have any options except for voting for Trump	I Get Depressed When someone says he doesnt have any options except for voting for Trump
AnthonyVVeiner: HillaryClinton And you have turned the Middle East into a living hell. https://t.co/c1CvUossWX	Hillary Clinton And you have turned the Middle East into a living hell.

Table 2: Further text preprocessing for K-Means Clustering.

Before Preprocessing	After Preprocessing
Hillary Clinton And you have turned the Middle East into a living hell.	hillary clinton turn middle east live hell
Angry patriots are about to paint Philly red. say no to lies corruption and tyranny! Like share; join! Philly	angry patriot paint philly red say lie corruption tyranny like share join philly

Table 3: Sentiment analysis of troll tweets.

Text	Sentiment
Sore loser Obama turns to Russian hacking to delegitimize Trump's triumph;	Very negative
Hillary Clinton And you have turned the Middle East into a living hell;	Negative
Trump Signs Obamacare Executive Order	Neutral
In light of Hillary's FBI investigation you can change your early vote in these states;	Positive

real Donald Trump is brilliant and has amazing insight; He's going to be a fantastic President;	Very positive

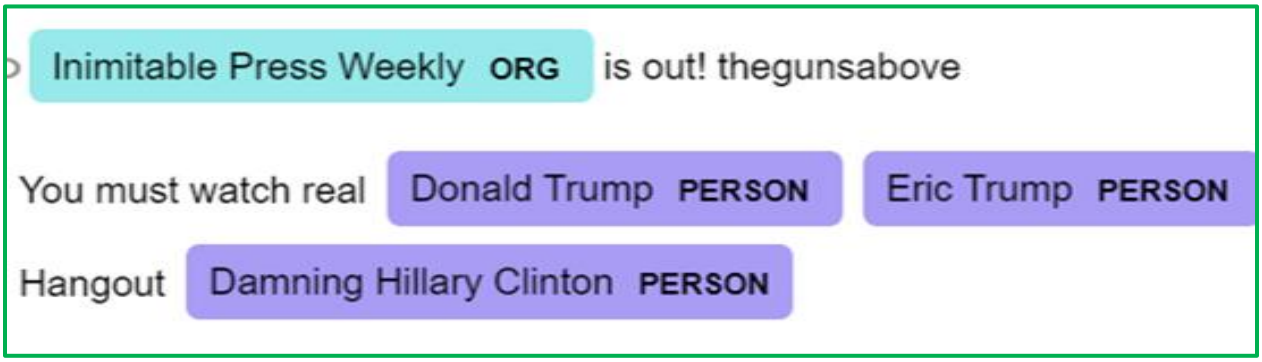

Figure 1: Named Entity Recognition and Label Visualization.

