Tutorial: step-by-step

1. ETL(Data Load steps):
2. Navigate to “etl” project under the main folder of the source files
3. Make sure that you have ‘a folder named “csv” under project “etl” and have “citations.txt” file in it.
4. Specify your path to dblp.xml in Main class of the project
5. Run the project (or Main class)
6. You will have csv files generated under csv folder
7. Use “init\_db\_final.sql” under dbscripts folder under the main folder and run it in MySql. This script will initialize the database and tables.
8. Use “load\_all\_data\_and\_index.sql” script to load the csv files into MySql. Beware that you need to change the paths specified in this script as they are in your system.
9. The ETL process should finish in under 10 minutes in total.
10. Running Nasa/backend:
11. Open the command prompt and enter the following commands:

> activator eclipse

> activator “run 9010” where 9010 is the port number

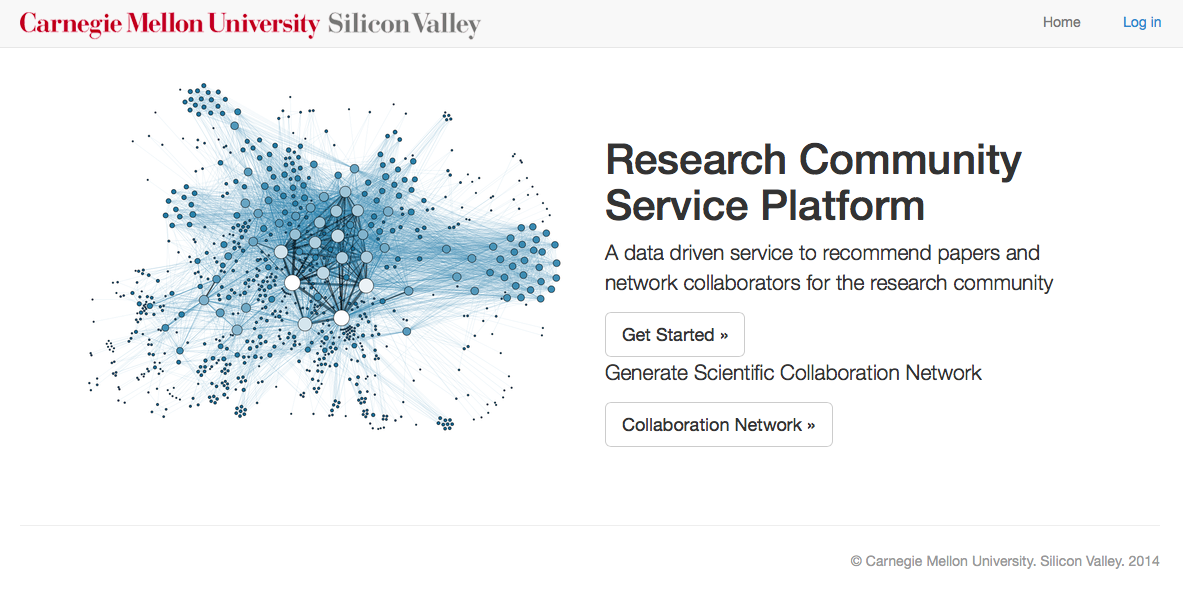
1. Running Nasa/frontend:
2. Open the command prompt and enter the following commands:

> activator eclipse

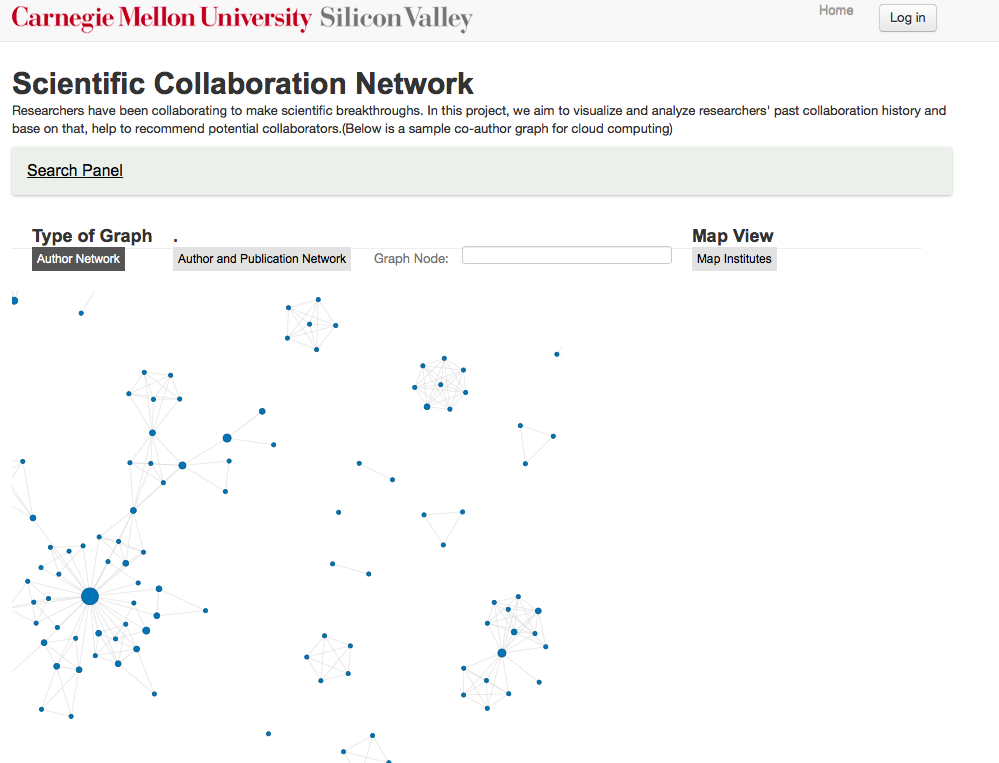
> activator “run 9009” where 9009 is the port number

