

Lucas A. Overbey*, Scott C. Batson, Jamie Lyle, Christopher Williams, Robert Regal, Lakeisha Williams

SPAWAR Systems Center Atlantic – PO Box 190022, North Charleston, SC 29419-9022

*lucas.overbey@navy.mil

INTRODUCTON – This work utilizes sentiment analysis and public event coding to develop tools for monitoring social and political unrest.

PURPOSE – To understand how geopolitical actors and/or politically-relevant events affect a population by quantifying their impact on the public sentiment and opinion expressed through Twitter. Applications include:

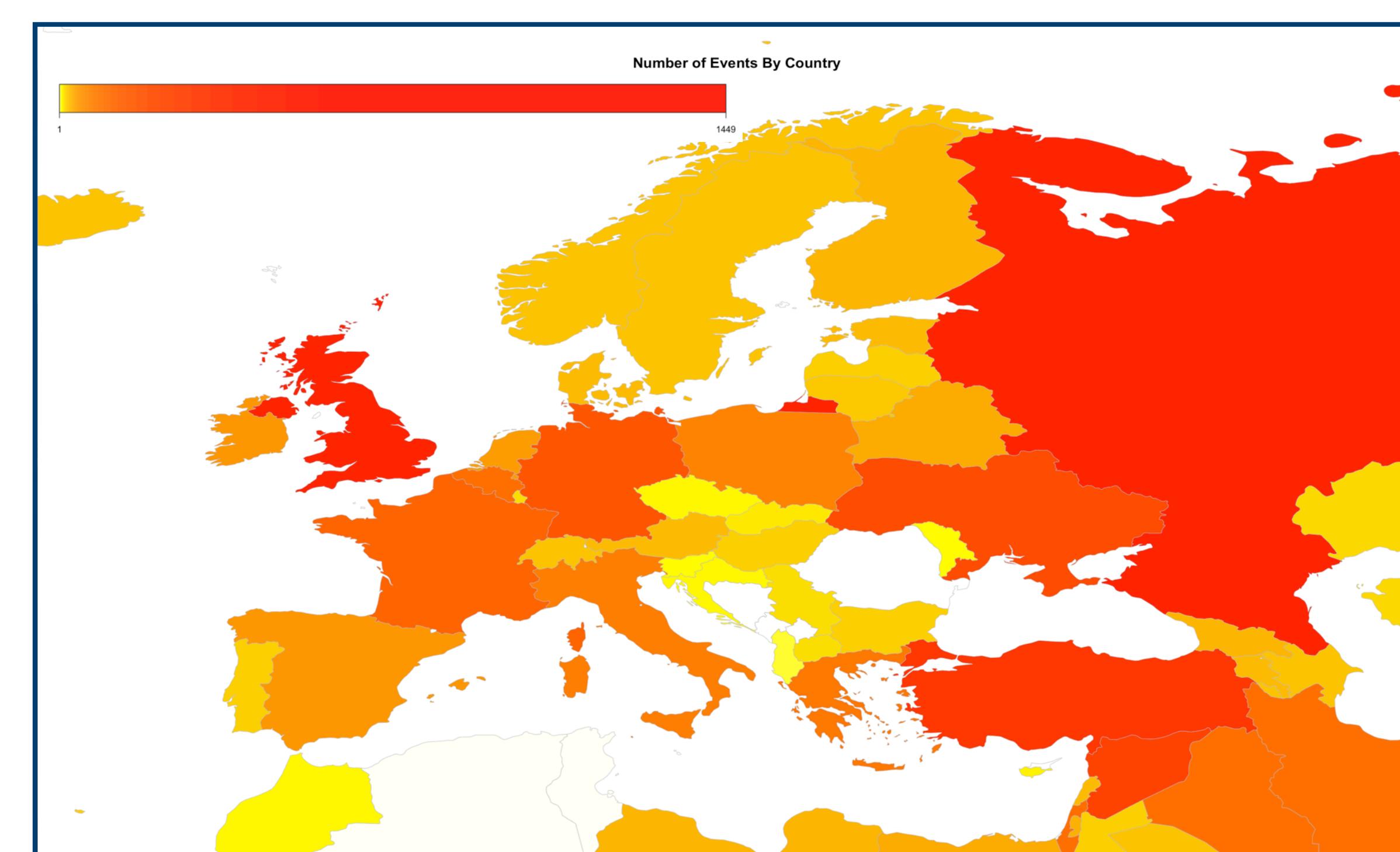
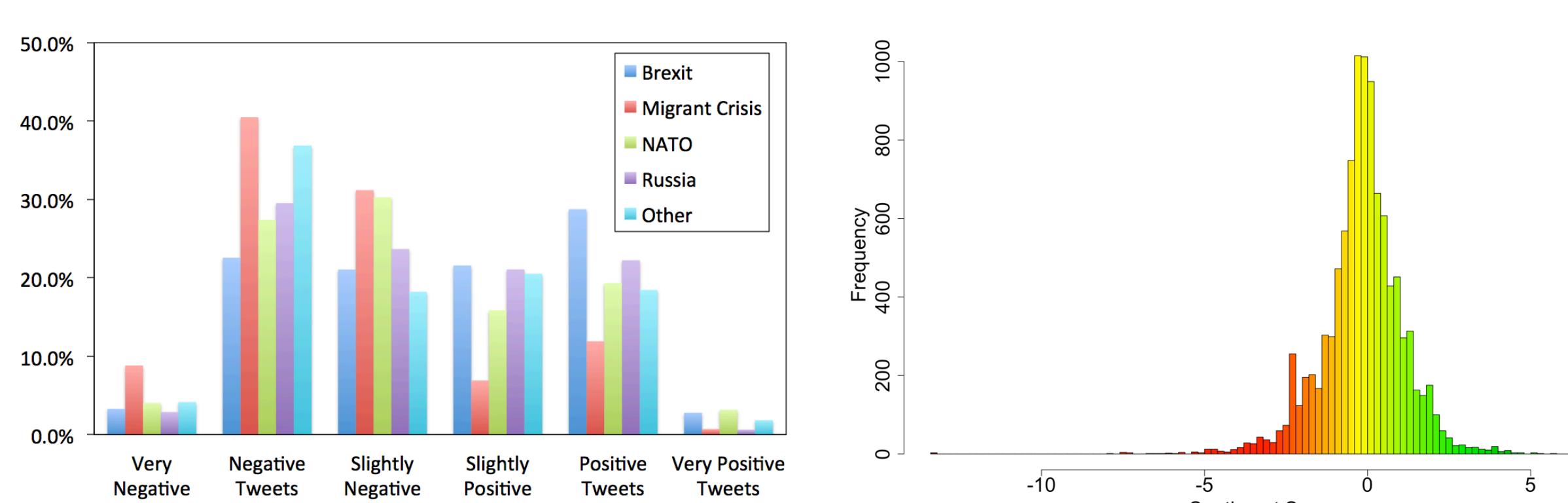
- Forecasting political instability
- Studying the escalation of conflicts and impacts of events and political actors
- Monitoring social and political unrest

