

RJ Duran
CCS120 (W2012) - Symmetry and Aesthetics in Contemporary Physics
Final Project
3/16/12

Data Visualization of Spacetime

For my final project I created a data visualization application in Processing using data from the ArXiv e-Print scientific paper database. The data set consists of 10,000 titles from recently submitted unpublished papers using the word “spacetime” in their titles. From this I was able to generate a 3D tree representation to explore the data and connections between common words used in association.

The data for each title is parsed using a partial FP-Tree algorithm into separate words based on a minimum support value. This value basically acts as a threshold for the minimum number of times a word must occur in the title to be of interest. Once the data is parsed it is drawn as a tree structure in 3D space and becomes navigable by a user.

Details

This is the query used to request data from the ArXiv API (<http://arxiv.org/help/api/index>). This query basically says, search for all papers with the word spacetime in their title and give me 0 to 10000 items. The results are returned in pure XML and have to be combed to extract the title from in between the <title> tags;

[http://export.arxiv.org/api/query?
search_query=all:spacetime&start=0&max_results=10000](http://export.arxiv.org/api/query?search_query=all:spacetime&start=0&max_results=10000)

The resulting data looks like a string of words, which are then saved to a text file.

```
ArXiv Query: search_query=all:spacetime&id_list=&start=0&max_results=10000
Finitary Spacetime Sheaves
Lorentzian spacetimes with constant curvature invariants in four dimensions
Globally Hyperbolic Flat Spacetimes
Energy-momentum non-conservation on noncommutative spacetime and the
existence of infinite spacetime dimension
Boundaries on Spacetimes: An Outline
...
```

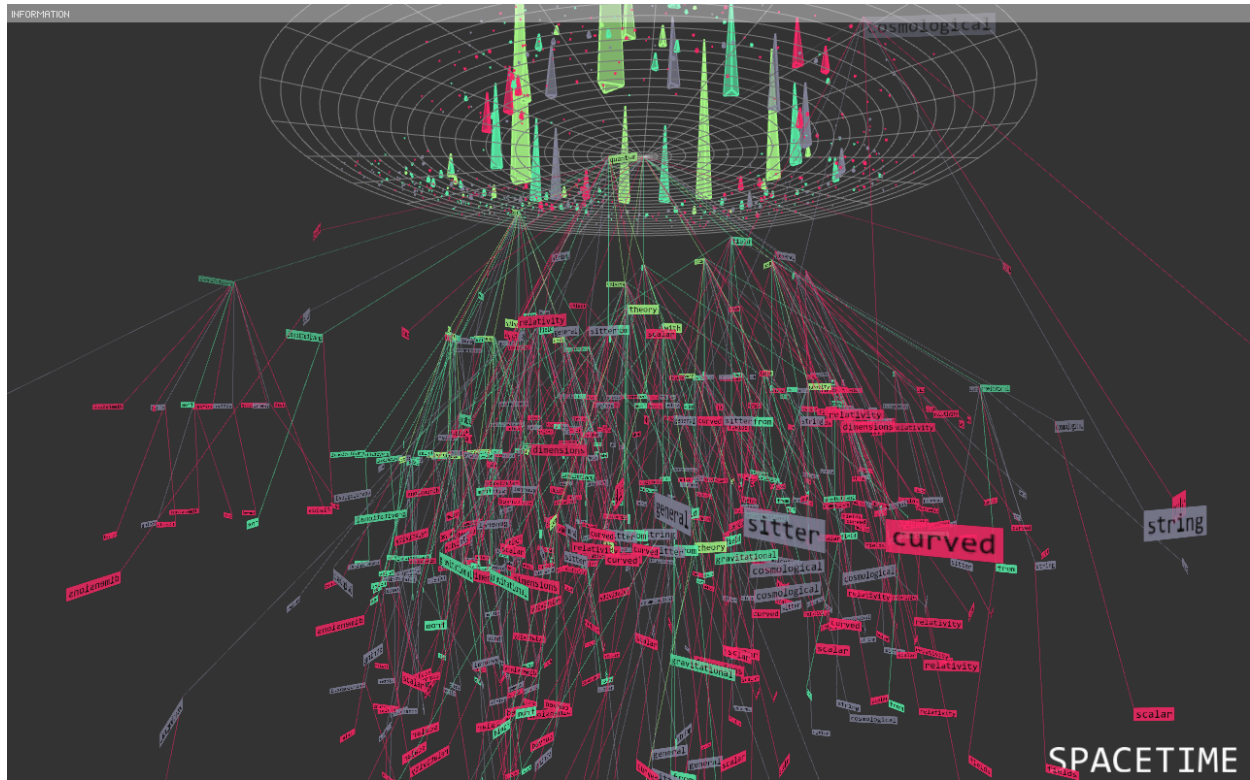
Once the data is passed through the FP-Tree algorithm the results for a minimum support value of 500 show that the most common terms associated with spacetime are the following: black:1094 quantum:938 with:816 theory:761 gravity:748 field:641 :580 gravitational:540 holes:519 from:518. This leads me to conclude that people are indeed writing about spacetime most likely in association with black matter, quantum theory, black holes, and gravitational fields.