1. Using the Web application:
2. Open the web application at <http://einstein.sv.cmu.edu:9005>

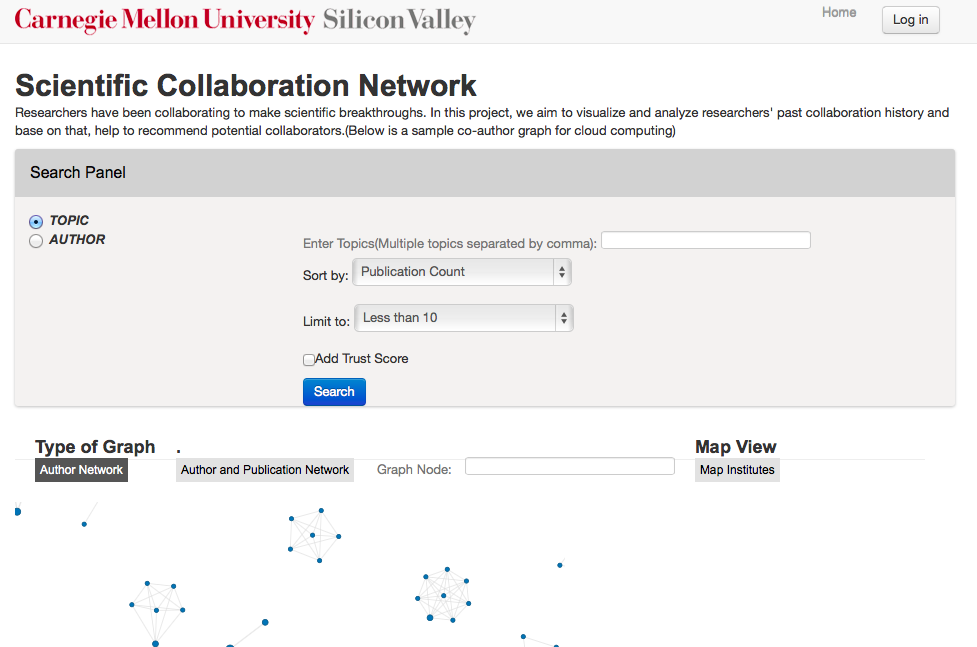
You should see the screen as shown below. Click on “Collaboration Network” button which will navigate to the next page.



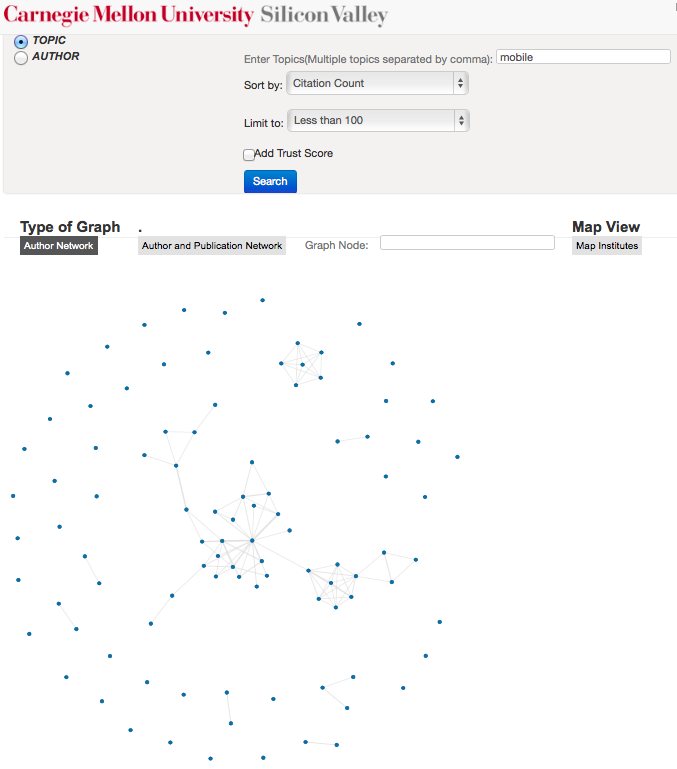
1. Clicking on “Collaboration Network” will navigate to page below.