HYPOTHESIS – Automated event coding may be used to gain information on sentiment aspect.



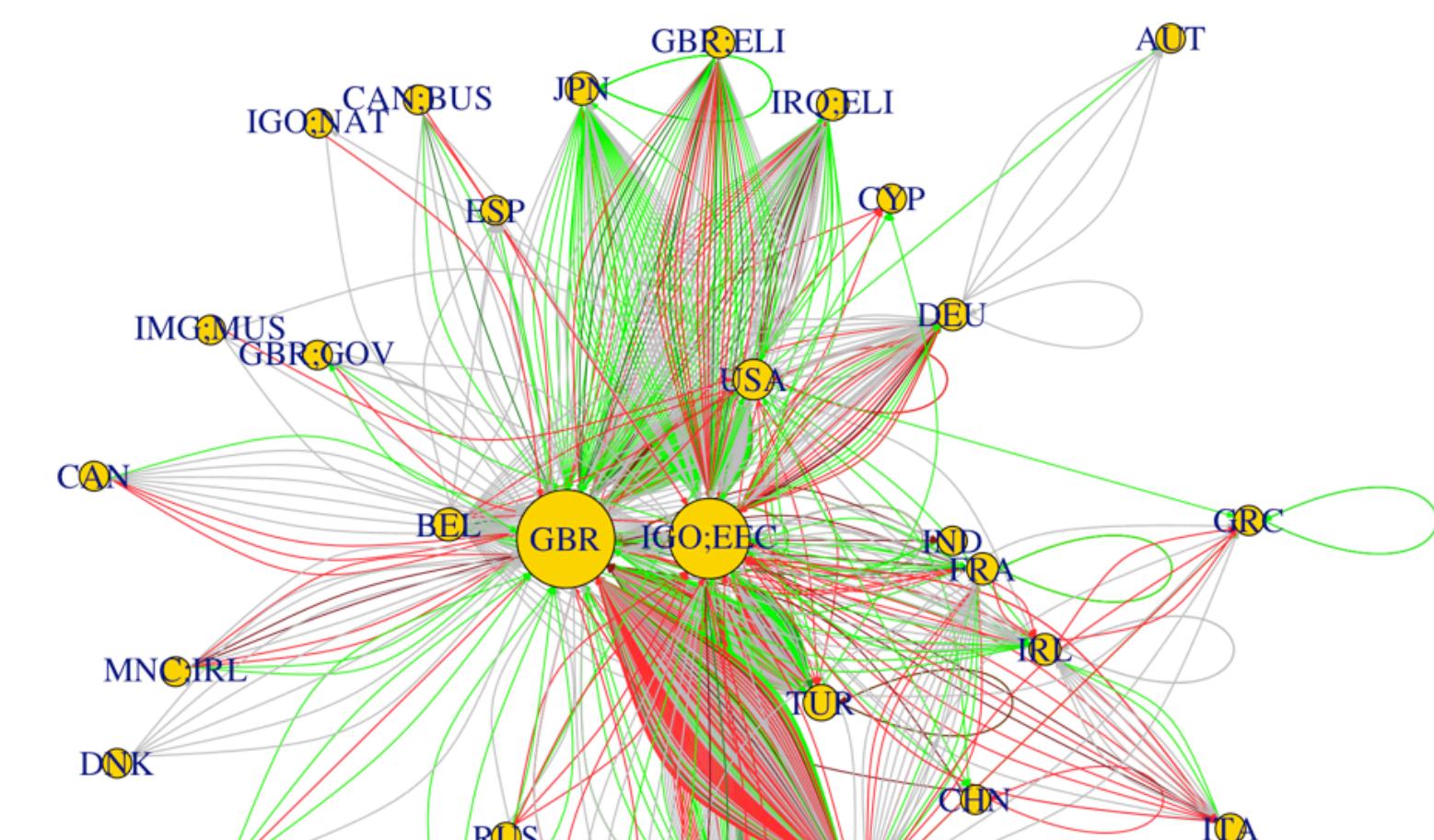
METHODOLOGY

- Self-collect tweets subject to keywords and manually bin them into five categories: EU Referendum in the UK (Brexit), Migrant Crisis, NATO, Russia, and “Other” for anything that did not fall into the above.
- Identify tweets linking to news content.
- Run sentiment analysis on the tweet text.
- Run the EL:DIABLO Scraper and PETRARCH (Phoenix data) on the linked news content to extract geopolitical events.
- Perform topic modeling on the event data and link it back to the sentiment of the tweet.



