Graph Algorithms for Visualizing High Dimensional Data

Abhinav Shankaranarayanan Venkataraman
Directed by : Prof. Ricard Gavalda and Prof. Marta Arias

Universitat Politecnica de Catalunya (UPC), Barcelona

27 June 2016

1. Outline
2. Introduction
   - Community Structure
3. State-of-the-art in Community Detection
4. Louvain Community Detection Algorithm
   - Louvain Method
     - Louvain Method
     - Louvain Method
     - Mode of implementation
     - Louvain Method
     - Louvain Method
     - Louvain Method
5. Visualization Libraries
   - Alchemy.js
   - Alchemy.js
   - Results
6. Overall System
7. Conclusions
   - Personal Learning
   - Softwares and tool kits that were used

Project Research Group

- This project is carried out within the LARCA research group at UPC .
- Researchers within LARCA have in the last two years began collaborations with hospital and health agencies for the analysis of electronic healthcare records [EHR].
- In previous work within the group, they pro- posed to organize the information in EHR in the form of graphs and hyper- graphs, which can then be navigated by experts and mined with graph and network theoretic tools.

What is Community?

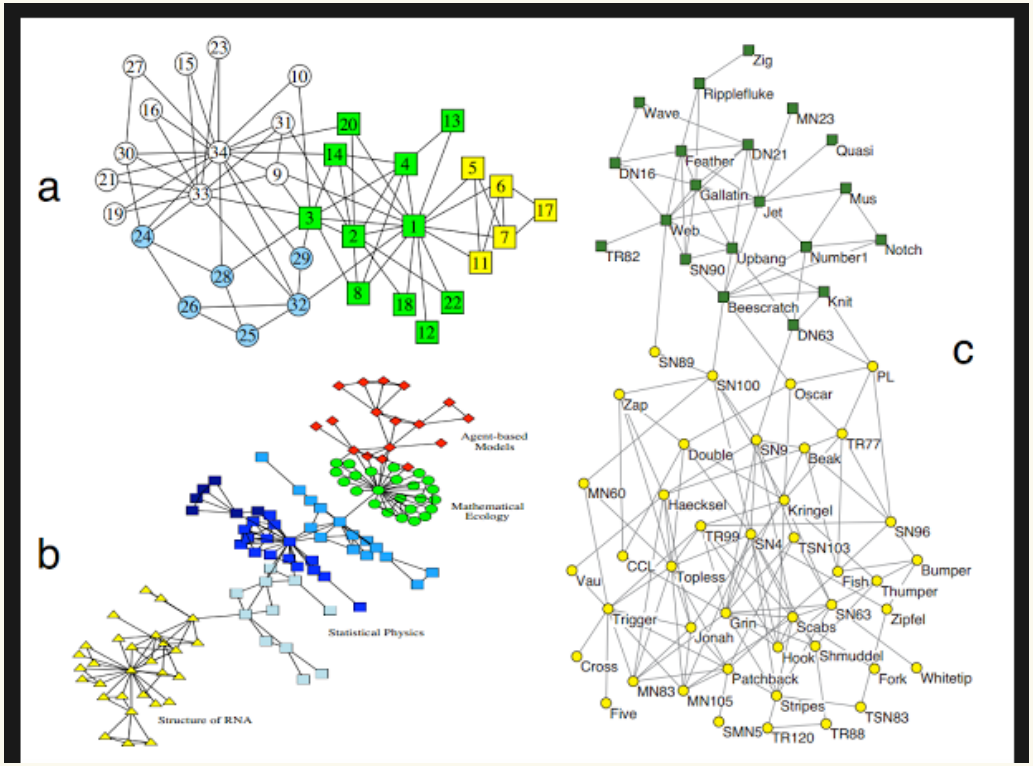


Figure : Communities: [2]

Goal of the Project

1. To survey a few algorithms that aim in community finding keeping in mind that the input is from the medical domain.
2. To choose an algorithms that benefit the purpose of organizing graphs from medical domain and for the purpose of visualization.
3. To implement the algorithms and test the efficiency of the algorithm using variety of graphs.
4. To build a Graphic User Interface (GUI) which enables visualization of the raw input on a web browser by drawing graphs.

Planning and Budget

1. Planning:
   - Required knowledge acquisition
   - Paper Analysis
   - Design and Implementation
   - Testing I
   - Testing II
   - Report Writing
2. Economic budget: Hardware budget, Software Budget, Human Resource Budget
3. Sustainability: Economically sustainable, Socially sustainable, Environmentally sustainable

State-of-the-art in Community Detection

Communities are a part of the graph that have fewer ties with the rest of the system. A community should be densely connected, well separated from the rest of the network and the members of the network should be more similar among themselves than with the rest.

