

Visualization of Twitter Data Using Visualization Tool Kit (VTK)

Büşra Mutlu
Computer Engineering
21000277

Merve İşleyici
Computer Engineering
21000179

Ömer Emre Kaya
Computer Engineering
21400857

Abstract: Online social networking plays an important role on today's social movements by its high accessibility, connectivity and freedom of opinion sharing. Millions of people use and follow social platforms for sharing daily opinions and following current trends. In this paper, we focus on Twitter as one of most popular platforms for public opinion sharing. We aim to visualize Twitter data using VTK(Visualization Tool Kit). In the line with the information that are collected from Twitter, by selecting specific subject, we visualized data. All in all meaningless data stack constitute a meaningful whole.

1. INTRODUCTION

By the development of web technologies, people are getting more included in online data generation process via social media¹. Social media is being used as a tool for public sharing, in which individuals gain the ability to make their voice to be heard. Either anonymously or with their real identities, millions of individuals use social media every day. As one of the most important social media platforms, Twitter lets the users to share their ideas in a limited 140 character long text. The infrastructure of the website allows users to use a symbol to identify a topic, which is called the hashtag (#). With 50 million tweets per day, hashtags are central to organizing information on Twitter[2]. Thus, hashtag can be used to draw public attention to a specific user-classified topic.

In this study, our aim is to visualize and make sense of Twitter Data using VTK (Visualization Tool Kit).

2. RELATED WORKS

Initially, before starting project, we did literature researches that related to visualization Twitter data. For instance,

- The volume of Tweets available through Twitter's streaming APIs is quite impressive. While these public streaming APIs provide a smaller subset of all Tweets by definition, the amount of real-time data they return can still be a bit overwhelming to process and visualize [3].
- Differently from this, the official Revisit project is a way to redefine how we look at Twitter. With this tool we are able to create custom line maps of data connecting tweets related to one or many keywords. We can additionally add a title to our graph and share the link online (even onto Twitter) [4].
- Some project focus on visualization of specific Twitter hashtag. Instance, use this interactive network to discover 'mutual connections' on Twitter among senators, governors and congressional representative and the 2014 candidates for a number of these offices. It shows the 1,168 politicians on Twitter sized by numbers of candidates who follow them back [5].

3. METHODOLOGY

A. *Development Methodology*

Firstly, we used Agile Software Development that is a methodology for modeling and documenting software systems based on best practices .[6] Agile methods are form part of the planning spectrum. In this way, while approach has a home ground within which it performs very well, and much better than the other. [7]

B. *Technologies*

We used Java 1.8 (32-bit) technology instead of 64-bit, because VTK platform support 32-bit system. And we used NetBeans IDE 8.0 during software, besides plug in twitter library and mysql library.

To store Twitter data that collect with regard to specific hashtag, we used MySQL Workbench 6.2 CE. Therefore, it helps to select, insert and delete Twitter data. In addition to this, to generate and configure VTK files, we used Cmake cross platform. After the

producing dll, we can use specific code that related to VTK in our project by creating VTK.jar file.

For visualizing, Microsoft Visual Studio C++ 2010 Express (32- bit) and NetBeans IDE 8.0 (32-bit) we used by adding .dll file, header file, .jar file to project. Moreover , for analyzing data and capturing different visual, we used JFreeChart 1.0.17 tool that is an open-source framework for the programming language Java.[8]