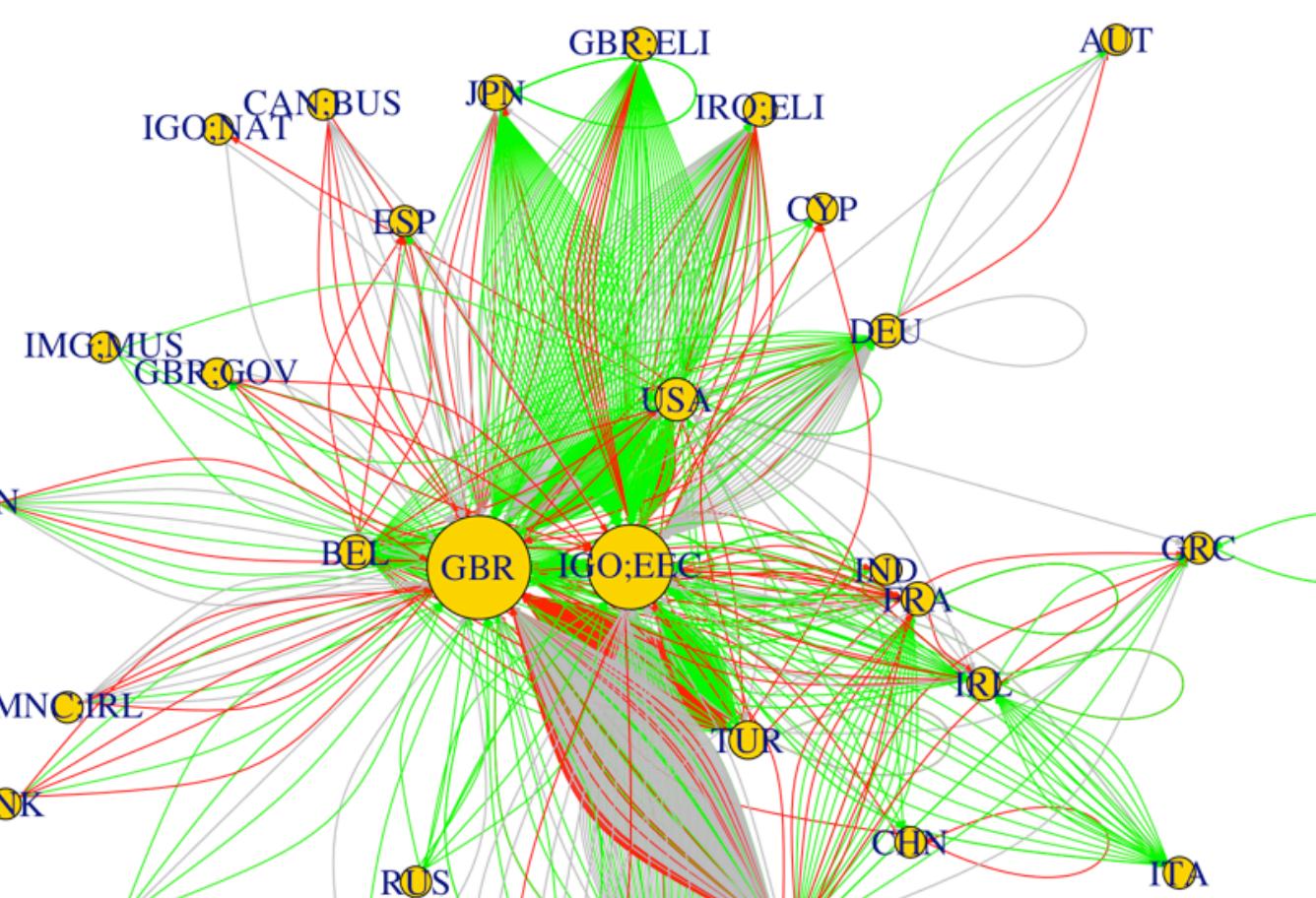
DATA ANALYSIS – From 4055 non-neutral tweets containing links to news content, we generated a graph of political actors for broad, geopolitically-relevant topic categories (note: “Other” category not displayed below). Edges on the left are colored by five sentiment categories from very negative (dark red) to very positive (dark green). Edges on the right are colored by the sign of the Goldstein score: negative (red), zero (gray), positive (green).