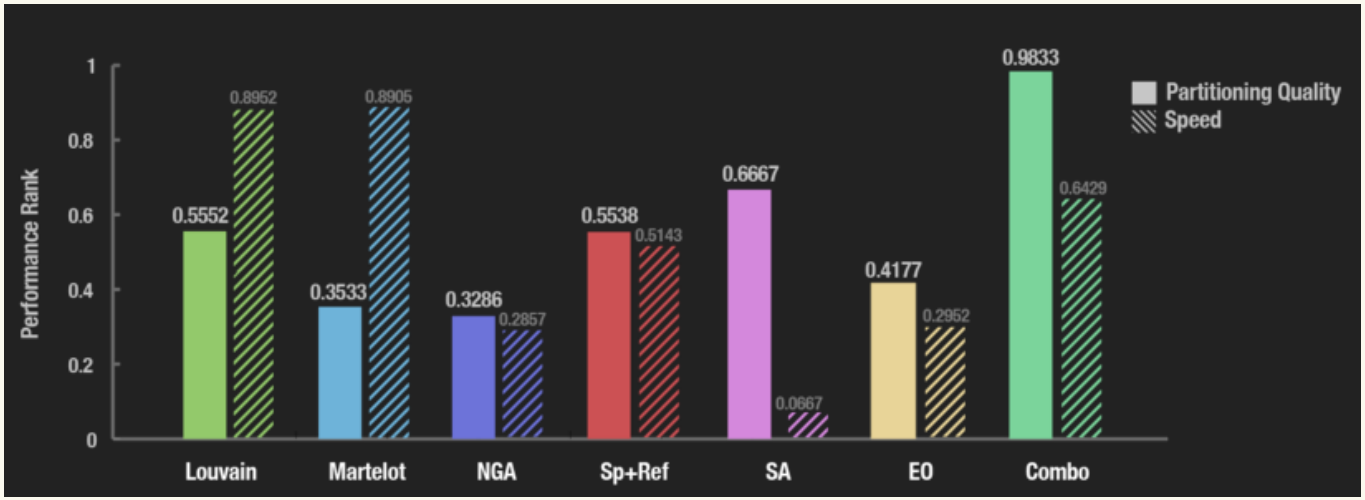


Figure : Exploring state of the art: [3]

Louvain Algorithm [2]

Louvain algorithms is the state of the art community detection Algorithm. Louvain algorithm attempts to maximize modularity. This algorithm has two phases. The diagram shows the

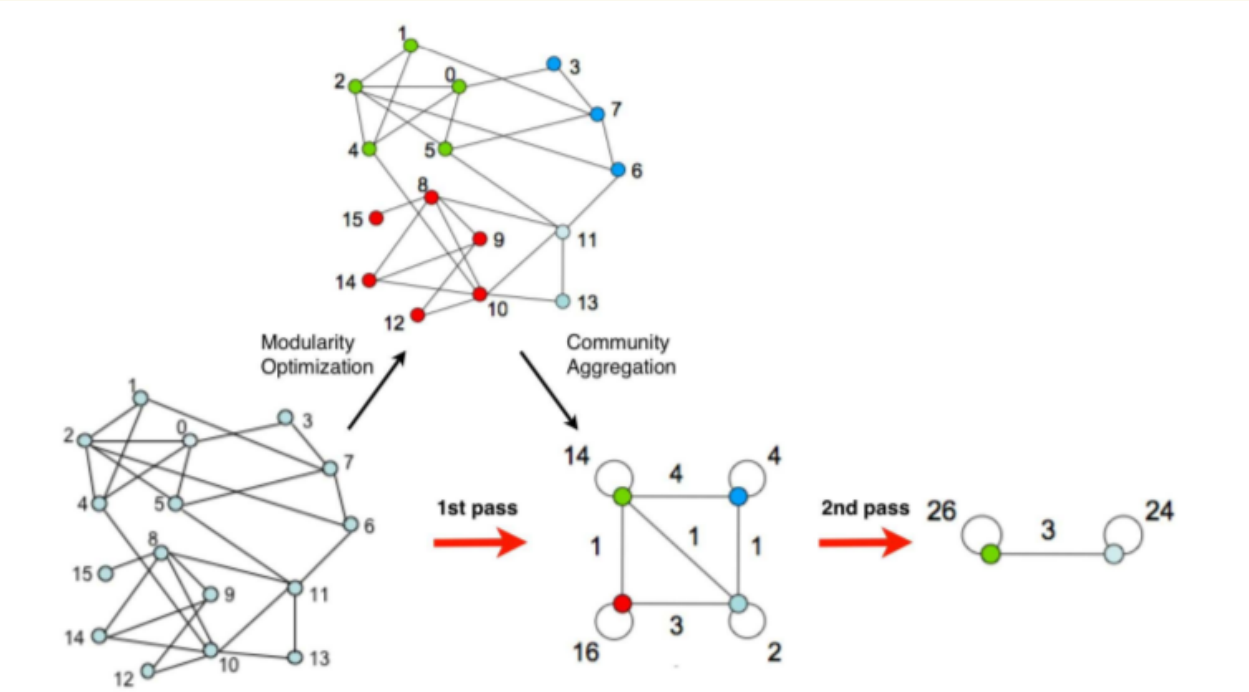


Figure : Visualization of the steps of our algorithm. This was taken from the paper ”Fast unfolding of communities in large networks” [1]

Louvain Algorithm Pseudocode

Louvain Algorithm Pseudocode:

1. Repeat until local optimum is reached:
    1. Phase1 : Split or partition the graph by optimizing modularity greedily
    2. Phase2 : Agglomerate the found clusters into new nodes

First phase in Louvain

Louvain Algorithm Pseudocode for Phase1:

1. Assign a different community to each node.
2. For each node v_i
   - For each $v_j \in N(v_i)$,consider removing v_i from community of v_i and place it in the community of v_j
   - Choose v_i into community of neighbour that leads to highest modularity gain (Greedy Choice).
3. Repeat until no improvement can be done

