ECS 163 Writeup

Team: Steve Avery, Gregory Guterman, Xuening Wang

To start our visualization, open visualization.html in your browser. At the top you will see a link to the alternative Timeline.js view of the timeline which is for reading the articles (does not support filtering). Next you will see the network graph. You can drag it around, and click on edges to show articles related to those edges in the timeline view at the bottom. The timeline view at the bottom can be filtered by news source. Some news sources are on the good guys’ side and some on the bad guys’ side so you can choose which you want to read. When you click an edge in the graph view, the relevant articles will show up in the timeline view.

Abstract

In this visualization project, we are trying to use visual tools to help figuring out the mystery of a fictional story – what happened on “GASTech” company on the company’s 20th anniversary ceremony and what may account for the staff’s missing – based on the data set from the Visual Analytics Community Challenge 2014. We incorporate two views of the visualization based on different level of details and enable the function of panning and zooming, interactively showing and hiding data, and filtering out some data to allow users generate both overview of the two involved organizations - Protectors of Kronos(POK) and GASTech, and the detailed news report narrated along a timeline.

About the visualization

Our visualization contains two views:

* A Force-directed network graph showing the interactions and connections among people
* A zoomable and panning timeline of news articles to navigate the fact details

In the network graph, we use points for people and use color to discriminate departmental or organizational information. Bigger points show the leader of the group. The darkness of links reflects the closeness between people, the darker links the closer. We give the options to users to choose whether to base on email connections information or on whether or not two names are mentioned in the same article. Colors of edges stands for types of connections: dark grey – assistance, light grey – subordinate, red – linked by events, black – email connection.

For the article timeline, we care about the historical story since 1995 and given the option to filter out the articles by their source, users can focus on only one or more presses that they believe in. Allowing for the extensity of the files, we have a categorical timeline to show only the titles corresponding to the selected resource(s) plus a more detailed timeline similar to a library of full contents. Both are zoomable to show in different scales.