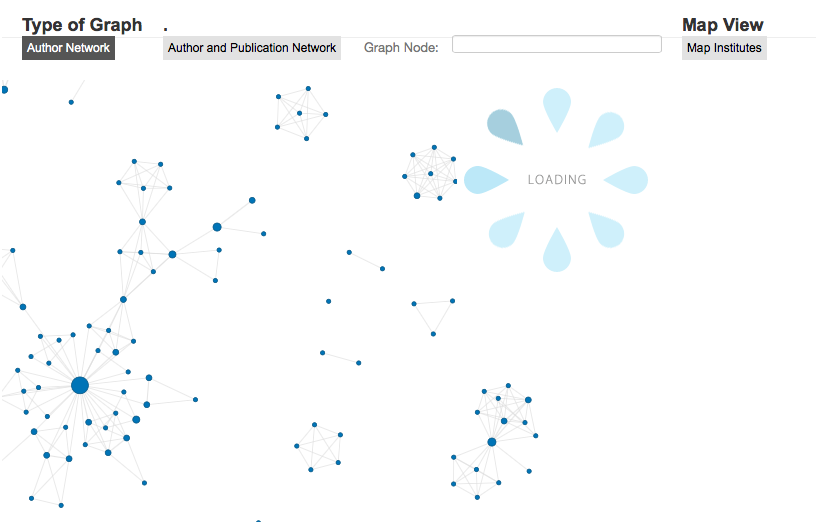
1. Clicking on “Search Panel” tab will show all the search options available as shown below. User can search either based on topic or author. If topic is selected, then the