Second phase in Louvain

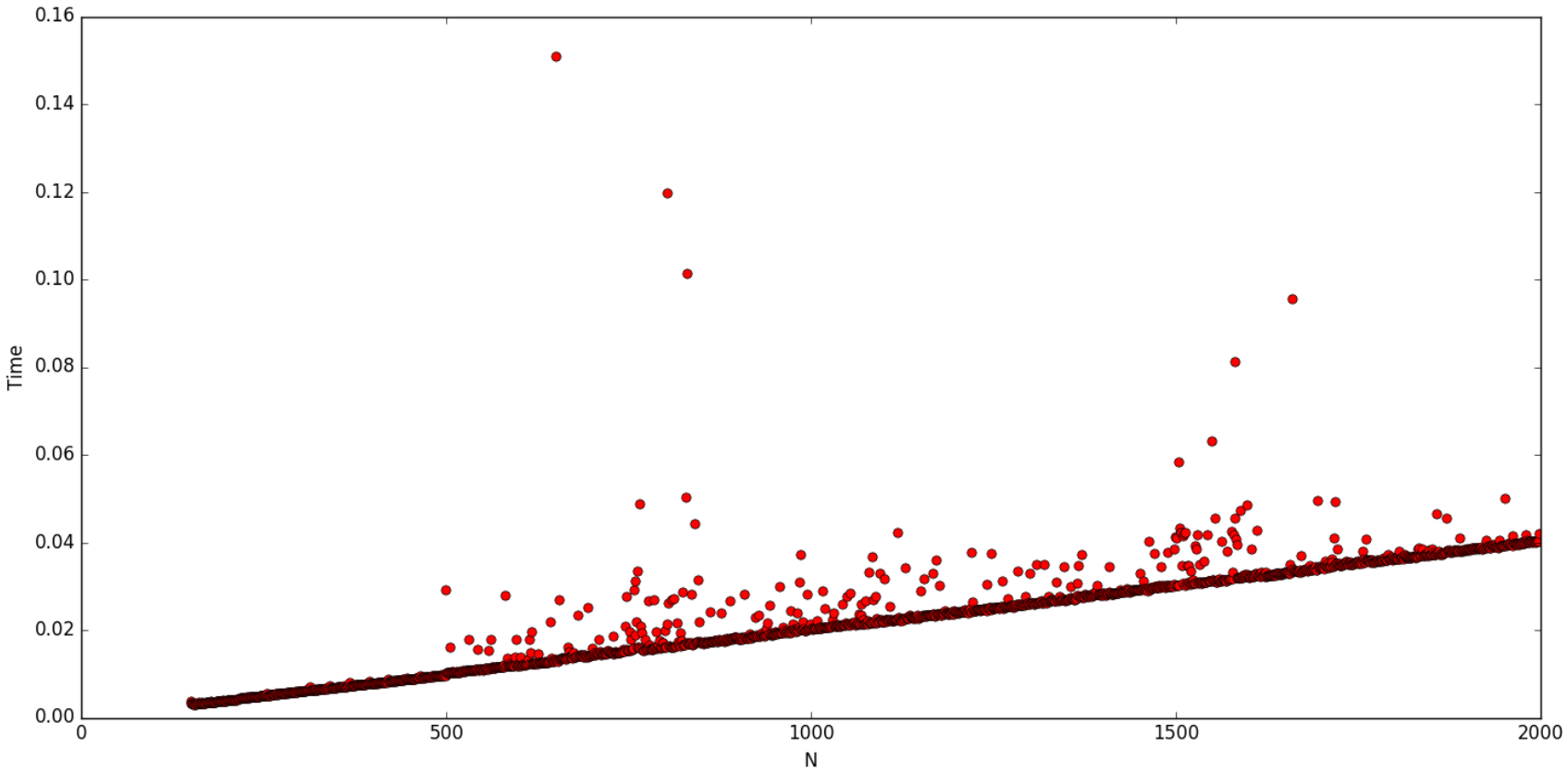
Louvain Algorithm Pseudocode for Phase2:

1. Let each community C_i form a new node v_i
2. Let the edges between new nodes v_i and v_j be the sum of edges between nodes in C_i and C_j in the previous graph

Mode of implementation

The implementation of the Algorithm is in Python. pyLouvain is code that is freely available although considering the task perfomed by the project it is tough to use pyLouvain directly. Hence modifications were made to pyLouvain and some part of the code was reused.
The input data structure was altered. The Input file is stored in a matrix and its transpose is used to get the node set. This is used to for a edge dictionary.