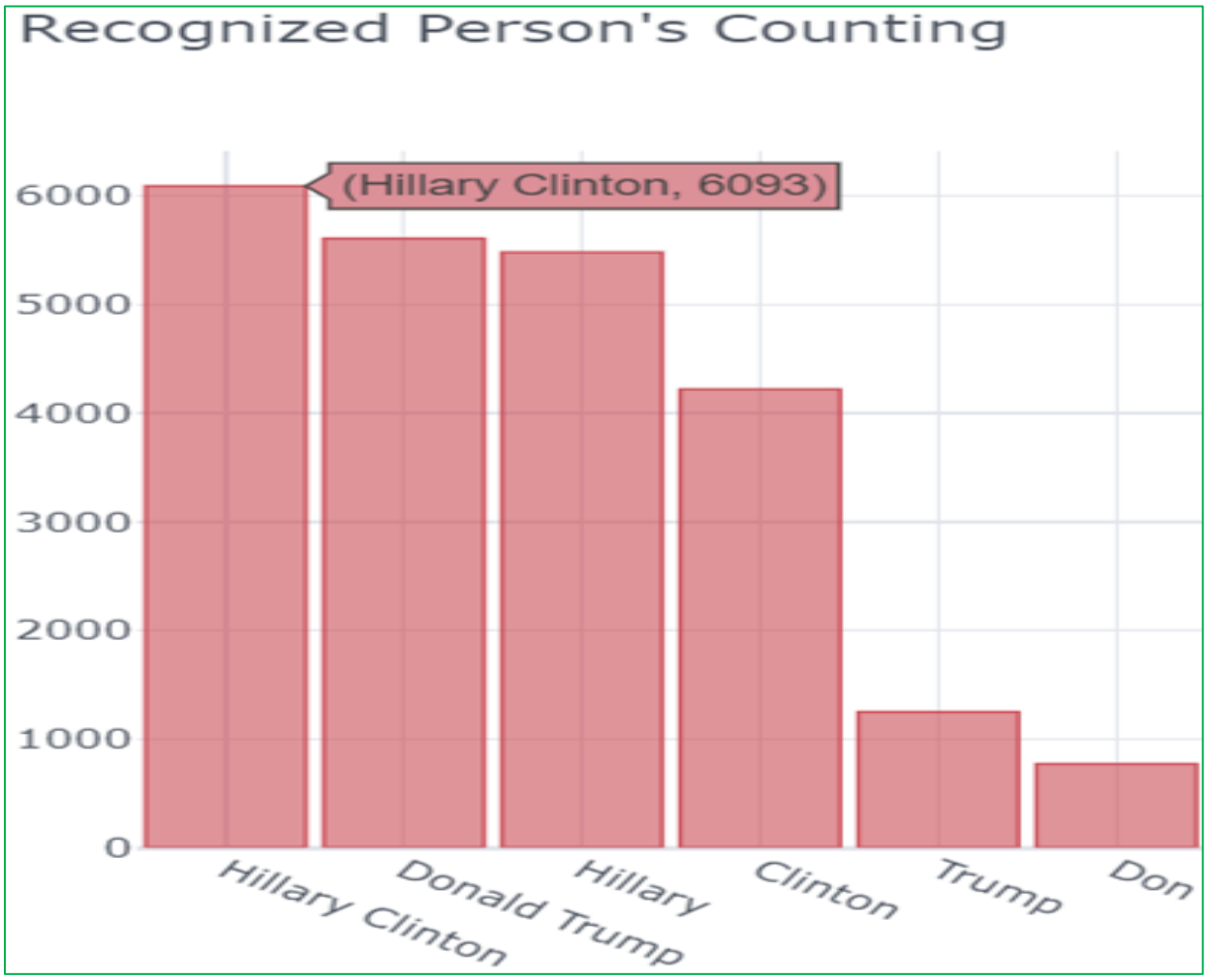

Figure 2: Top 6 Frequencies of 'PERSON' entity.

Table 4: Total Named Entity types and counts.

Entity Type	Count
Person	99,273
Org	60,596
Gpe (Geopolitical location)	30,466
Cardinal	30,354
Date	23,727
Norp (Nation/Religion)	17,455
Time	4,803
Work_Of_Art	3,318
Fac (Facility)	2,451
Percent	2,359
Money	2,241
Product	2,191
Loc (Other location)	2,124
Ordinal	1,965
Event	1,817
Law	1,268
Quantity	771
Language	144