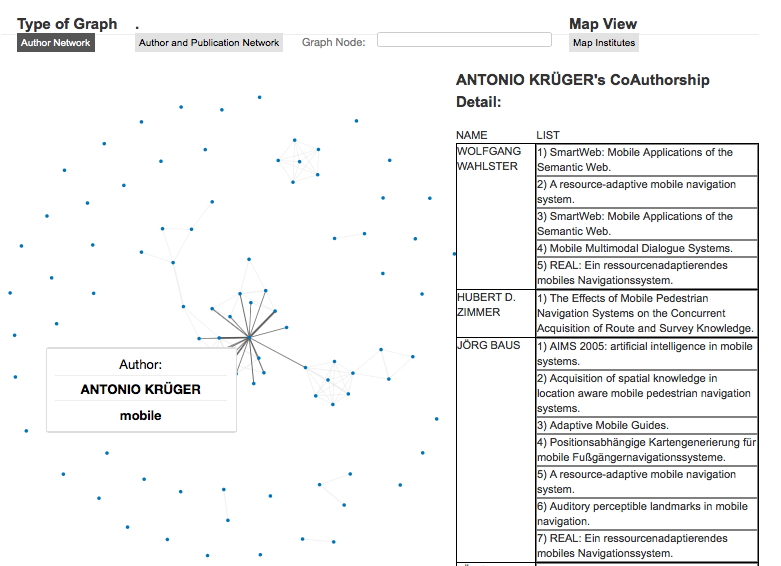
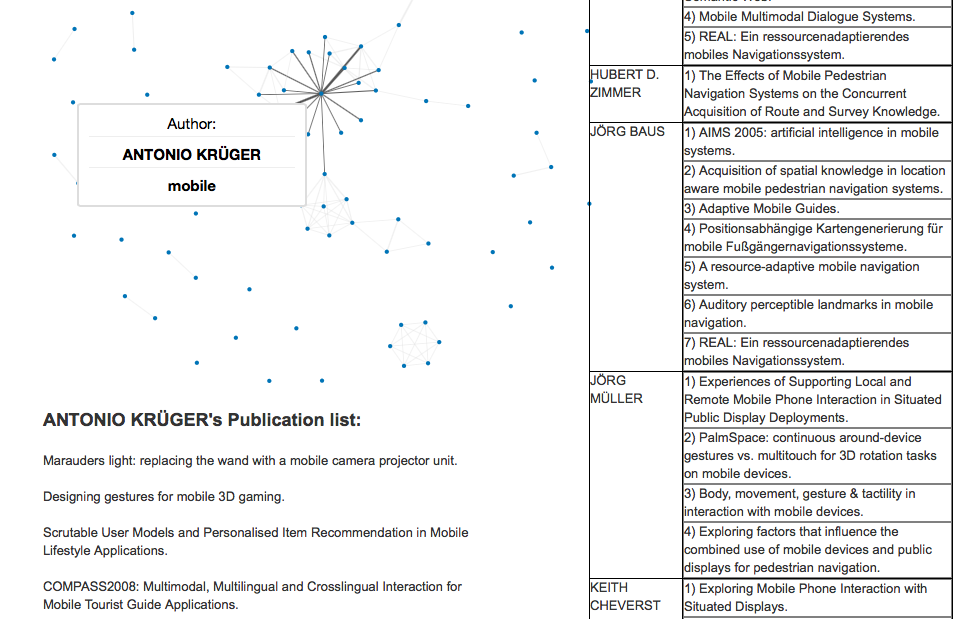
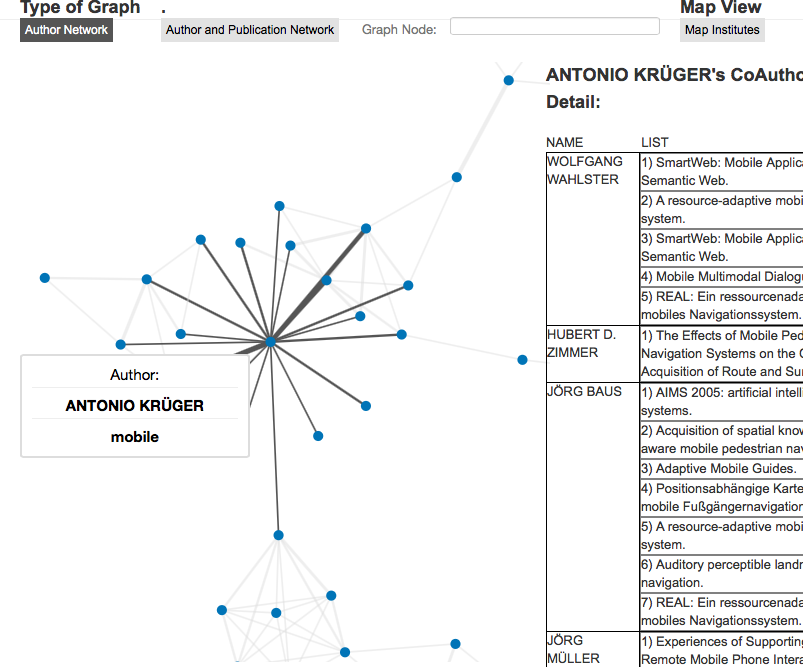
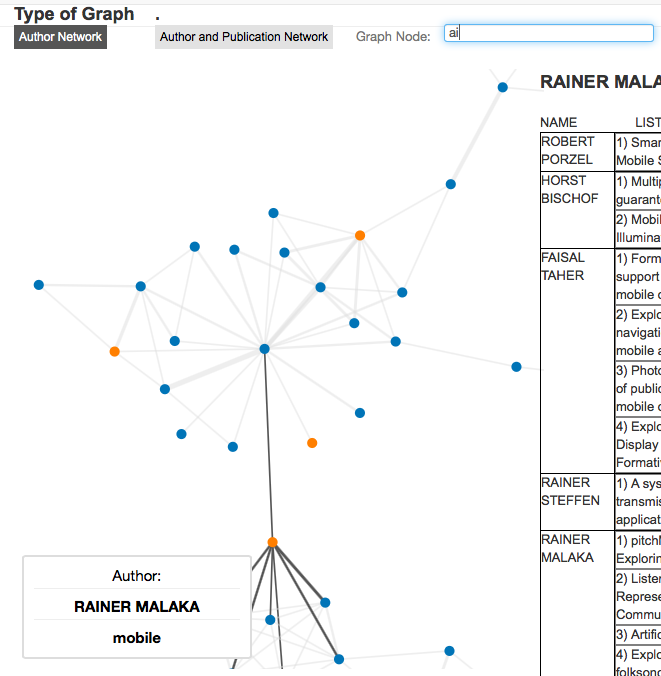
user has options to enter either single topic/multiple topics separated by comma. User can also sort by either publication count/ citation count for a topic. To see only the top 10/30/50/100 authors in that particular topic, the user has a Limit to option too.