SENTIMENT EDGES

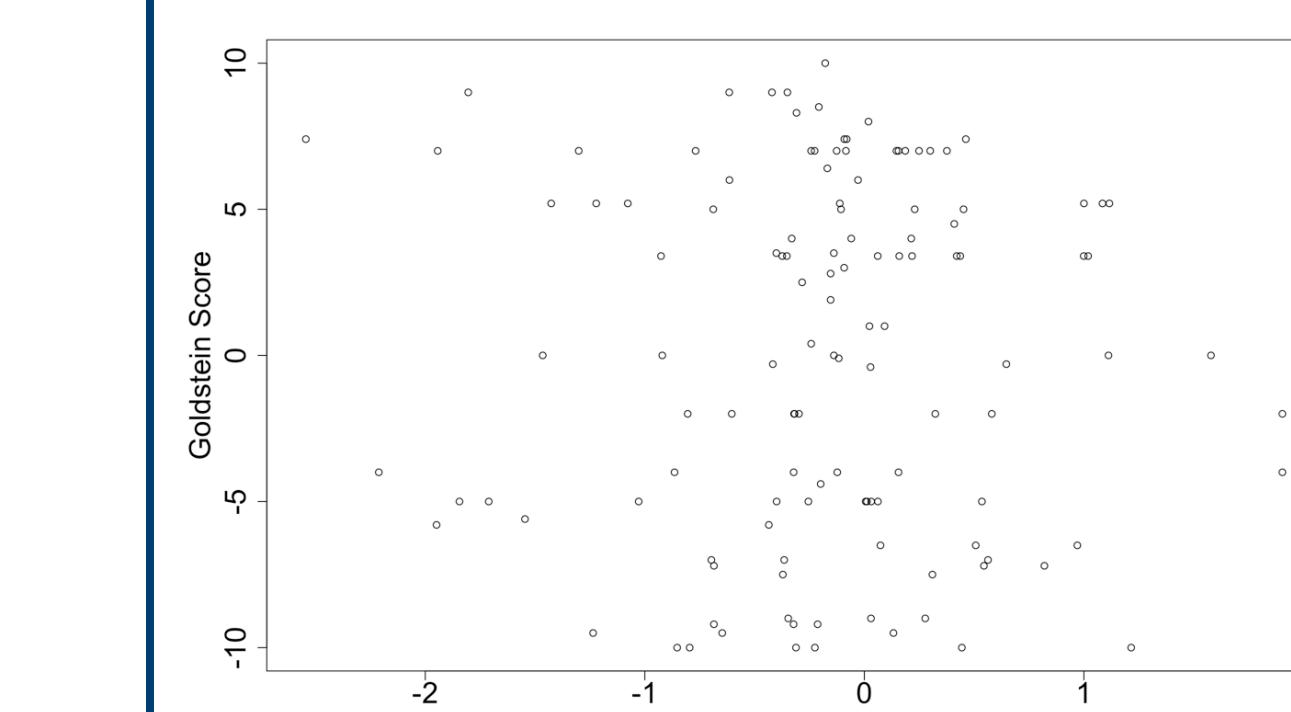


BREXIT

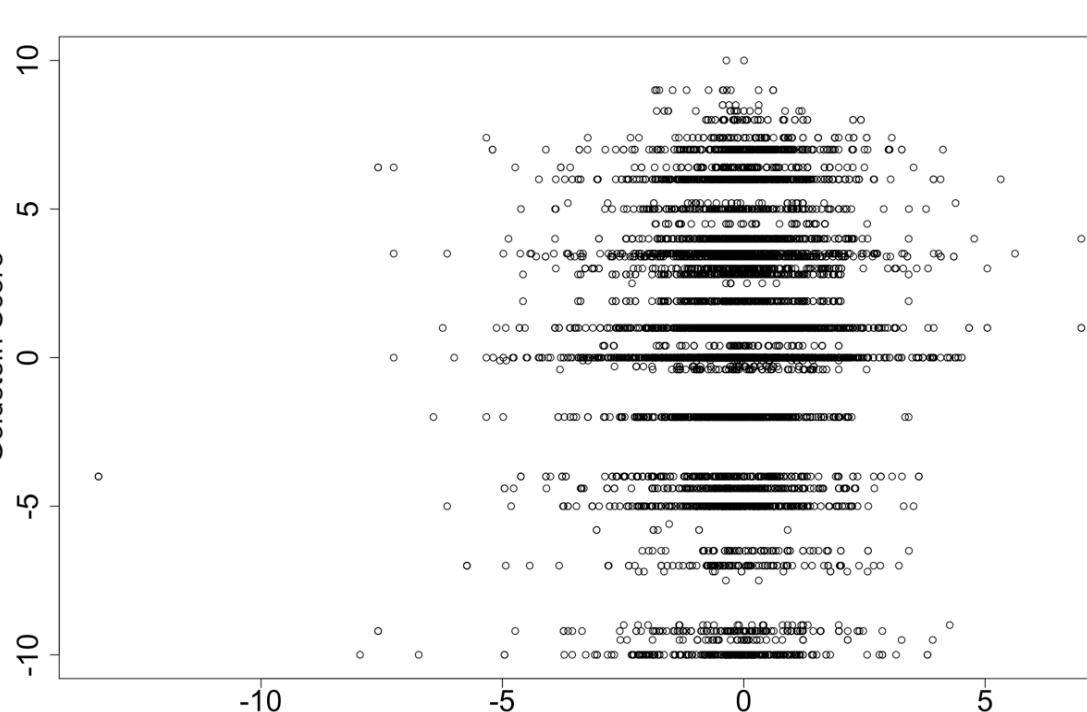
GOLDSTEIN SCORE EDGES



CORRELATING SENTIMENT AND GOLDSTEIN SCORES

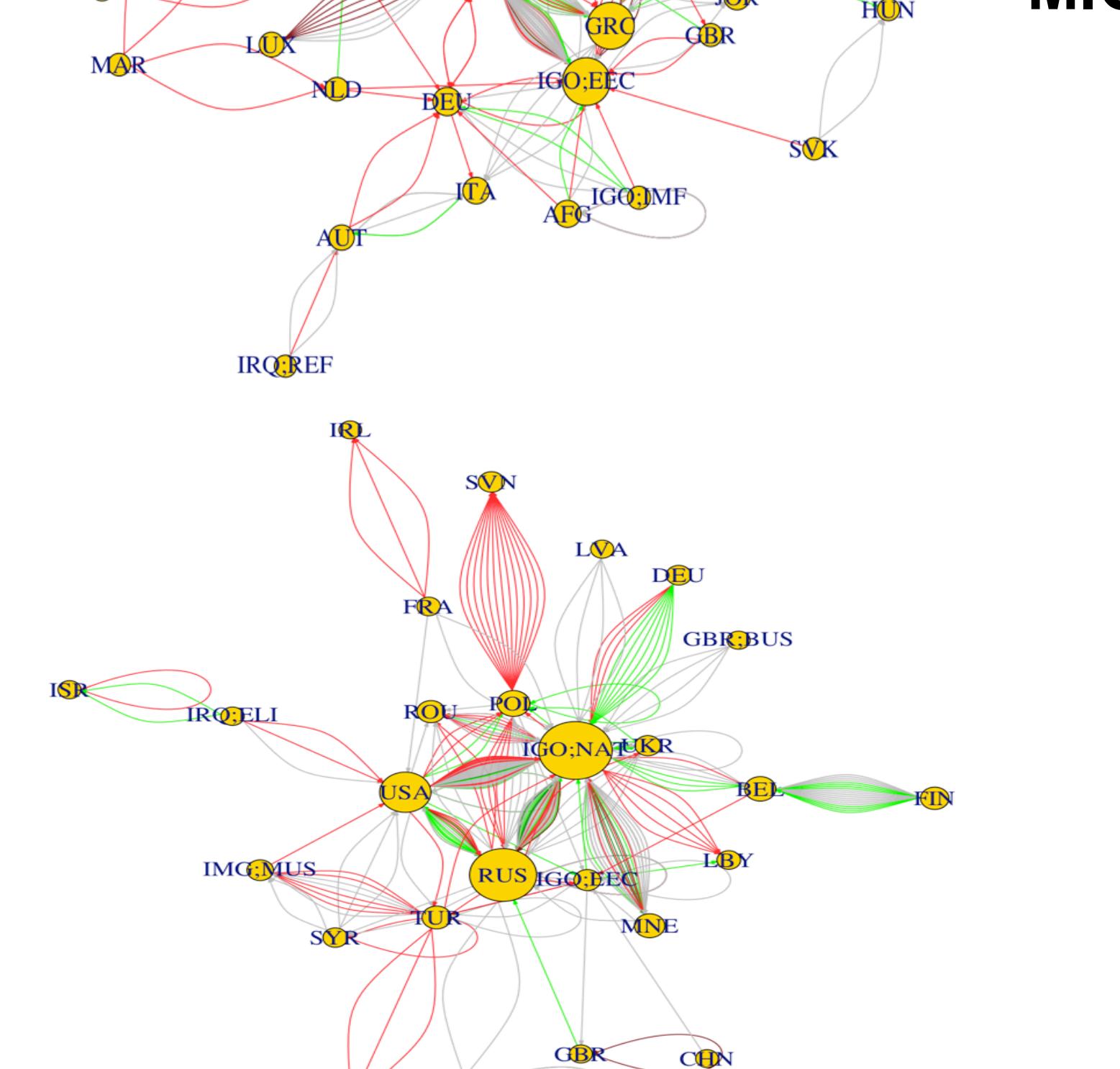


Correlating by Event

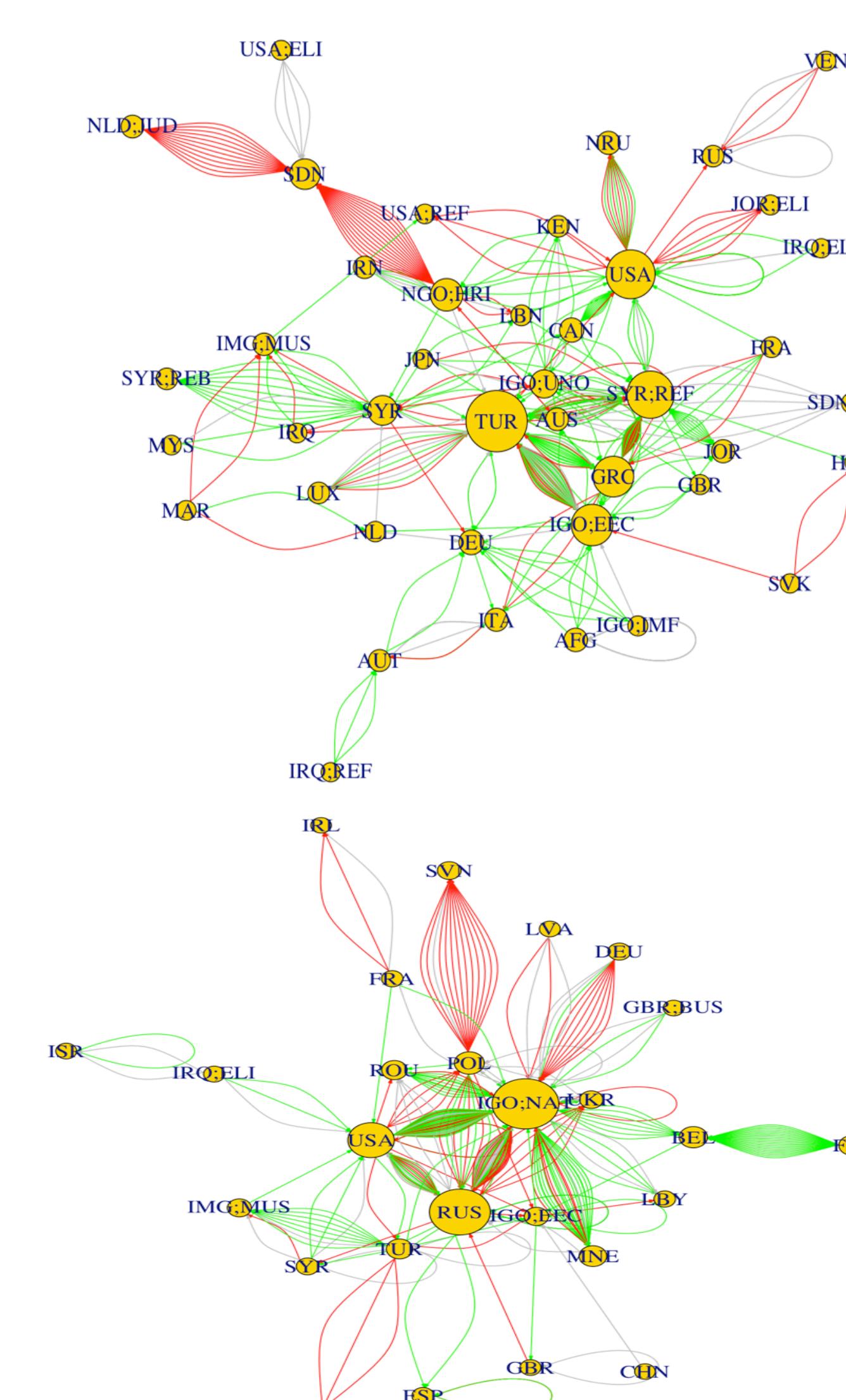


Correlating All Tweets

MIGRANT CRISIS



NATO



RUSSIA

ACCURACY – A supervised decision tree model was used for topic classification based on event information. 