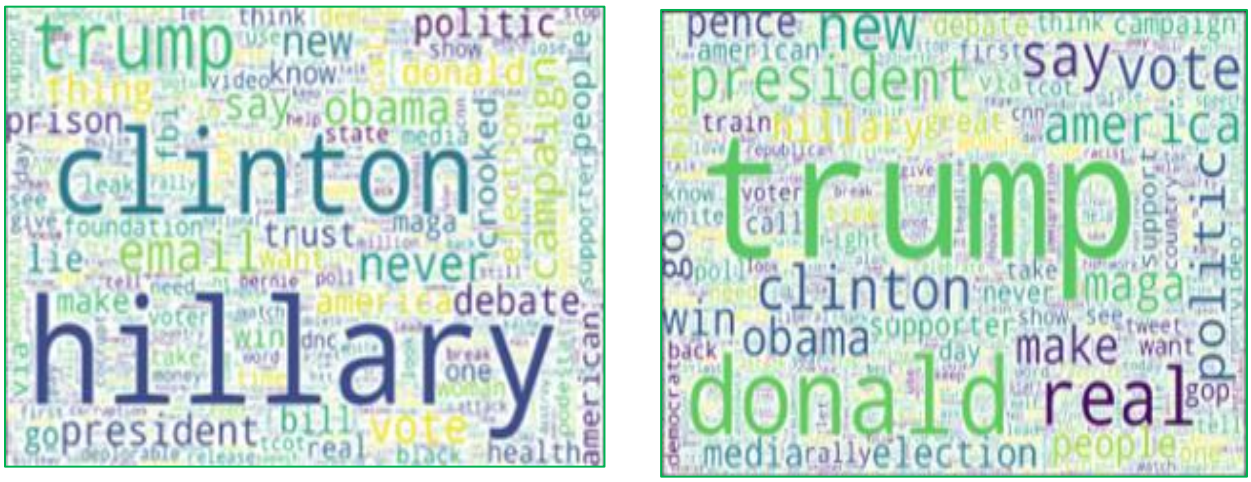

Figure 3: Word cloud of Clinton (left) and Trump (right).

Table 5: Top 15 terms in each cluster.

Cluster	Frequent terms
1 (Trump)	trump, donald, real, president, vote, politics, say, america, clinton, maga, win, media, make, obama, pence
2 (Clinton)	hillary, clinton, trump, email, campaign, vote, prison, president, trust, obama, thing, debate, say, politics, crooked
3 (Others)	obama, word, make, people, day, like, thing, say, know, love, news, black, life, new, want