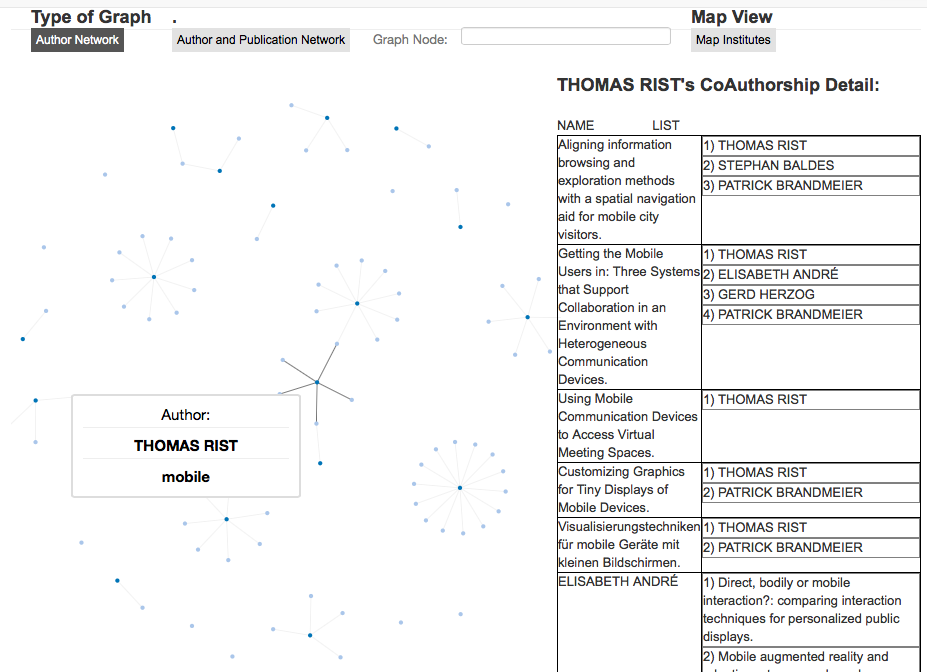
1. After selecting all these options graphs are generated- for “author-author” network and “author-publication” network.(Both the graphs are generated and rendered as and when the user clicks on author-author/author-publication network).

