

Subgraph mining for conflicts and mediation forecasting SBP Data challenge abstract

Conflicts and mediation are two of the most important social process in international relation with broad involvement of domains varied from economy, commerce, political powers and legal environment to human rights, cultural reservation, and refugees. However, the outbursts of conflicts are hard to predict attributed to sophisticated causes, unmanageable triggers and aggregated pressures [1].

Problem

We address at discovering the patterns denoting the conflicts and mediation relationship from history records and predict political events. Specifically, the research questions stressed in the project are:

- Which countries/organizations will most probably have conflicts in a given timeframe?
- What kind of conflicts will they have, and what kind of mediation will be conducted?

What will be the third-party entities involved in the conflicts? What will be the relationship among them? If the conflicts could be predicted, it might be an optimized option for people to arrange the activities in advanced and reduce the probabilities of conditional risks.

A series of conflicts and mediation can be represented as a graph pattern. Syria crisis has been focal since 2011, involving more than 4.5 million people and over hundreds countries. United States and Russia are the two major countries controlling the situation, while the relationship between these two countries is as intense as the crisis. United States has been conducting strikes on Syria since 2014 for annihilating the terrorists, and Russia began to launch strikes in September 2015. The powerful strikes made by the strongest countries in the world successfully suppressed the terrorism and regime. However, the military operation conducted by Russia stimulated the disagreements and the oral conflicts affected other relations between Russia and United States. In February 2016, a peace talk mediated by United Nations, Russia and United State reached an agreement of ceasefire plan. The future moves of Syria crisis greatly depend on the strategies playing by Russia and United States. From Figure1, it is clear to see that the events have formed a graph with multiple interconnected subgraphs. USA-RUS-SYR triangle is the strongest subgraph among them which illustrates significant correlations caused by events.

The influences of syria crisis show significance on business, trade, refugees, human rights violations and cultural heritages. If the tendency could be illustrated by prediction of conflicts and mediation between Russia and United States, it would be helpful for governments, enterprises and non-profit organization to prepare the countermeasure.

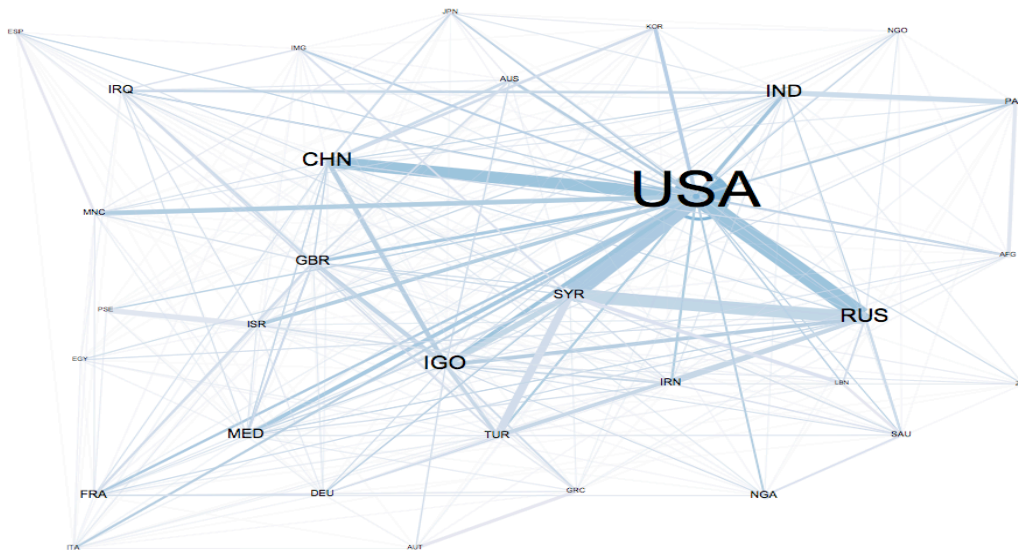


Figure 1. Graphical overview of events dated from Feb 20th to 26th, 2016

Dataset

The dataset we are using is Pheonix, a set of conflicts and mediation news coded by CAMEO. We will also crawl users' tweets from Twitter as a support to evaluate the seriousness of the conflicts.

Related work

Political events forecasting has been a popular topic. There have been scholars working on using statistical model to forecast the civil conflicts and events of interest. In the work[2], Schrodte et.al used Latent Dirichlet allocation for events classification. Work[3] used Bayesian model, random forest and ADABOOST to generate the posterior probabilities given historical data. Shearer[4] was using Hidden Markov Models for fitting dynamic behaviors of conflicts to transition states and using events classification to forecast the states. Most of the works were targeting on combining time series and statistical method to build classifiers and predict the events based on that classifications. However, fewer works have taken the internal forces among conflicts entities and fine-grained classifications into accounts.

We are going to formulate the problem into subgraph mining problem and apply Apriori-like algorithm[5] into implementations. The Apriori-like algorithm is an efficient pattern mining method that being able to discover the frequent single items in the pattern as well as subgraphs isomorphism. It is natural to form the relationship as graphs with entities as nodes and types of conflicts and mediation as edges, and fit them into the algorithms.

Team members

The team members are Yusan Lin, Hui-ju Hung, Tao-yang Fu and Yanjun Gao from Computer Science and Engineering, Penn State University.

Reference

- [1] Levy, Jack S. "The causes of war and the conditions of peace." *Annual Review of Political Science* 1, no. 1 (1998): 139-165.
- [2] Schrodtt, P.A. Forecasting Political Conflict in Asia and the Middle East using Latent Dirichlet Allocation Models *. (2011).
- [3] Arva, B., Beieler, J., Fisher, B., et al. Improving Forecasts of International Events of Interest . (2013).
- [4] Shearer, M.R. Forecasting Israeli-Palestinian Conflict with Hidden Markov Models. 0-27.
- [5] Huan, J., Wang, W., Prins, J., and Yang, J. Spin: mining maximal frequent subgraphs from graph databases. *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1 (2004), 581-586.