At the moment the colors are mapped to the data such that the top 10 connections from each

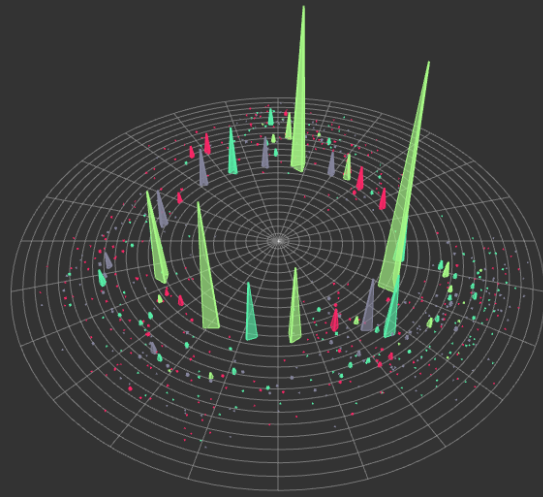
node are of a particular color out of a palette of 5 colors. The important thing is to see that the pyramid heights and position of the nodes below are the same data, just represented at a different scale.

All the views show the same basic data in different ways and reveal slightly different patterns based on how you color nodes or corresponding shapes. It can be interpreted many ways, which is the beauty of data visualization!

Screen Captures

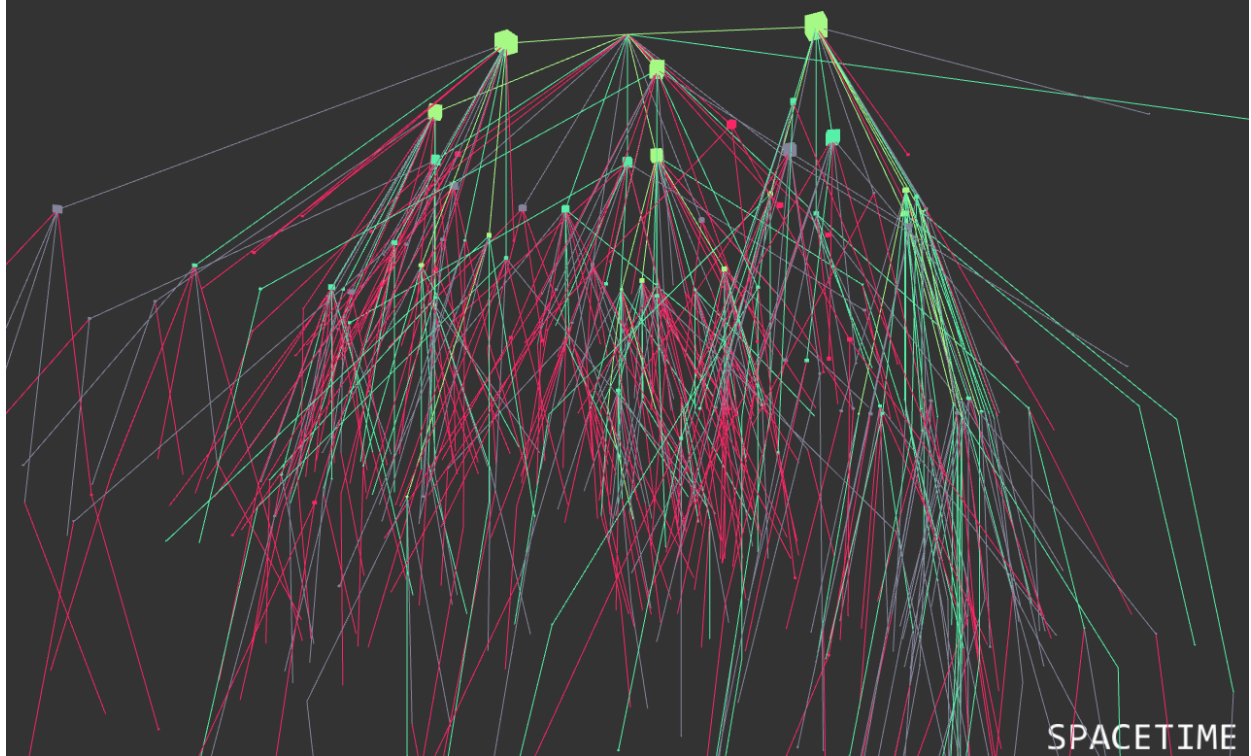


INFORMATION



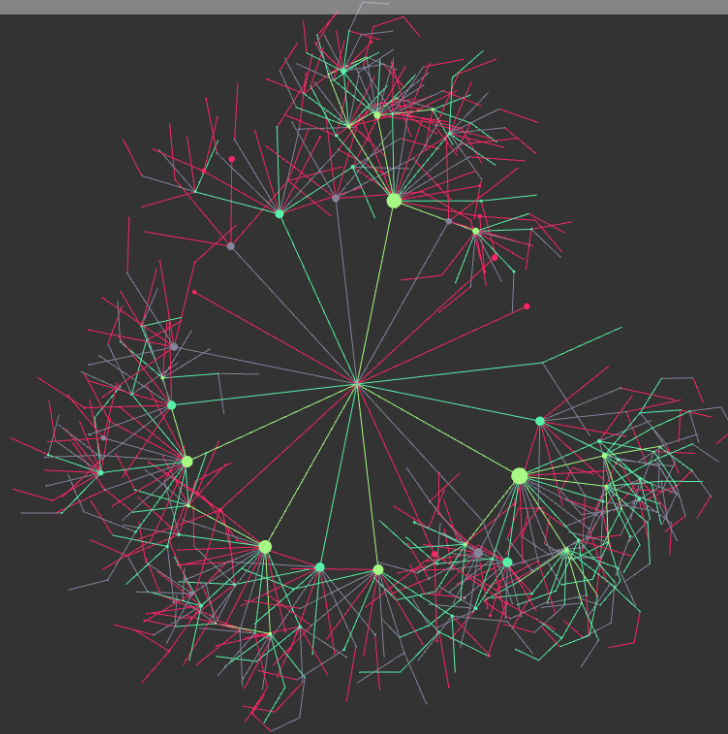
SPACETIME

INFORMATION



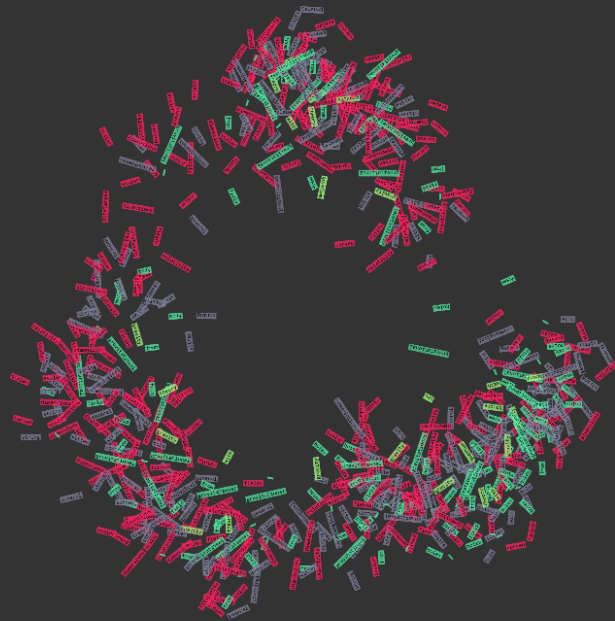
SPACETIME

INFORMATION



SPACETIME

INFORMATION



SPACETIME

Resources

<http://processing.org>

<http://arxiv.org/>

<http://arxiv.org/help/api/index>