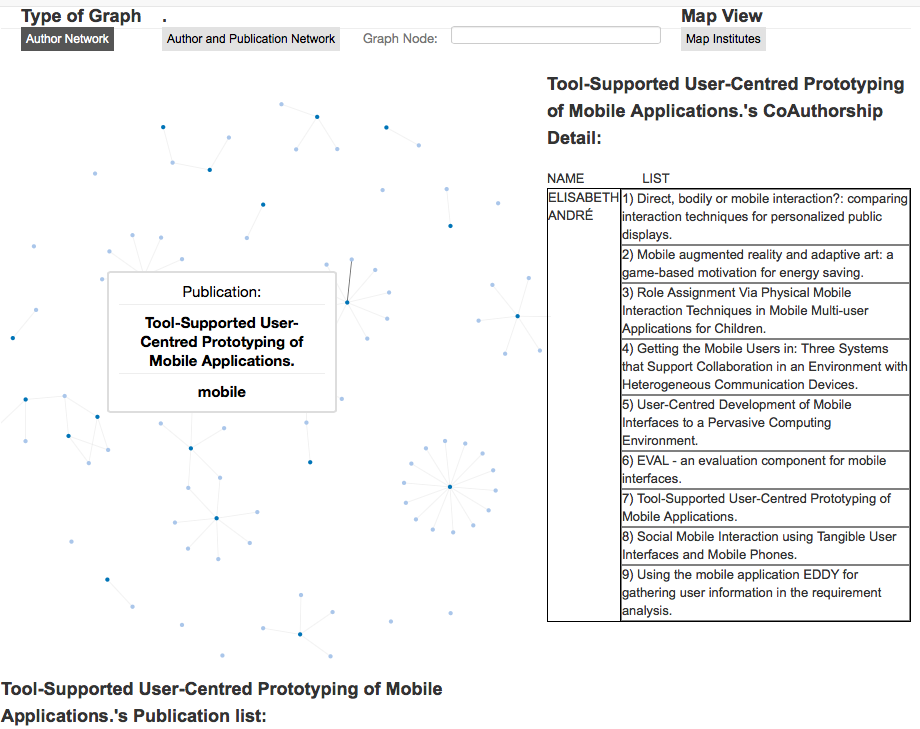


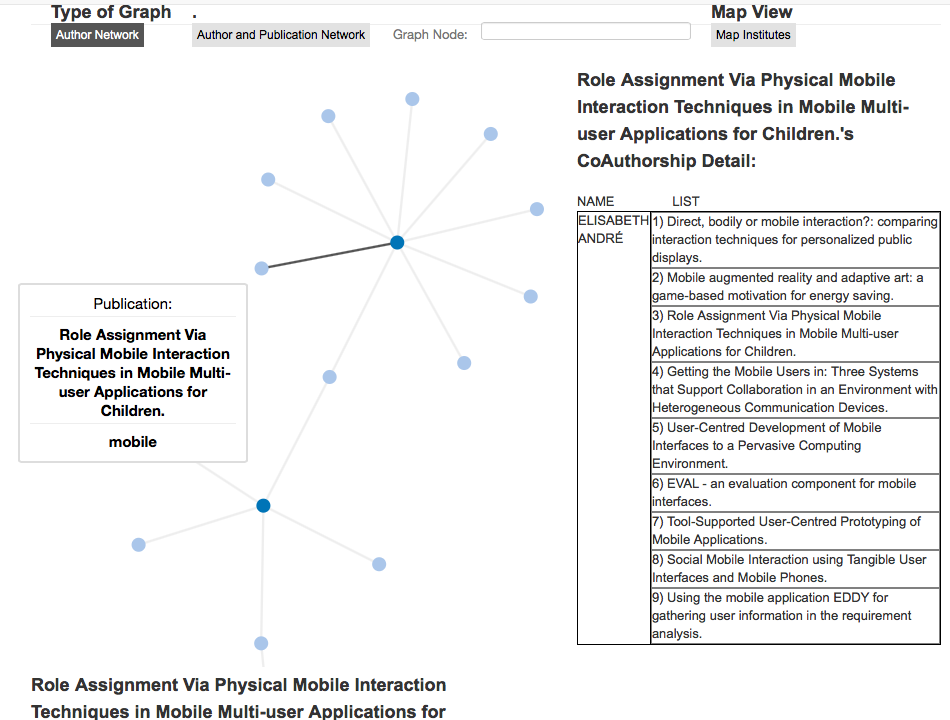
1. Even after the author network is generated still loading image is visible as author-publication network is being generated in the background but the user can interact with generated network.

