# GateViz - Interactively analyzing social media trends

In December 2019, the intendant of Germany's largest public broadcasting institution apologized in a special broadcast. The notion was, a satiric song had insulted the broad mass of their viewers. Analysts showed only a few thousand right-wing twitter accounts carried the outcry [1]. There are already a variety of tools that could have helped to analyze the situation on the fly. But they are not designed to be used by a layperson with no statistical and mathematical knowledge. The goal of GateViz is to create a tool usable by data science laypersons to asses trends and their evolution in social media.

Backend

With Vizdom an interactive data exploration system is already at hand. Although the focus of GateViz does not lie on data pipelines, the underlying optimization techniques are still helpful [2]. The analysis task will mostly resolve around cluster detection and labeling of accounts or posts. Opposed to the original design, a new model training for the labeling tasks should not be necessary each time. Meaning, pre-trained models for tasks like sentiment analysis are required. These additional functionalities pose the main technical challenge

GateViz
––––––––––––
Vizdom/Tupleware
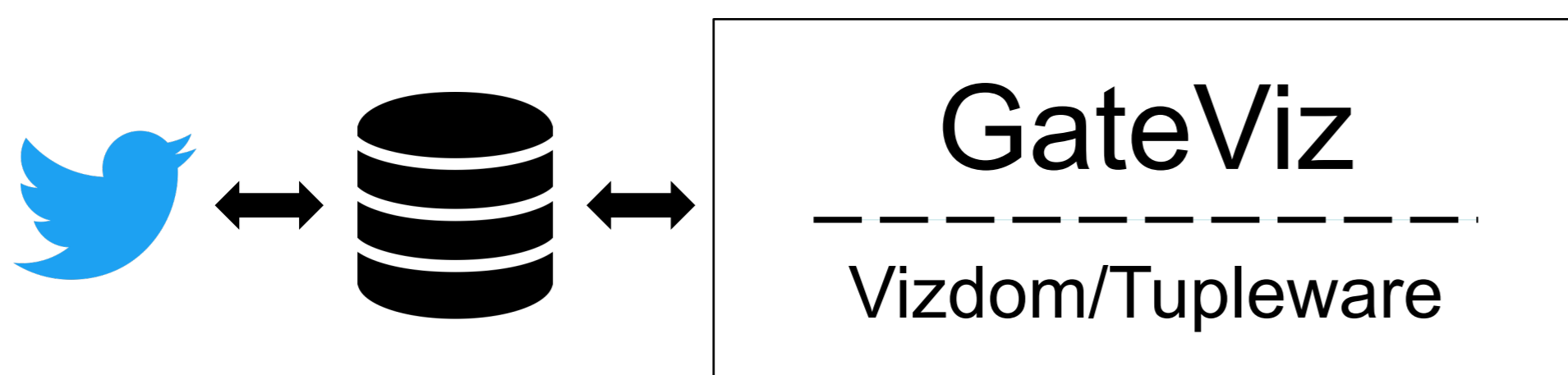
Data explorative tasks do not need a real-time stream of the data. Posts get pulled through search APIs into a local database.

