Advanced Visual Data Analysis

Ahsan Iqbal, Magdalena Proszewska

1 Overview

Mini-Challenge 1 from VAST 2014 is a text-based data collection concerning the kidnapping of the GASTech employees by members of the social movement group Protectors of Kronos (POK). Here, we discuss techniques used for data analysis and results that we obtained for that challenge. Additionally, we briefly describe history of POK in order to give more context for the kidnapping.

2 Dataset

Dataset of this challenge contains

- a map of Kronos,
- a chart describing GAStech organization, in PDF format,
- a spreadsheet of GAStech employee records, in Microsoft Excel format,
- email headers from two weeks of internal GAStech company email, in CSV format,
- resumes and short biographies of 35 (not all) GAStech employees, in Microsoft Word format,
- historical reports and descriptions of Kronos and Tethys, in Microsoft Word format,
- 845 historical news reports from multiple domestic and translated foreign sources, in text file format.

3 Context

In 1997 citizens began to be concerned about an abnormal increase in the occurrence of illnesses. They believed that it is due to the water condition in a nearby river. For that reason, in 1998, 7 citizens formed a health organization called Protectors of Kronos (POK), whose main focus was water purification. Later that year, in August, 10 year old girl, Juliana Vann, died due to illness associated with benzene toxicity, which resulted in increase of popularity of POK.

The government, except minister of health, did not support POK's request and after sudden death of that minister in 2001, POK and government were not able to find the middle ground. After Elian Karel bacame POK leader in 2001, organization became more aggressive and openly blamed the government and GAStech (a company that drills gas) for water pollution. On June 19th 2009, Karel was arrested and later was murdered under suspicious circumstances. Silvia Marek became POK leader. On January 20th 2014 GASTech employees have gone missing.

4 Questions

We managed to answer all the questions in this challenge.

4.1 Protectors of Kronos

Who are the leaders? In order to find leaders, we searched for sentences that contain keyword "leader" in both historical documents and articles. From historical documents, we learned that the leader of initial group was Henk Bodrogi and the second leader was Elian Karel. Based on articles, we find out that Elian Karel was murderd on June 19th 2009 and Silvia Marek took the leadership position after him. Example of article that contains a keyword and allowed us to answer that question is shown in figure 1.

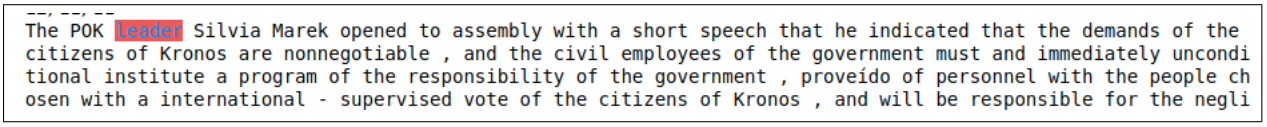

Figure 1: Article with keyword "leader".

Who is part of the extended network? We started by finding all names in history documents (two consecutive words started with capital letters) and sentences that they occur in. With that approach we found

- Family of Juliana Vann: Mandor Vann, Isia Vann,
- Cesare Nespola - minister of health, died in 2001,
- Lorenzo Di Stefano – professor of environmental science at the University of Abila,

which support POK, hence can be consider as the extended network.

We found more interesting results when we were looking at most often used words that are written in capital letters. The idea is that words written in capital letters are either names of organizations etc. or words used in articles' titles. We ended up finding keywords "APA" and "MDMC". From articles that contain these keywords, we learned that MDMC is a drug and APA stands for "The army of people of Asterian" and is an organization associated with drugs trafficking. Besides, while we searched for links between POK and GAStech, we found exchange of emails that contain word "kronos". The exchange was between 6 people: Isia Vann, Ruscella Mies Haber, Inga Ferro, Loreto Bodrogi, Minke Mies, Hennie Osvaldo.