C. System Design

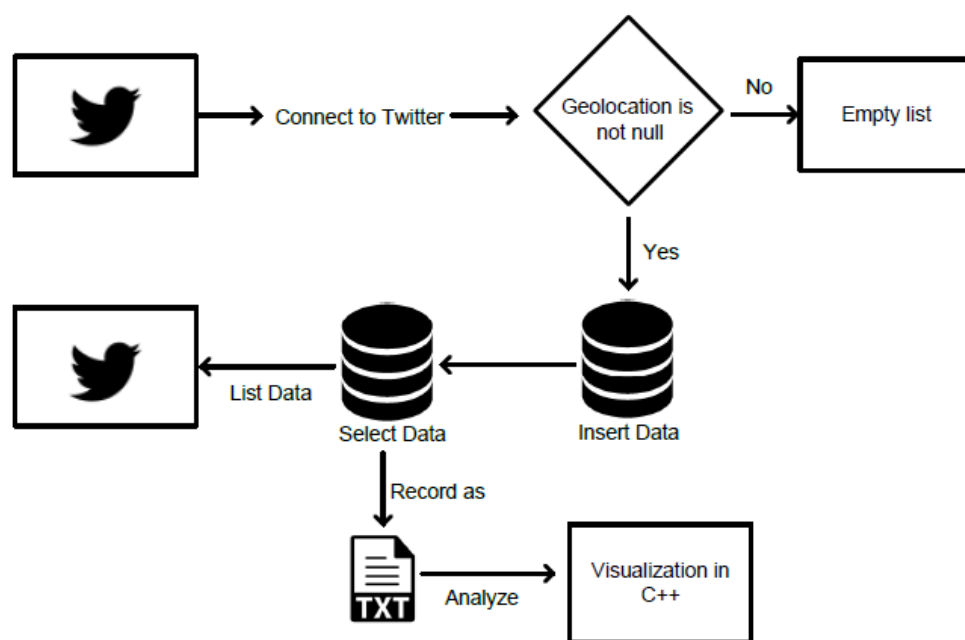


Figure 1. Flowchart of Twitter Data Visualization

As is seen above, flowchart expresses how to get, store and visualize data. After to connect Twitter with key and id, data is check. If geolocation is available, all data store in database, but if geolocation is null, list is always empty. When data is selected from database with appropriate query, either is listed at window or is converted and stored as txt file. Thereafter, txt file is read and data is edited by using VTK tool in C++.

- At first window, user enter the hashtag that he/she wants to search. Hashtag is searched in database, if name is in there, is listed at second window. If name is not in there and different hashtag is in database, previous hashtag and data are deleted and new information are inserted, as the following,

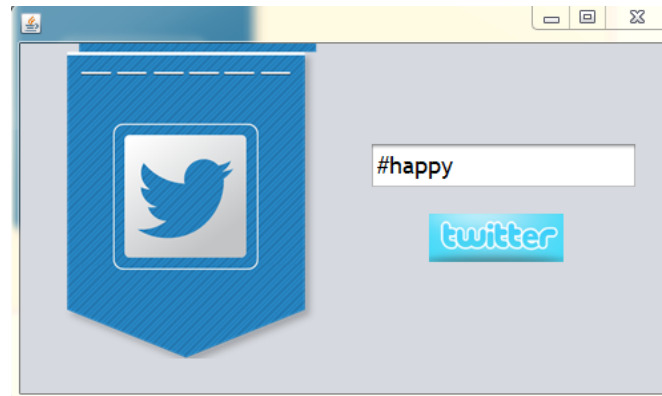


Figure 2. Hashtag Window

- At the second window, information that related to specific hashtag, are listed on the arrangeable table. As follows,

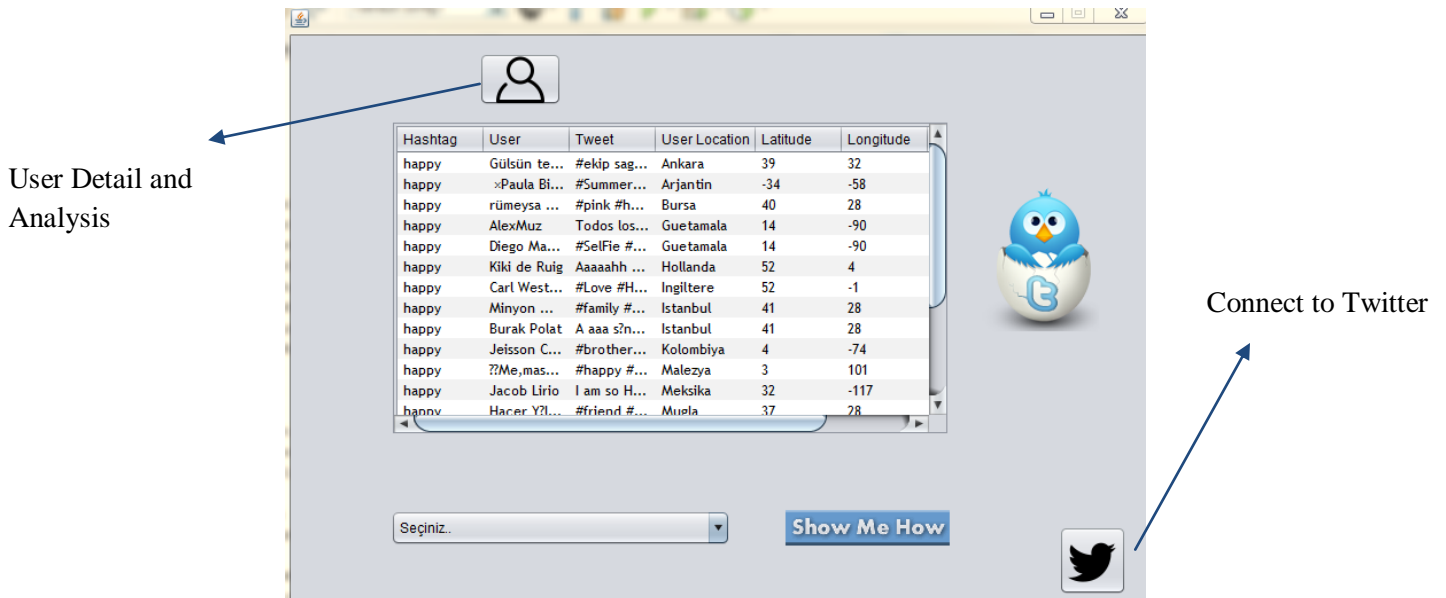


Figure 3. List Window

Generally, location of user is not exact in Twitter. Because of lack of this situation, we determined location line with the longitude and latitude (GeoLocaiton).