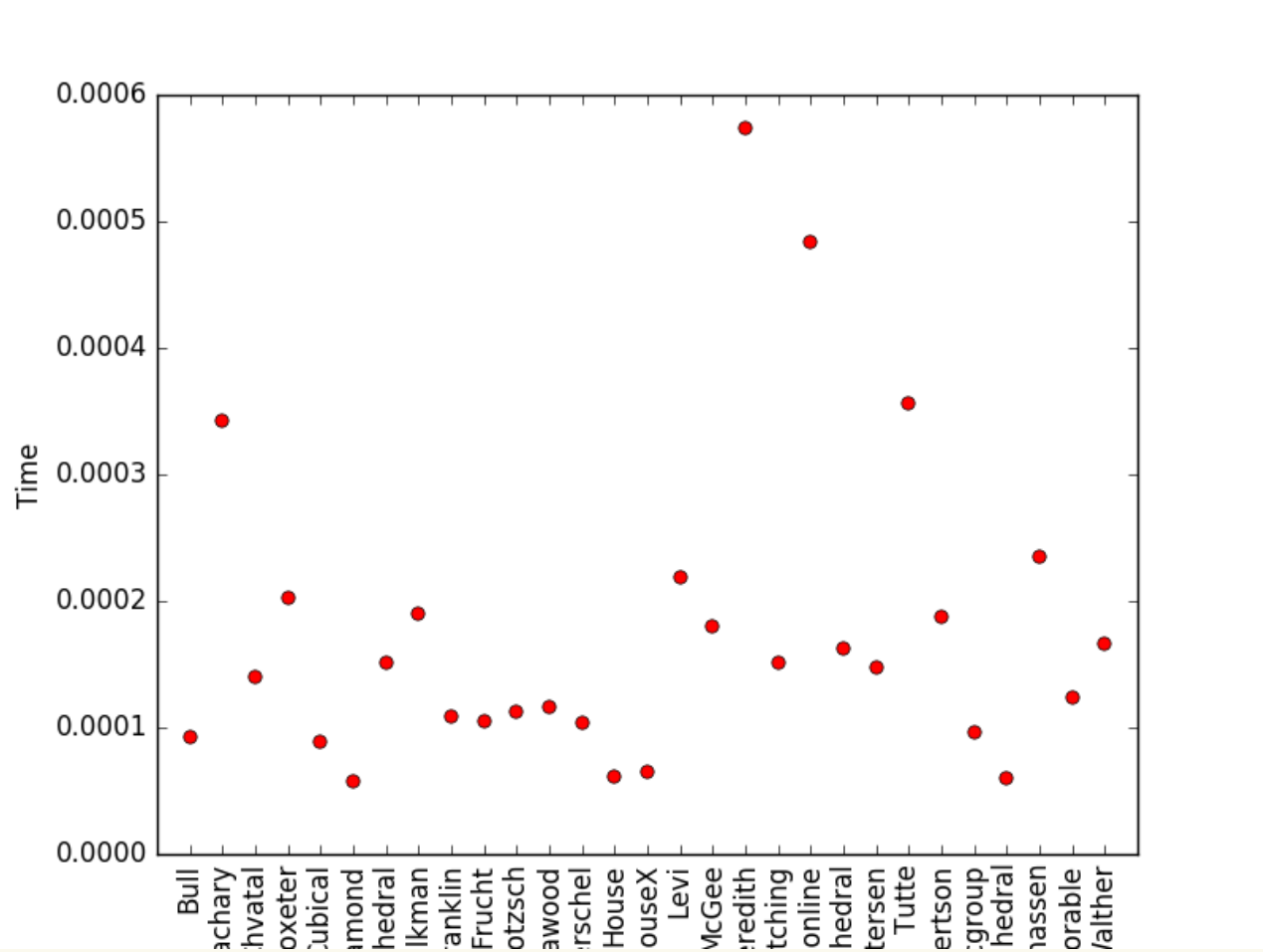
N until 2000 and Q=0.4, Scale-Free distribution Experiments



SNAP Experiments

Description	Nodes	Edges	communities Known	Time
Email network from a EU research institution	265214	420045	---	2.24
DBLP collaboration network	317080	1049866	13477	5.58
Amazon product network	334863	925872	75149	5.75
Youtube online social network	1134890	2987624	8385	17.54