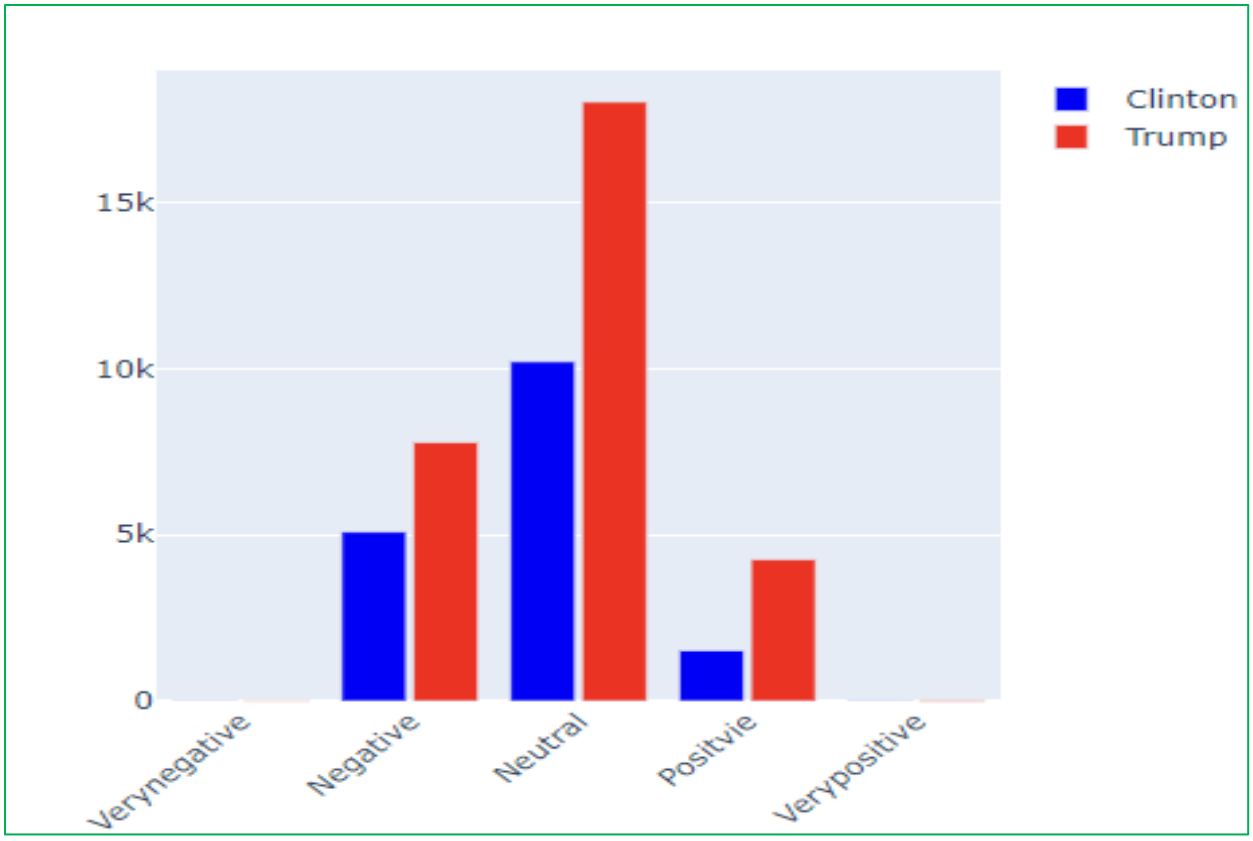

Figure 4: The sentiment distribution in Trump cluster (red), and in Clinton cluster (blue) between Very Negative and Very Positive.

We used the concern index, that we defined (Ji et al., 2013), as follows:

CI = N/(N+P), N=#(negative tweets + very negative tweets); P=#(positive tweets + very positive tweets)

Trump CI = 65%, and for Clinton CI = 77%.

According to the Z-score and p-values, this difference is statistically significant!

Table 6: Example triples extracted from the tweets.

Example Triples from Clinton cluster		
subject	**relation**	**object**
Hillary Clinton	delete	email
Clinton	order destruction	email
Clinton foundation	is most corrupt	enterprise in political history
Clinton foundation	employ	Muslim brotherhood official
Example Triples from Trump cluster		
Trump	lash out at	media
Trump	trash	mainstream media

Trump	is	worst president
Trump	Keep workplace protection for	LGBTQ Americans

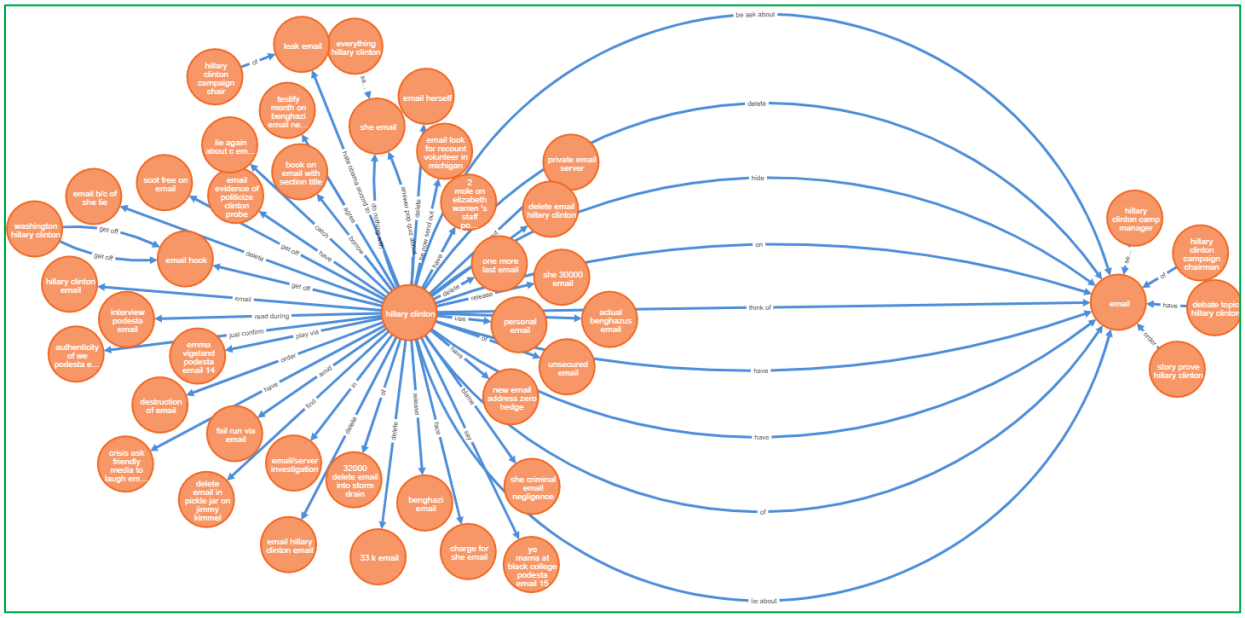

Figure 5: The triple visualization on Neo4j shows part of the relations between Hillary Clinton and email.