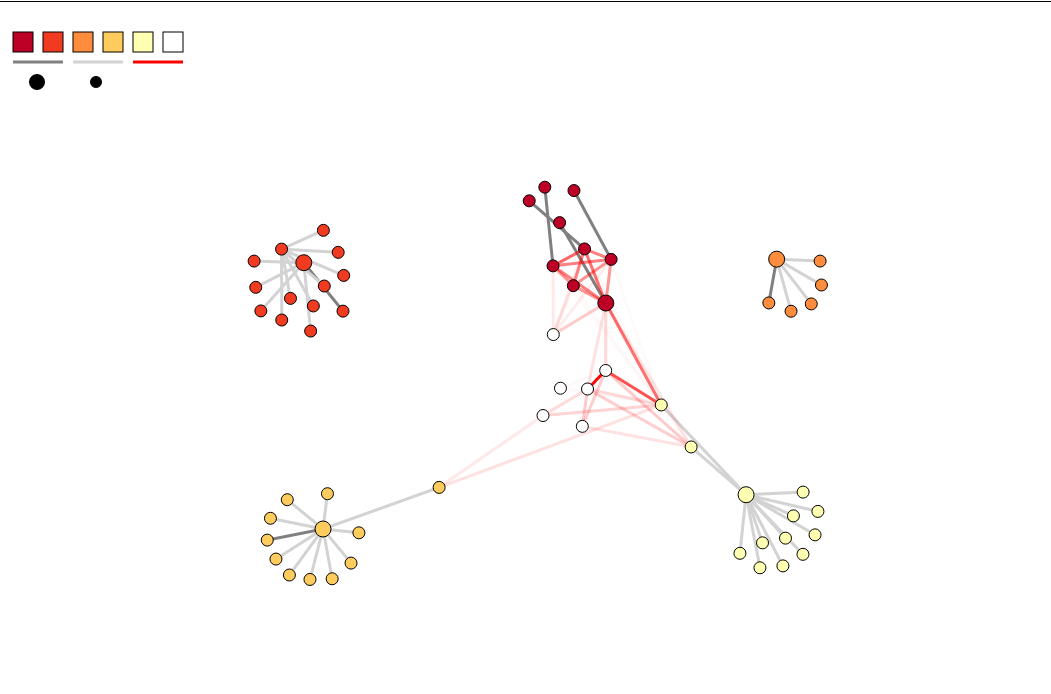


Fig. 1 Network graph

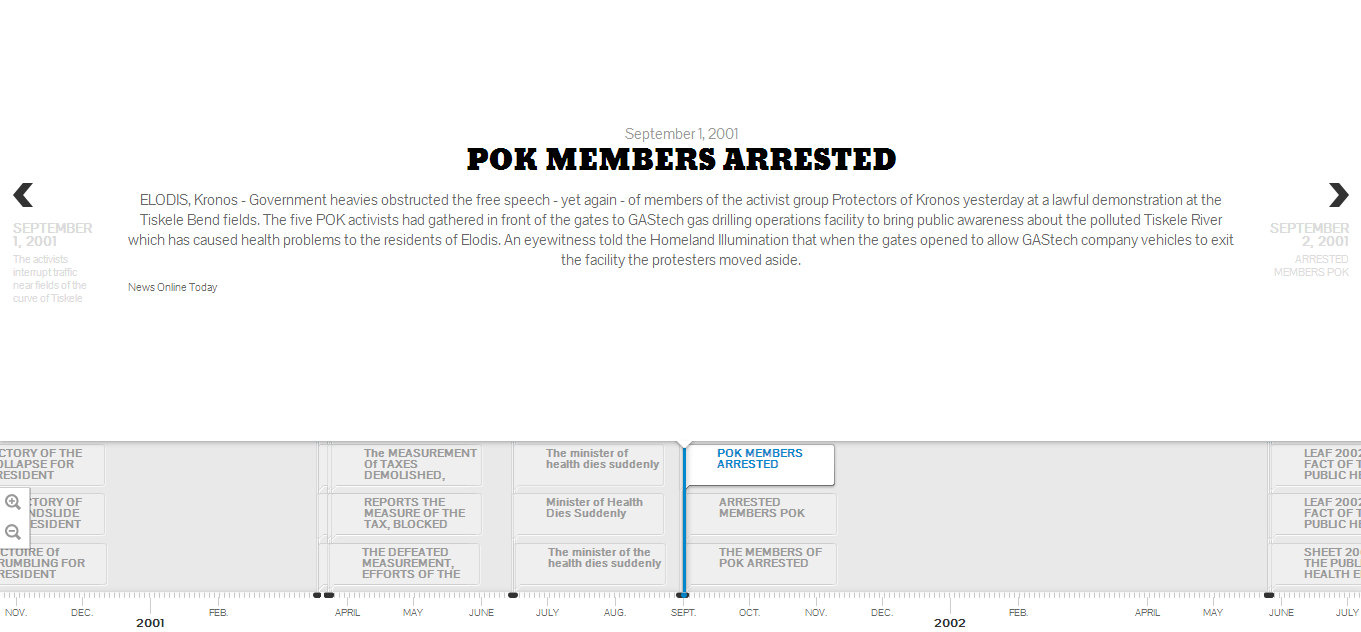


Fig. 2 News Timeline

Background

This fictional story mainly involves two organizations – a Tethys-based oil drilling company GASTech declared to have 14 employees disappeared on the event to celebrate their 20th anniversary as well as successful IPO on January 20, 2014, and a radical organization POK which is thought to be responsible for the kidnapping. Provided with information as follows:

* A map of Kronos.
* The GAStech company hierarchy, employee records and resumes/ short biographies.
* Two weeks of GAStech internal email headers.
* News articles.
* A master's thesis about the evolution and state of the Protectors of Kronos

We are required to use visualizations to better analyze and explore the structure of POK and the relationship between the two organizations so that to uncover the core as well as the extended conspiracy network.

Dataset analysis and preprocessing

As a typical criminal network, we focus on the two weeks email headers(1170 records stored in a csv file) and the news articles(844 files) to measure the relationship and structure of the network. We filter out meaningless Emails by subjects’ keywords such as “Haha” or “Cute!”, while for the articles, the challenge comes from the amount of plaintext files and the discretion about the reliability of different sources. We found a powerful Natural Language Processing(NLP) API developed by The Stanford NLP Group. This tool helps us to identify the entities (person, organization, location, etc.) mentioned in each article.

Some findings

In the pre-analysis stage, we browsed through part of the dataset and found some suspicious clues. For example, some staffs share the same last name with members of POK which may be a sign of relative affiliation. The names of people (eg. Sten Sanjorge Jr. the president, Vasco-Pais Willem the Environmental Adviser, etc) and usage of words from particular resources are Italian-like but neither Kronos nor Tethys’s official language mentioned Italian, what is more, “coincidentally”, some reports have it that one of the two planes taking off from Tethys Airport is heading to Rome, Italy. Besides, all the emails from the so-called Mr. Environmental Advisor are talking about maximum profit but none refers to the environment, president of GASTech sent files to the truck driver… By visualizing the communications and event-based connections( mentioned in the same news article), our doubts have been proven or denied by solid evidence. Here are what we concludes:

From the network graph, we find the executives have very close connections with each other. Besides that, some unexpected inter-departmental communications are noticeable such as the link between Sten Sanjorge and the truck driver Henk Mies. The police interrogated Edvis Vann and released him while no article has mentioned there is an interrogation about another Vann’s – Isia Vann, who also has the POK related family name and plays the role of perimeter control. However, based on our visualization, the closely connection between Isia Vann and non-GASTech member, especially POK member shows up by the dark red link! Another interesting pattern, the out-of-GASTech connections involves mainly with three important groups – the executive, the security group and the truck drivers.

Also, by navigation the source of articles, we find some good/reliable source like International Times and Homeland as well as bad source like the Aliba Post which seems to try to cover the fact and cause distraction.