Famous Graphs Experiments

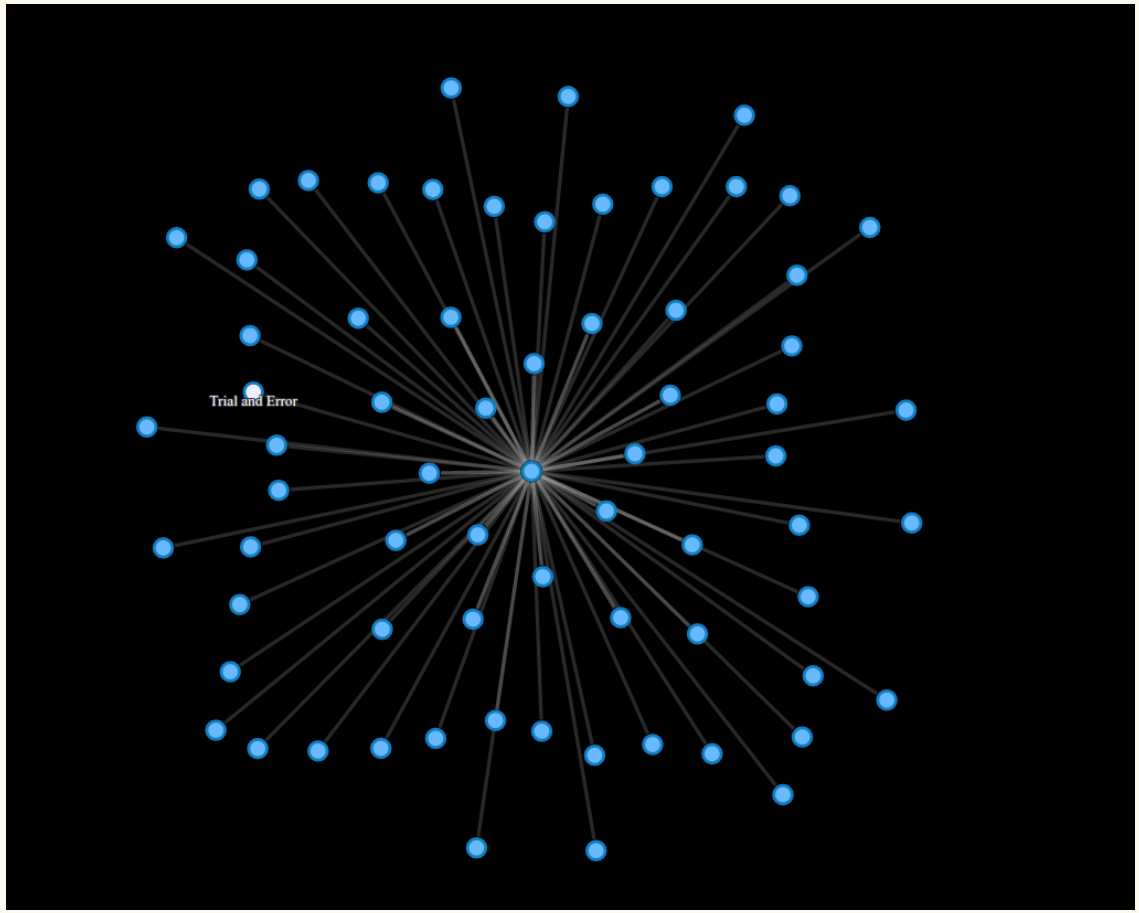


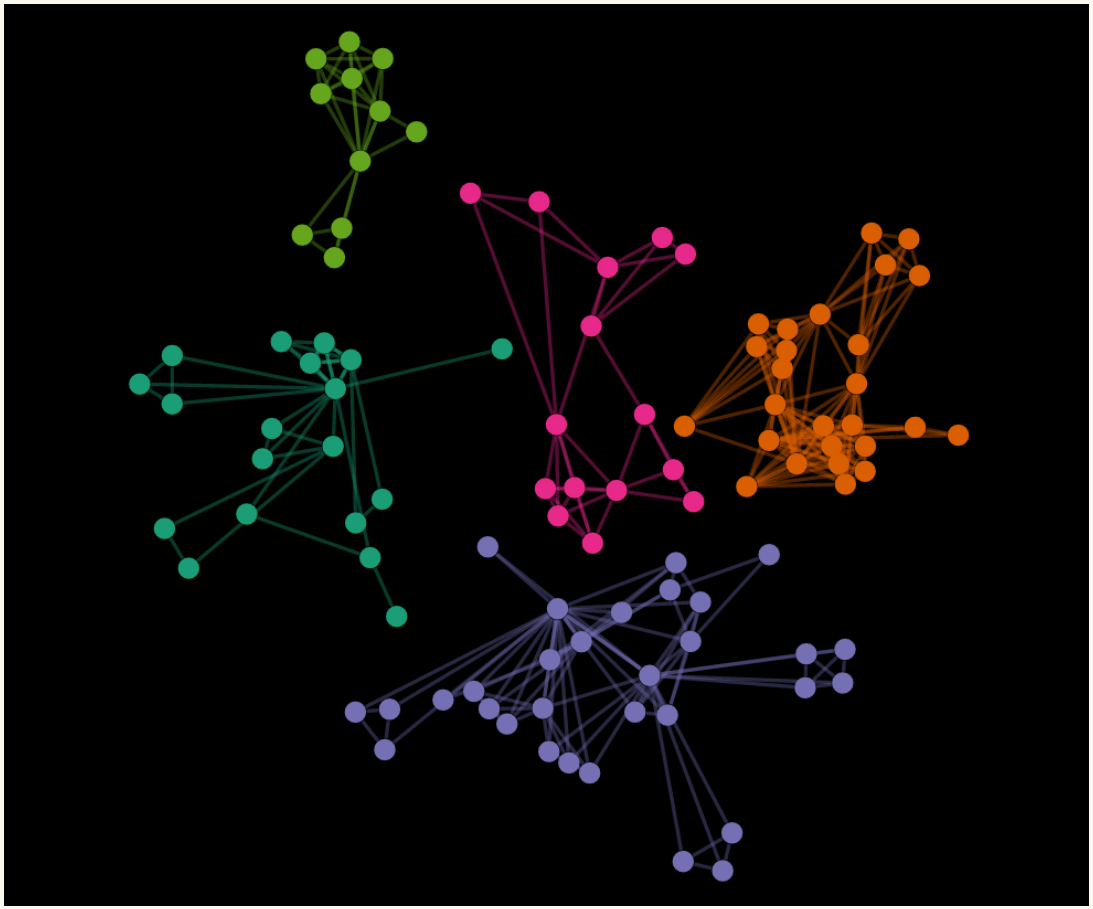
Visualization Libraries