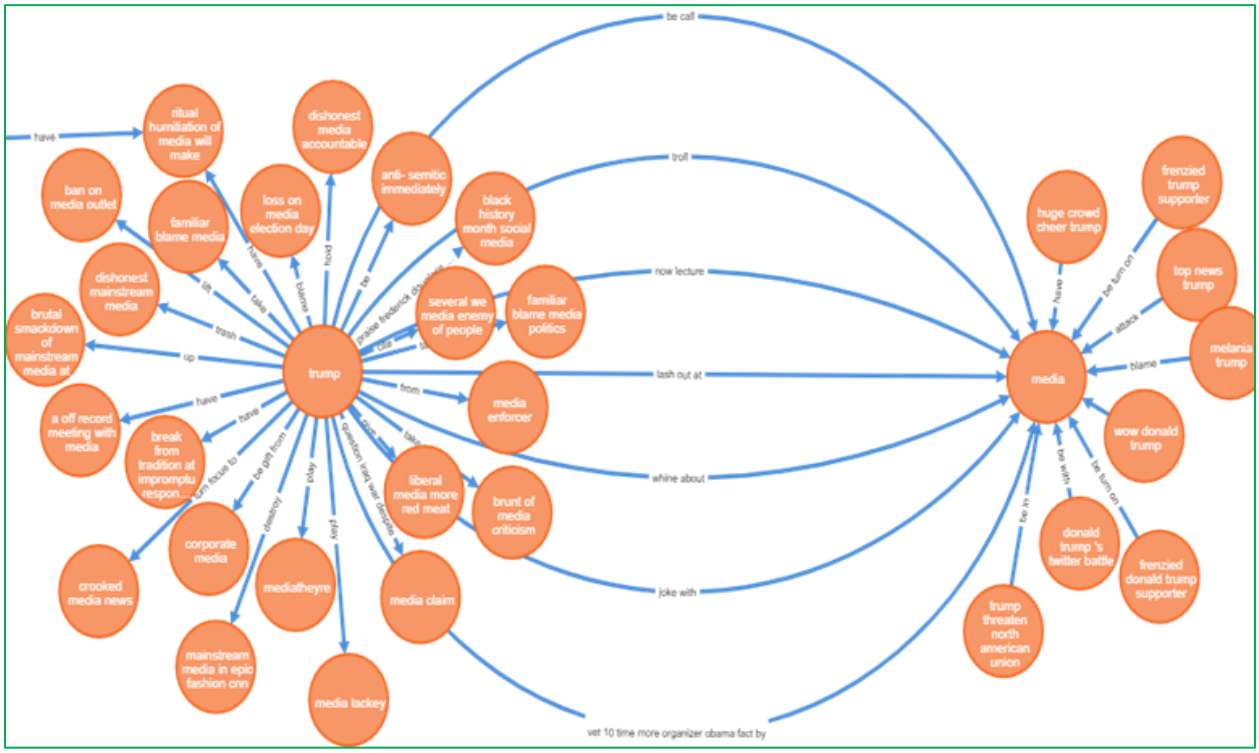

Figure 6: The triple visualization on Neo4j shows part of the relations between Trump and media.

- In this work, we presented a Knowledge Graph approach to identifying relationships among different entities in troll tweet sets from the 2016 Presidential Election, which was apparently influenced by Russian Trolls. The trolls targeted one candidate, Hillary Clinton, and her family, by repeatedly accusing her of the "email-gate scandal" and of misuse of the Clinton foundation, etc., which could be crimes if they indeed happened. Such misinformation might have greatly influenced a number of voters. The concern index in the Clinton cluster of troll tweets was the highest, and over 10% higher than in the other two clusters, which shows that Russian Trolls had used many more negative terms portraying Clinton than Trump.
- An elaborate combination of Knowledge Graph and Natural Language Processing methodologies, such as Named Entity Recognition, triple extraction (of subject, predicate, object triples), and sentiment analysis, has been applied in this work.