Their emails mention "defenders of kronos", which supposedly stands for POK, and ARISE, which based on information found in articles, is APA's website. Graph that represents exchange of emails and list of emails' headers are shown in figure 2, article that mentions ARISE is shown in figure 3, article from figure 4 proves possible connection between APA and POK. We conclude that APA and these 6 employees from GAStech are part of POK's extended network.

Where are the potential connections between the POK and GAStech? Based on results obtained for previous question, we continue with more analysis regarding found 6 GAStech employees that are somehow connected to POK. Vann, Mies, Bodrogi and Osvaldo are surnames that occur in POK's historical documents, hence we conclude that some of them might be family of official POK members. Besides, from employees records we know that all except Haber work in security, also Bodrogi and Ferro were recently hired. All of that, leads to the conclusion that

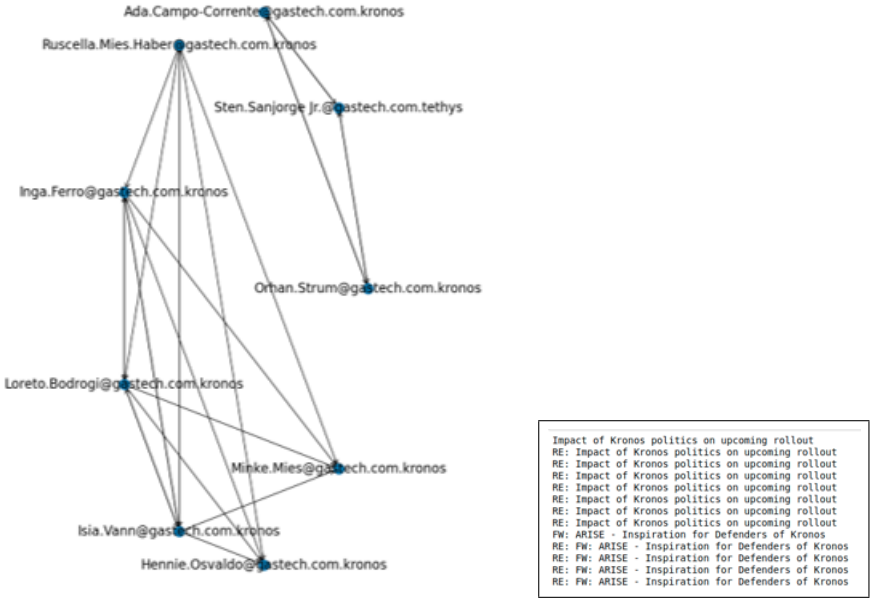

Figure 2: Exchange of emails with keyword "kronos".

13/10/22
The last years their mark has had APA by Internet sites and publication of professional , hoogstaand illustrate
d magazine to build called Arise .

Figure 3: Article containing keyword "ARISE".

articles/389.txt
05/15/11
The concerns over the increase of drug related arrests are two - fold : first , violent activist groups such as
the Protectors of Kronos could be using drug sales as a means to fund their activities ;
articles/576.txt

Figure 4: GASTech and POK connection

those people are connected to POK and contributed to the kidnapping. 7th possible suspect can be Edvard Vann, which is also part of Vann Family, also works in security and was also recently hired.

How has the group structure and organization changed over time? Analysis of POK's structure is based on historical documents. By looking at sentences that contain datas, we found out how the structure changed over time. Results are shown in figure 5. Moreover, we include results from previous questions and create a chart shown in figure 6 that shows current structure of POK.

Additionally, we look into how POK was seen by others based on words that occur in articles that mention organization's name. We learn that before 2009 POK was associated with health and positive activism and after 2009, it became associated with terms such as arrest and violence.

Figure 5: Change of POK's structure over time.

