

Topogram: a web-based toolkit for spatiotemporal network analysis of online activities

Clément Renaud
Telecom ParisTech
UMR LTCI

Topogram is a web-based and open-source toolkit to extract and visualize social, semantic and spatio-temporal dynamics within large sets of data. The purpose of this tool is to bring elements of contexts while studying and exploring large sets of text data that describe online activities, understood as online enunciation acts. Topogram link together different dimensions of the data : words (lexical analysis), relationships (networks), time (changes and evolution) and space (geographic mapping). It has been created during a 3-years study about Internet memes spread on the Chinese online social network Sina Weibo (Renaud, 2014).

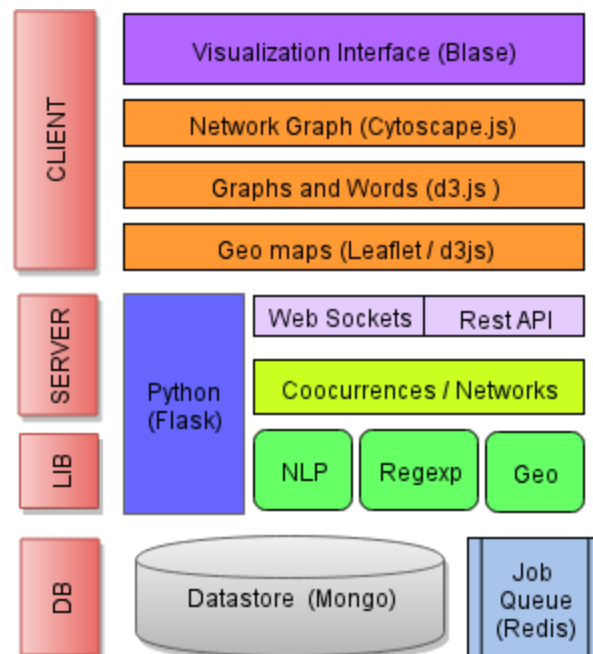


Fig. 1 - Topogram : General structure of the application

Topogram originates in an effort to provide a data mining¹ and visualization² framework that will allow researchers to minimize routine tasks and focus on the exploration of relationships

¹ Topogram analysis library, code available at <https://github.com/topogram/topogram>

² Topogram visualization client, code available at <https://github.com/topogram/topogram-client>

between different dimensions in data. Users can import data containing text, dates and locations by using Topogram’s generic importer (CSV files) or custom scripts using the public API. The Topogram server will process the data to extract a network of words by segmenting sentences (with pre-processing for asiatic languages) and building networks of cooccurrences by analyzing rows (ex. cooccurrences of words in a single message for online social networks). A network of relationships and citations (defined by regular expressions) can also be added. Results are indexed to allow full-text search by terms, time and places. In Topogram, the text analysis is based on a plugin system to easily allow multiples languages to be used (Python NLTK). Currently English, French and Chinese language are supported.

Data preparation Select the correct format for your data.

Describe this dataset

Which column contains the text to process ?

Text Column 微博内容

Which language is the text ?

Language zh

Which columns contains the source / author(s) ?

Author Column 作者ID

Which column contains time information ?

Time Column 发布时间

What is the format of the date ? (using `strftime` pattern)

Time Pattern %Y-%m-%d %H:%M

ex. 2015-03-31 00:59 | result: 2016-01-31 11:59

Additional Columns (optional)

The columns selected below will be stored but not processed.

微博内容
转发数
评论数
原微博ID
原微博URL

Citations & patterns

Select the citation patterns from existing or add a pattern.

Update Description

Add patterns

Pattern Citation format...

Regular expressions to match citations in your text (formatting for Python).

Sample from row 1 / 10 << prev | next >>

Title Citation name...

A short description of your citation patterns.

Save Pattern

Fig 2 - Screenshot of Topogram data description process after import

Once information have been extracted from the data, the visualization interface allow a browsing with different views (accessible in a web browser): a network of words based on lexical analysis (word cooccurrences), a representation of time distribution in the data (interactive bar graph), a geographical map of the provided locations and potentially a network graph of other entities extracted from the data (ex. citations graph in social networks). Data can be browsed through any of those different dimensions : selection of a specific timeframe or geographic zone, lookup of specific words and terms, selection of parts of the graph, etc. Each selection criteria will update other views of the data accordingly, allowing an easy exploration of the data. Figures

and results can be exported easily as images (png or svg), subsets of data (csv) or web-view by sharing directly a URL to the visualisation. Topogram interface is designed to minimize user information overload while studying those. The web interface also allows for real-time edition and annotation of a single document by multiple users, making easy for teams to work together on the analysis of large documents.