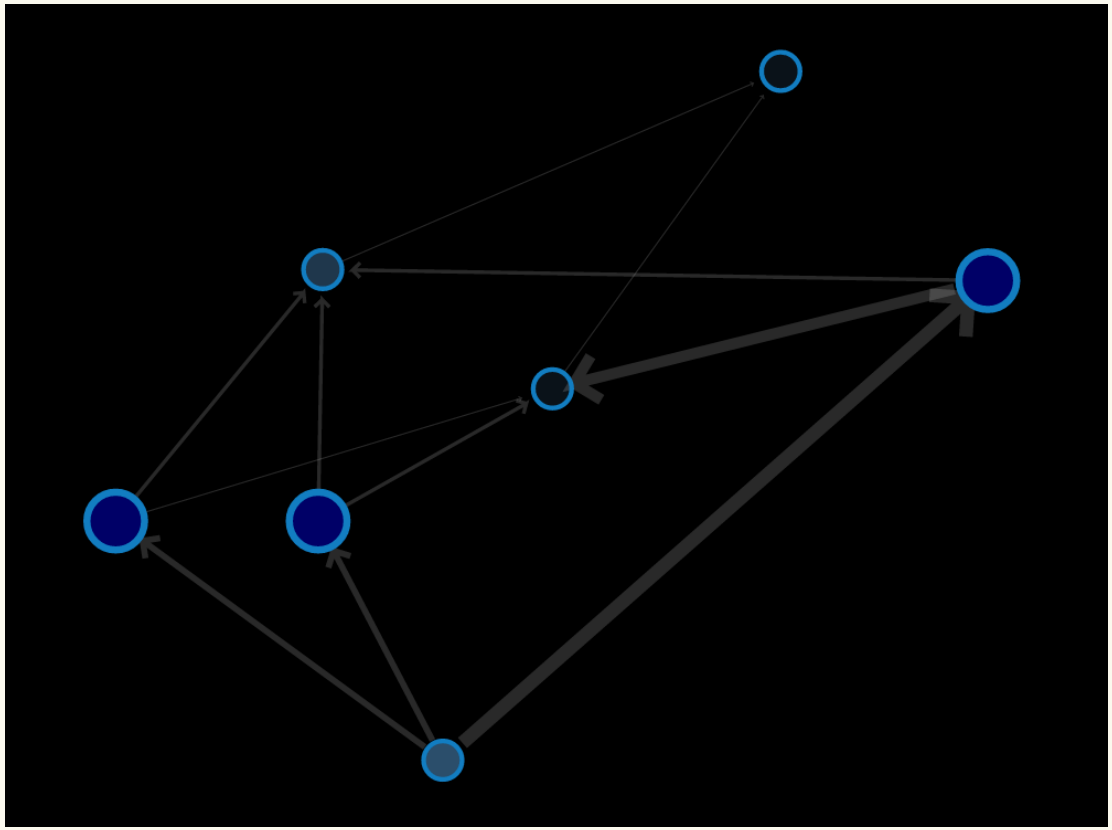
	Protovis.js	D3.js	Alchemy.js	Gephi
JavaScript	✓	✓	✓	
JSON Object	✓	✓	✓	
Robust		✓		✓
Less Overhead			✓	