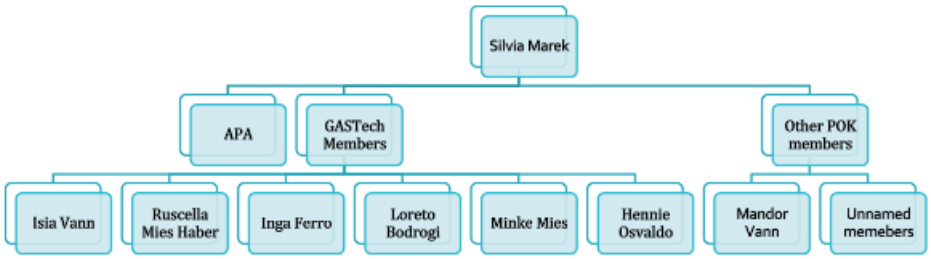

Figure 6: Current structure of POK organization.

4.2 Timeline of events of January 20-21, 2014

After finding and understanding the structure of POK organization and its relation with GASTech our next task was to expose the timeline of events of the 20^{th} and 21^{st} of January 2014. For that we had at our disposition 845 news articles from locale and foreign sources. We had quite a simple strategy which consisting in a string matching process through all the articles. We looked for "20 January 2020", "20/01/2014", "01/20/2014" and similar keywords. It appeared that only "20 January 2014" was useful since all articles about the events had the date in this format in their headlines. In the figure 7 we can see an example of the results we get with this method. It appears that the articles are not sorted in a chronological way and that they were updated through the day. Also most of the articles start their news with a time code referring to the moment in the day. Last thing to notice before processing our list of articles about the events, is that there are a lot of redundancies between the articles since they are from different sources covering the same event. The next step we did consisted in sorting the articles in a chronological order and keeping only the lines starting with a new time code.

Now that we have a preprocessed timeline of the events, we will look at the most used non trivial words that appears. We get graphs like figure 8 in which we can see that from 09:26 and 11:30 the most used words in the news was fire. With this method we can select few keywords that seems representative of the events and see their evolution through out the day. The results we get with this method are shown in figure 9. One conclusion that we can get with this is that at the beginning of the day the first hypothesis was a fire at GASTeach headquarters abut it rapidly vanish in the afternoon to be replaced by a theory about GASTeach executive feeding the country by jet and that finally it appeared that employees were missing.

Finally after having the chronology of the events and how they were covered by the media, we looked through our preprocessed events (which was a text file of approximately 40-50 lines) to cut what we judged to be useless and construct the timeline of figure 10

```
1532 MODERNIZATION - an employee of Kronos - of GAStech that leaves the construction of the centers who refuse
Update 1:00 PM:  Police have come and gone from the GAStech building, but up to this point no GAStech employee
1018 - A fire alarm has gone off at GAStech headquarters.  People are evacuating the building, and so far no s
Modernization, 3:05 PM:  A police P2o has left the scene, but others remain to the external part to fix the co
1025 - Up to now no test of smoke or fire has been marked.  The fire unit has to arrive hardly. 20 gen. 2014
1744 - The means relationships are coming in how much a sure number of employ you of GAStech have been rob to
```

Figure 7: Examples of articles

4.3 Explanations

4 out of 10 missing employees are GASTech executives, the other 6 are members of POK working in security that we identified. Executives might have been kidnapped by POK or APA. On one hand POK could kidnap them in order to blackmail GASTech and force the company to stop polluting the water. On the other hand, it could have been a operation of APA in order to obtain large amount of money from GASTech or GASTech's CEO, who is the only non-kidnapped executive.