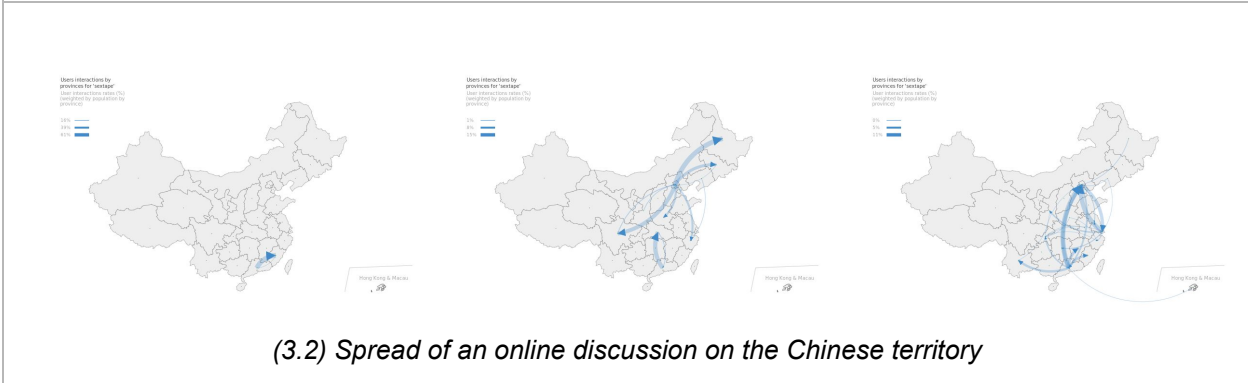
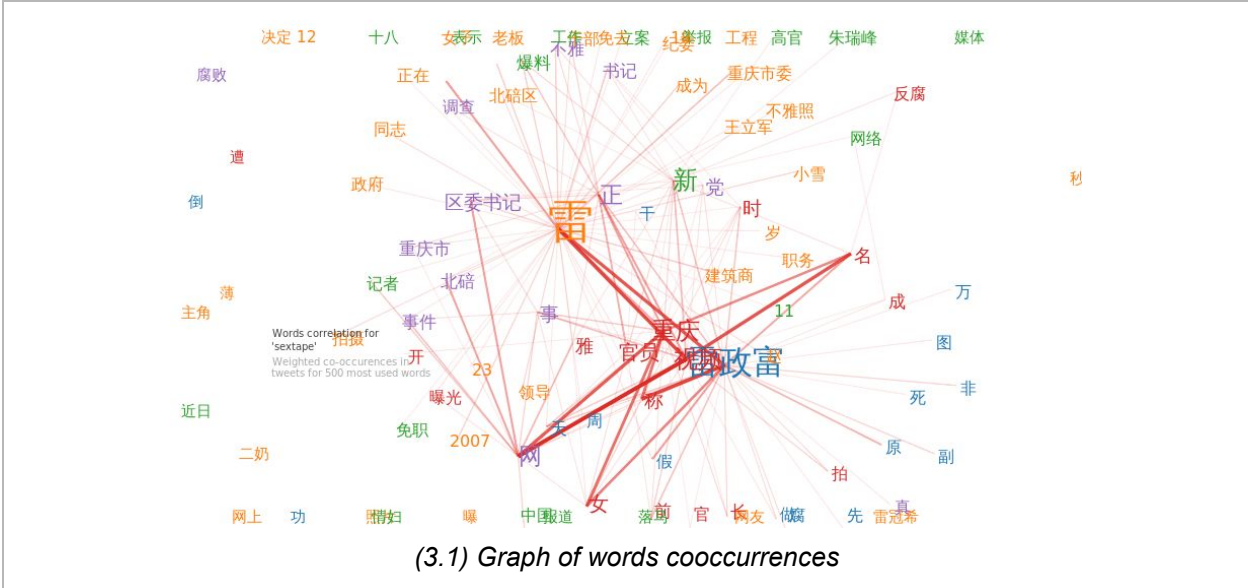


Fig.3 - Figures made with Topogram- the spread of an Internet meme on Sina Weibo (Renaud, 2014)