Client

Different from the case of Vizdom, the schema is similar among different datasets. So the visualization is vital to get an understanding of a dataset. According to studies, a graph like visualization is most intuitively for social media data. For smaller sections of the network, matrix visualizations can be used [3]. Filtering happens through text-based search or interactive manipulation of the graph. A time slider allows the exploration of temporal development.
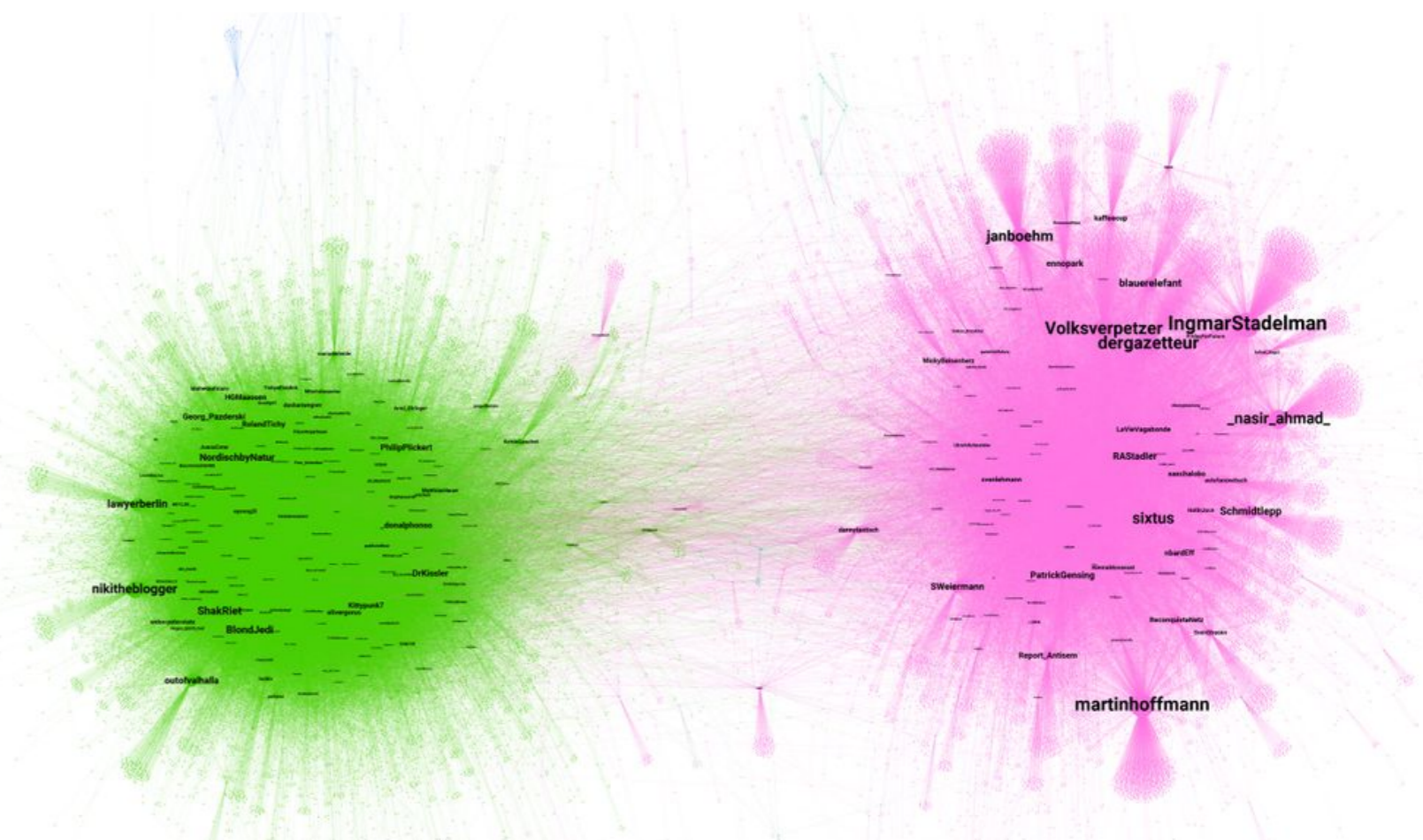
Graph visualization algorithms like ForceAtlas2 can already cluster and structure the data on a visual level [4]. They are designed to be run locally on the client and produce continuous results in the form of force-directed graphs.

Evaluation

A hard measurement criterium is the interactive threshold of 500 ms [5]. Apart from evaluating the technical fulfillment of this criterium and the fulfillment of the tasks in general, an evaluation is necessary on how effective laypersons can work with this tool. Based on this evaluation, the feature set can be refined and or extended for future work.

Ethical discussion

While the usage by journalist serves the society as a whole, the automatic labeling of content can be misused. Besides form validating the legal ground, further investigations are necessary. How can this tool be protected against malicious usages?

Nils Thamm

Student IT-Systems-Engineering (Master)
Practical Data Engineering WS 19/20
Hasso Plattner Institute, Potsdam, Germany

E-Mail: nils.thamm@student.hpi.de

[1] Horn, E. & Peteranderl, S. (2019, December 30) Die Empörungsmaschine läuft heiß. Spiegel. https://www.spiegel.de
[2] Crotty, A., Galakatos, A., Zgraggen, E., Binnig, C., & Kraska, T. (2015). Vizdom: interactive analytics through pen and touch.
[3] Henry Riche, Nathalie & Fekete, Jean-Daniel & McGuffin, Michael. (2007). NodeTrix: a Hybrid Visualization of Social Networks.
[4] Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.
[5] Z. Liu and J. Heer. The Effects of Interactive Latency on Exploratory Visual Analysis.
Image credits: Freepik, Smashicons, Luca Hammer

HPI Hasso Plattner Institut