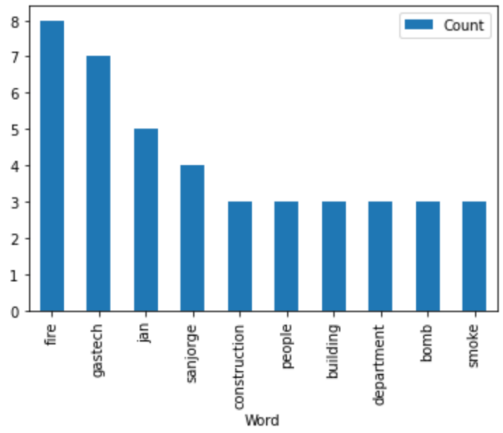

Figure 8: Most used words between 09:26 and 11:30

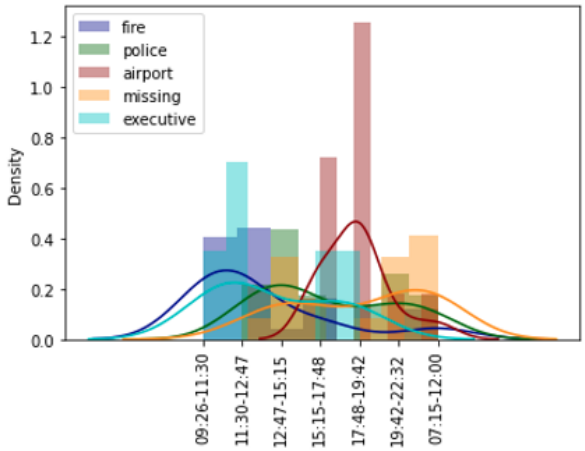

Figure 9: Use of keywords over time

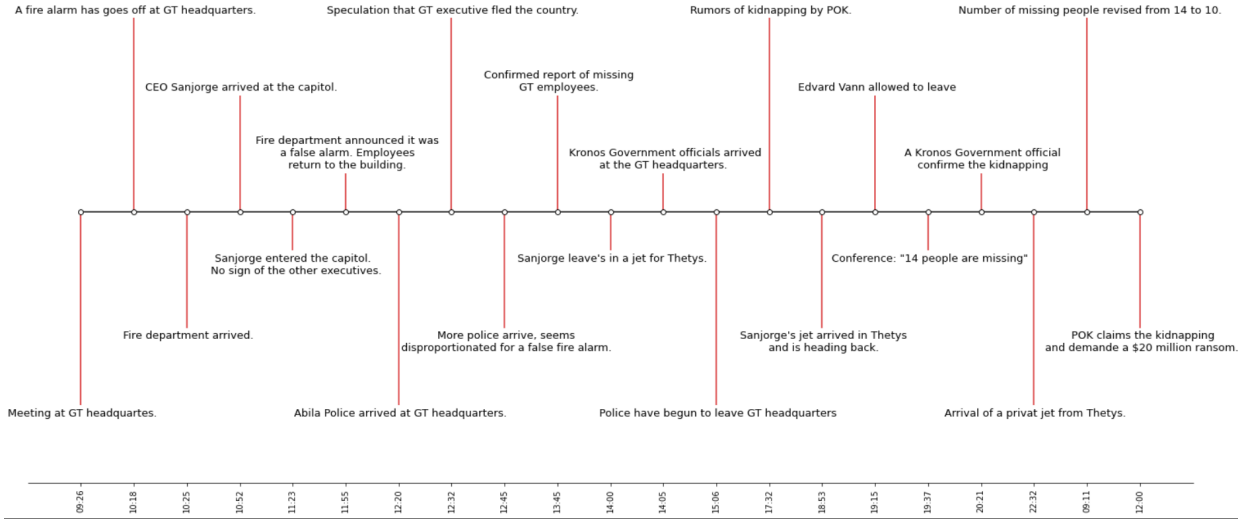

Figure 10: Timeline

Another hypothesis could be that this was arranged by GASTech's executive so that they could run away from responsibility, but that is the lest probable theory considering that POK is claiming the kidnapping. Nonetheless it is still possible that POK saw an opportunity and decided to claim the action to put pressure. The fact that CEO got on a plane and seemingly run away from the kidnappers seems suspicious and could imply his participation in the events, but we think that it is more likely a coincidence since Kronos is an island and it seems that CEO of GASTech often uses planes for travels.

5 Future work

One way to improve our results could have been to search for lexical fields instead of precise keywords. For example the fire threat in the new report was connected to a potential bomb threat but that does not appear in our graphs. But even with more precision we do not think that it could led us in another hypothesis or to a contradictory information compared to what we have now.