5-fold cross validation was performed across the 4055 tweet/event data points. The total accuracy for the model was 82.4%. Precision, recall, and f-measure metrics are provided.

Precision, Recall, and F-Measure			
Class	Precision	Recall	f_1
Brexit	0.88	0.96	0.91
Migrant Crisis	0.75	0.81	0.78
NATO	0.72	0.56	0.63
Russia	0.81	0.84	0.83
Other	0.79	0.62	0.70
Macro-Avg Total	0.79	0.76	0.77

SUMMARY – The length of tweets makes identifying sentiment aspect in Twitter data difficult. However, many geopolitically-relevant tweets expressing opinion have links to news articles with more information about the topic and/or event(s). Event coding can be used to inform public sentiment aspect. This form of sentiment aspect may be used to predict the political actors involved as identified through event coding. Future work will be conducted to also address whether or not relevant unsupervised aspects and corresponding topic graphs may be generated by political actors and event coded information.

REFERENCES

1. C. Musto, G. Semeraro, & M. Polignano. A Comparison of Lexicon-based Approaches for Sentiment Analysis of Microblog Posts. 8th International Workshop on Information Filtering and Retrieval. 2014.
2. M. Ringsquadri & D. Petkovic. Analyzing Political Sentiment on Twitter. AAAI Spring Symposium. 2013.
3. P. Schrot, J. Beiber, & M. Idris. Three's a Charm?: Open Event Data Coding with EL:DIABLO, PETRARCH, and the Open Event Data Alliance. International Studies Association Meeting. 2014.