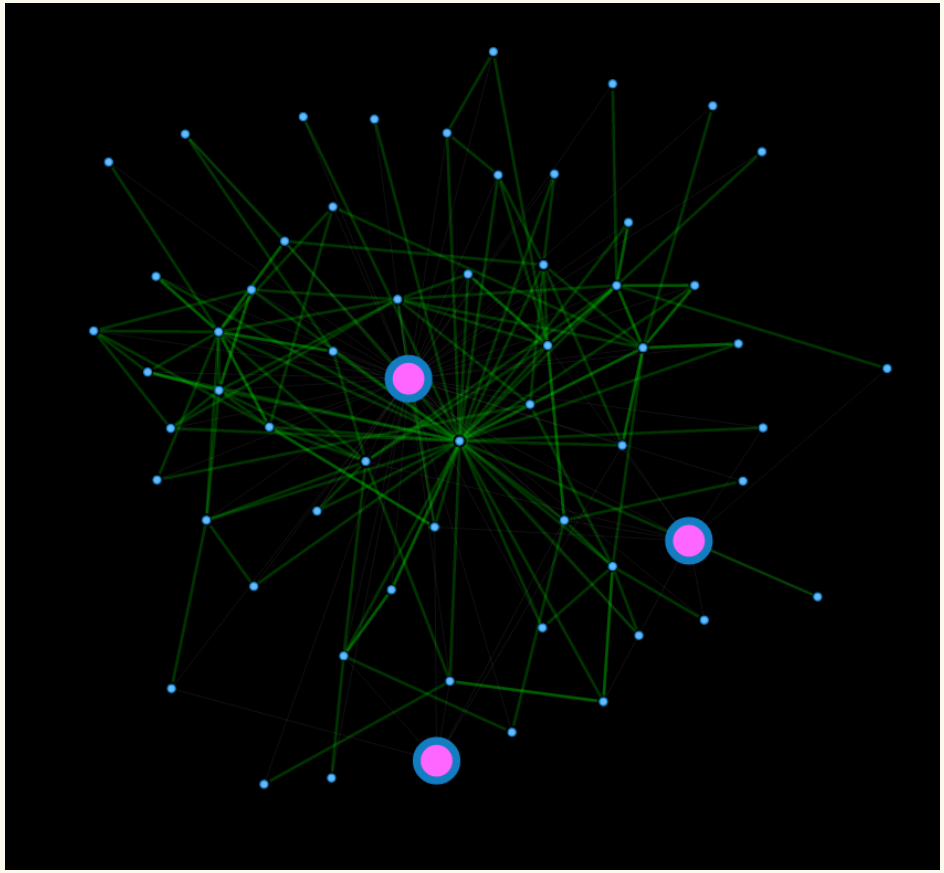
Table : Comparing Visualization methods

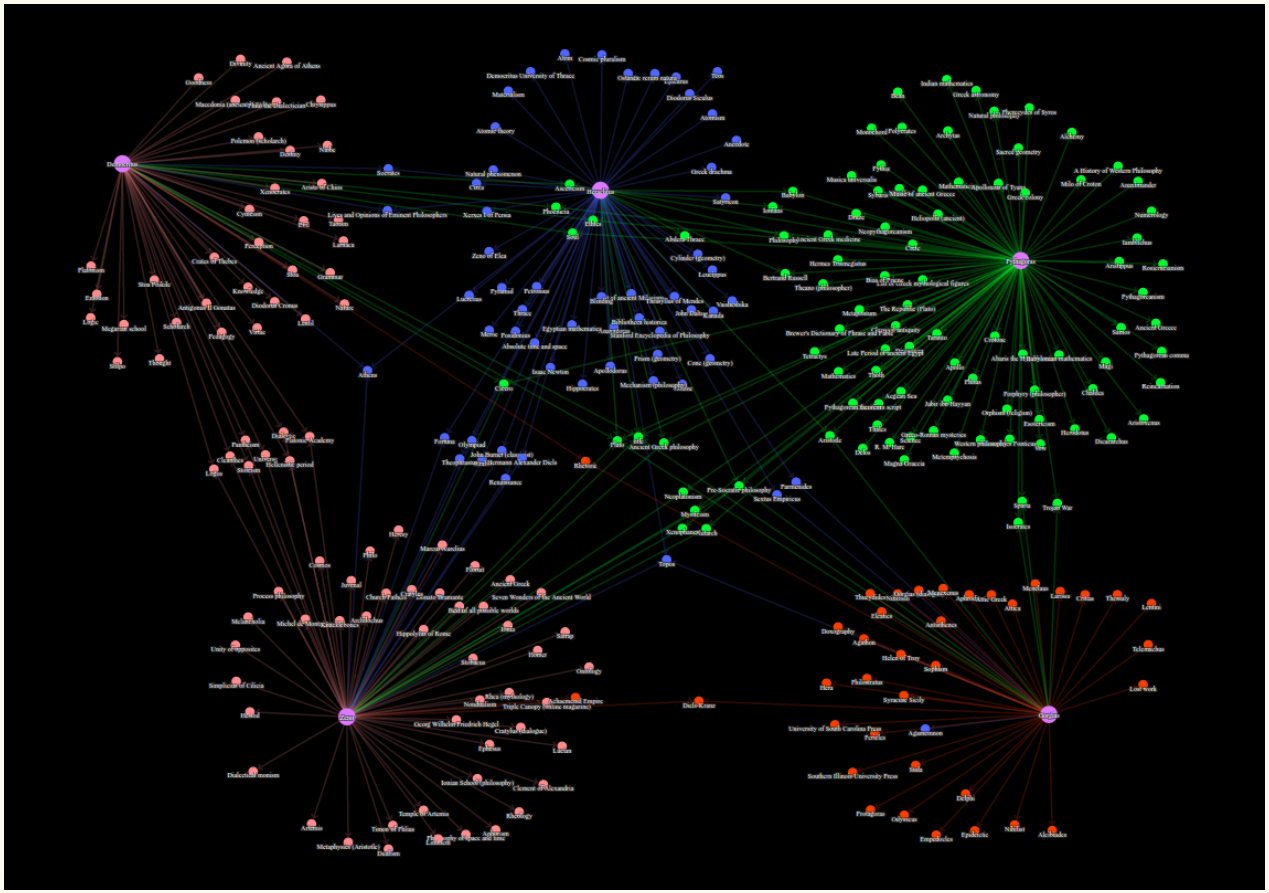
1. Alchemy needs three main units to form as an application namely: alchemy.css, alchemy.js and data.
2. Five simple steps to connect the JSON object to draw the graph.
3. Tests:

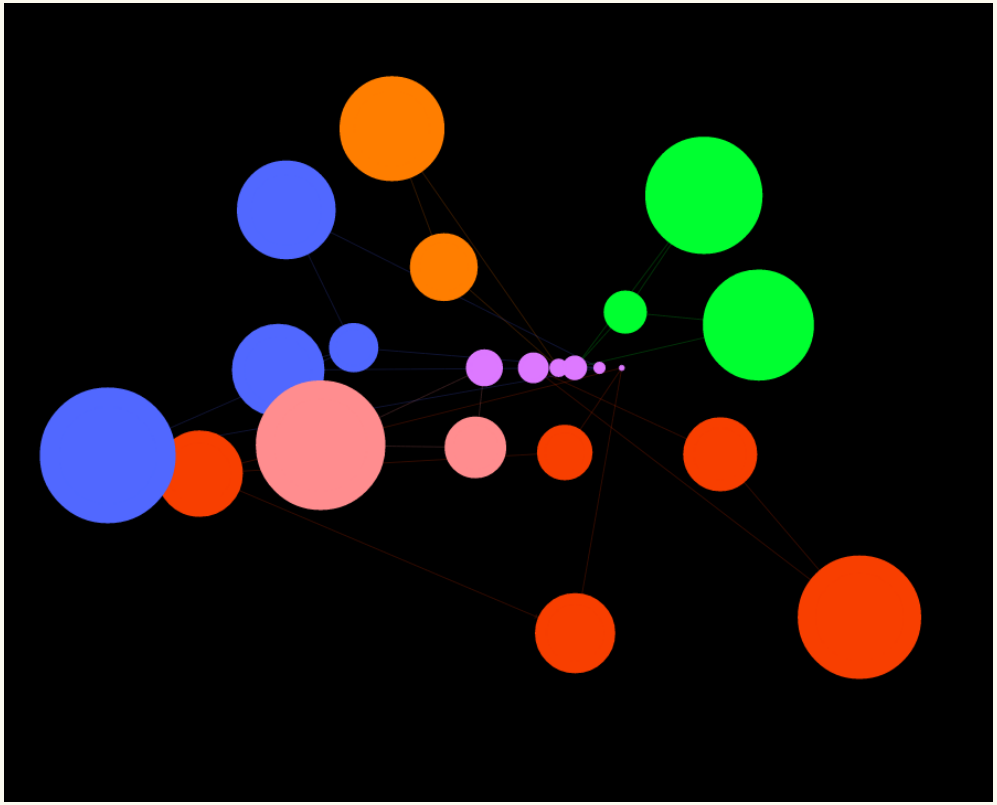




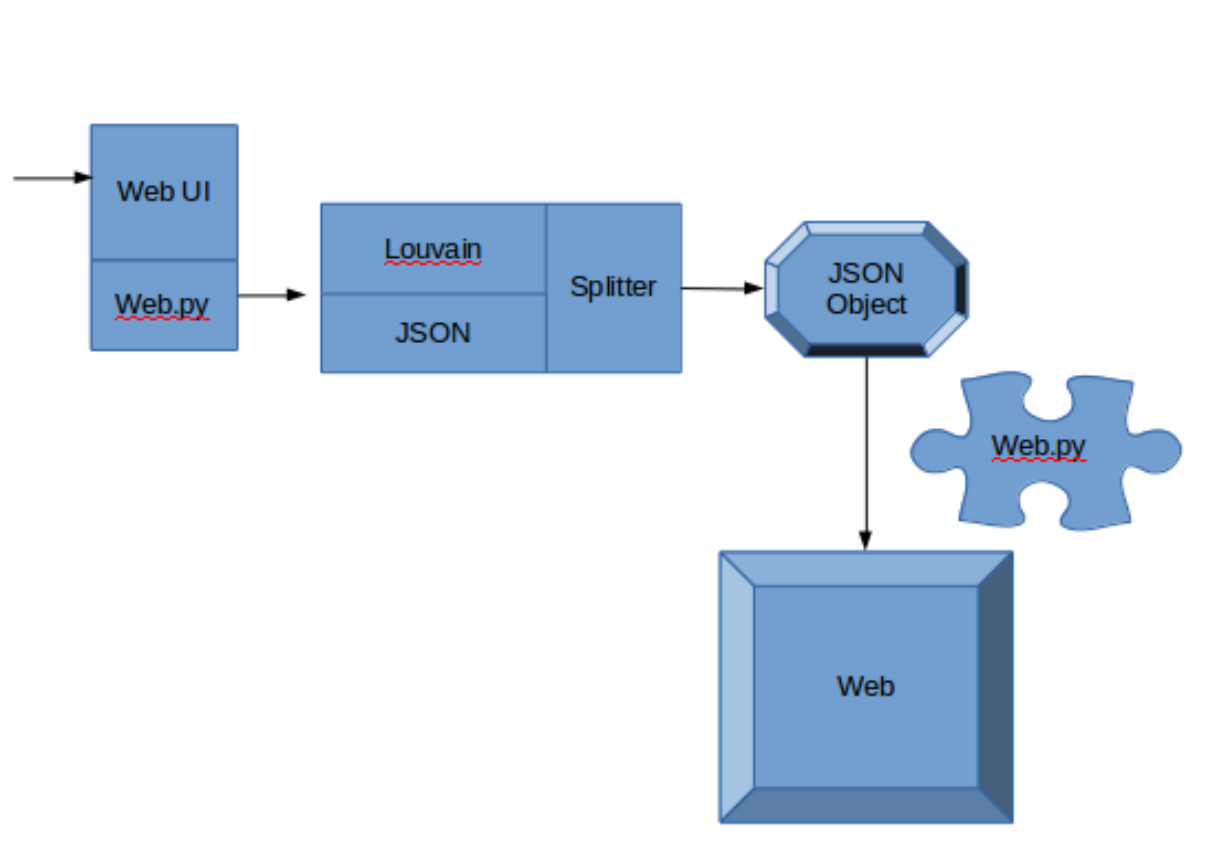








How it works?



Personal Learning

The project gave a huge scope for exploring many softwares and trying out several of them. It gave an oppurtunity for me to learn Python to depth. My interest foe Data visualization and graph algorithms has increased. By trying a web application I got accustomed to using web frameworks and web technologies. Since the project had more scope for exploration. My interest in Data Visualization has increased. My interest in graphs has increased. My python programming skill has also increased along with that I have also learned to code for web technologies on my own.

Software tools

1. git
2. github pages
3. Linux OS

List of References that were used

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre.
Fast unfolding of communities in large networks.
Journal of statistical mechanics: theory and experiment, 2008(10):P10008, 2008.

Santo Fortunato.
Community detection in graphs.
Physics reports, 486(3):75–174, 2010.

Stanislav Sobolevsky, Riccardo Campari, Alexander Belyi, and Carlo Ratti.
General optimization technique for high-quality community detection in complex networks.
Physical Review E, 90(1):012811, 2014.

Thank you

Thank you for all those who supported me throughout the project. It was a Great time at Barcelona working with Prof.Ricard and Prof.Marta.