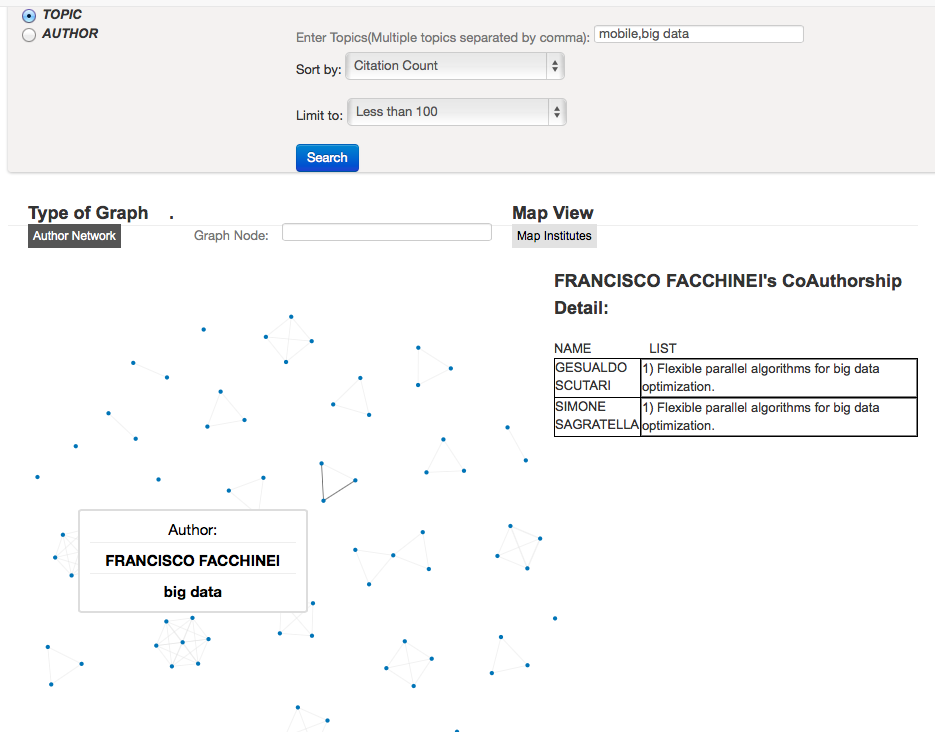
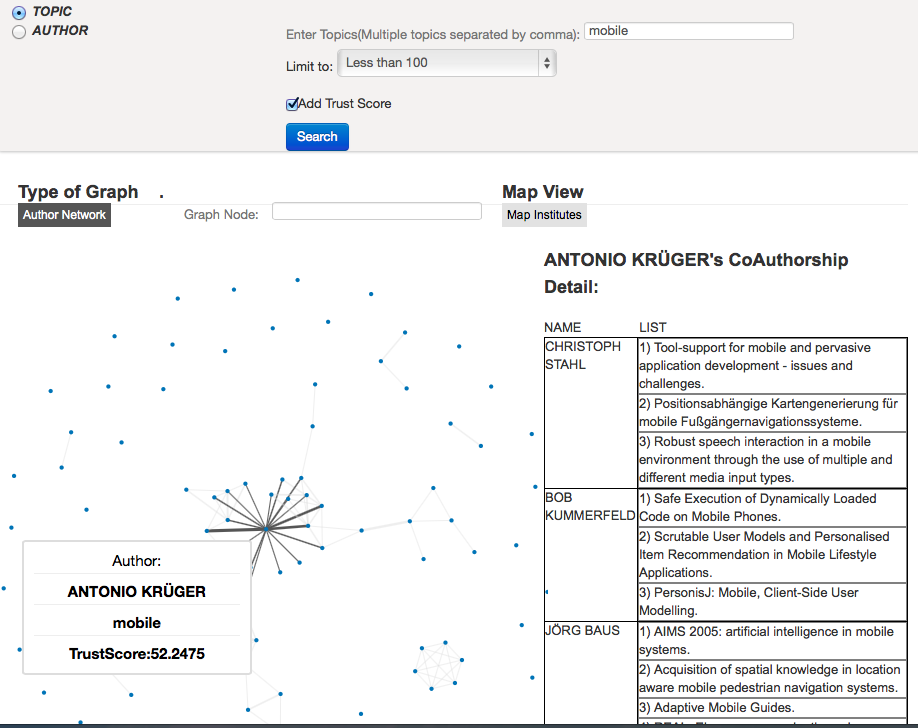
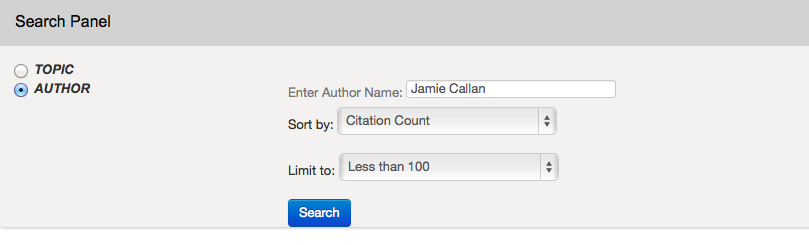
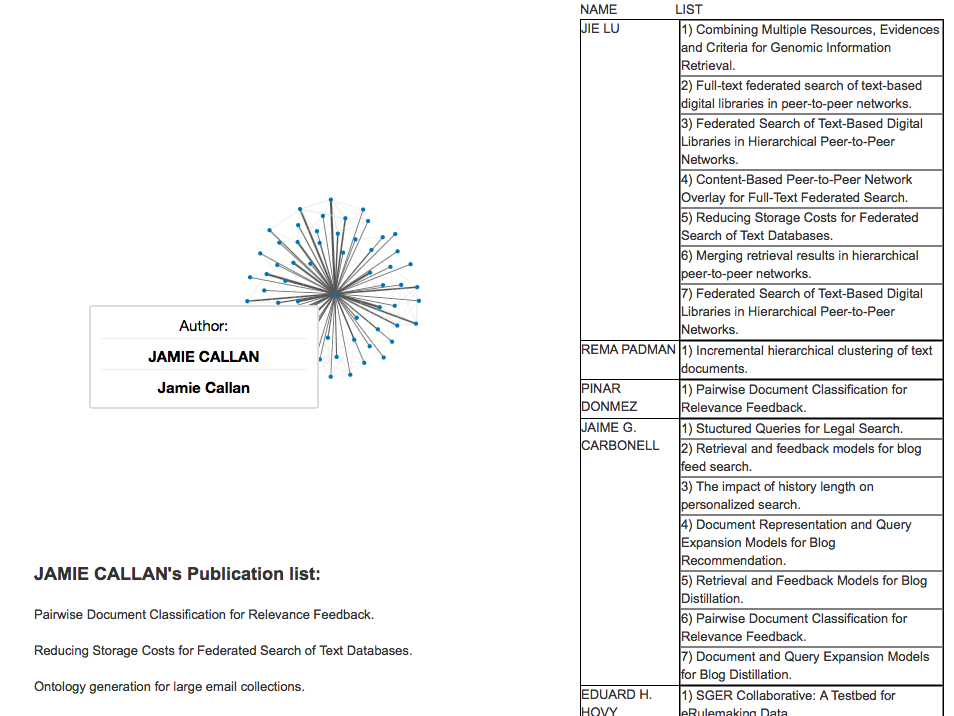
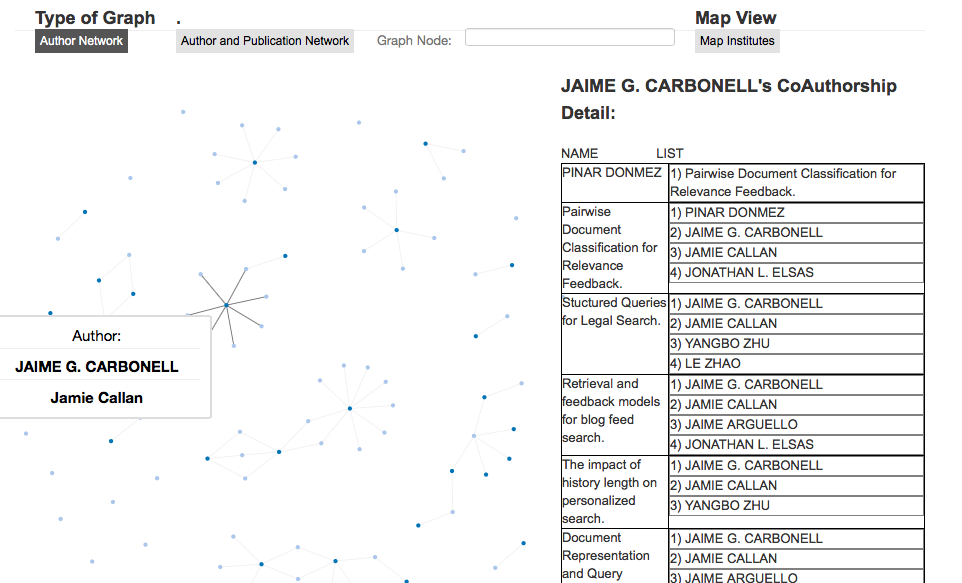


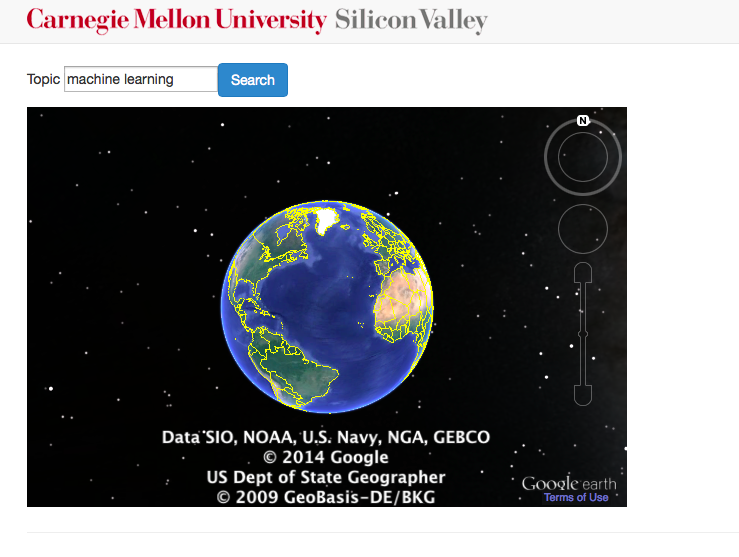
1. User can interact with the graph using mouse as follows-
2. Hovering on the nodes, will display all the coauthors of that particular author towards the right in a tabular format as shown below.
3. On clicking on the node itself, author’s own publications are displayed below the graph as shown below.
4. User can see where the graph is dense as it is zoomed out and then he can zoom-in to particular sections of the graph by mouse wheeling. The user can even drag and drop the graph to move the graph and look at other sections of the graph when zoomed-in 
5. User can search for a particular author name in the graph too by entering the name of the author or just initials in the graph node textbox and the nodes in the graph containing those letters in the name of the author node will be highlighted as shown below(all nodes with “ai” in the name are highlighted as orange in the graph). 
6. The other kind of network is displayed by clicking on “Author and Publication network” on top and the author nodes are dark blue in color while the publication network is light blue in color. User can interact with the graph in the similar way as author-author graph. Similarly the table on the right side displays details about the authors who have coauthored that particular publication on hovering on a publication node.(first figure below shows the author node and the second figure below shows a publication node)







1. User can search for multiple topics too as shown below(eg. mobile,big data below).
2. User can see the trust scores of different authors on a particular topic search by clicking on the checkbox which says “Add trust score”. This score depends on the author’s publications and his coauthors. 
3. User can search for a particular author to see his network of coauthors by selecting author search option. This also gives the similar kind of sort and limit options.
4. Below graph shows Jamie Callan’s publications and his coauthors.
5. User can also check author-publication graph of a given author by clicking on “Author and Publication Network” button on top of the graph. 
6. Clicking on the “Map View” button takes you to the following page.



This search is for schools that produce most number of PhD thesis in the world on a particular topic. on searching for “machine learning” the following schools are highlighted.