- Next stages, when ‘Show Me How’ button is pushed, different type of VTK Tool will be exhibited. Studies about this subject are continuing.
- After that, to get much more information about user, we created tab window as the following,

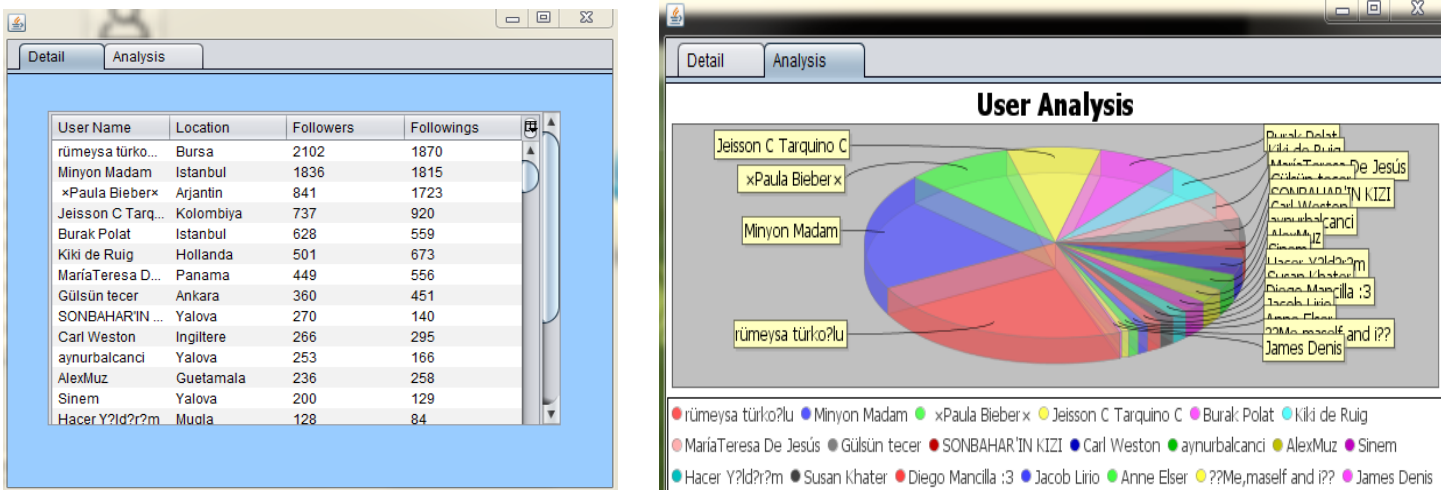


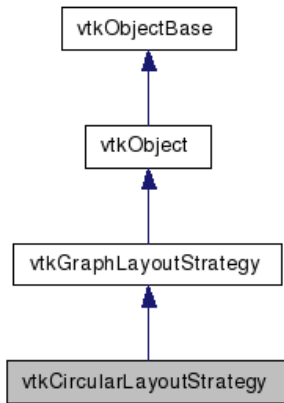
Figure 4. Detail and Analysis Window

D. Project Phases

- To be able to collect Twitter data, we plug in twitter .jar files.
- By using MySQL Workbench, we store data that we collect from Twitter and thanks to this database, we controlled data whether exist or not. Therefore, it helps to select, insert and delete Twitter data.
- When we call Twitter's Search API, we can retrieve up to 100 tweets at a line. But , by using different method, we can attain more than 100 tweets. So that, we can use these data for aim of testing and analyzing.
- We used structured of VTK in C++, because of limited source of java. Initially VTK is an open-source, freely available software system for 3D or 2D computer graphics, image processing and visualization.[9] Simultaneously, it includes data structures, graph algorithms, several layout algorithms [10]. We used to visualize for layout of VTK, as the following,

1. VTK Circular Layout
2. VTK Fast 2D Layout
3. VTK Force Directed Layout
4. VTK Random Layout with vtkGlyp3D

1. VTK Circular Layout



```

VTK_CREATE(vtkCircularLayoutStrategy, strategy);

VTK_CREATE(vtkGraphLayout, layout);
layout->SetInputConnection(edgeCategory->GetOutputPort());
layout->SetLayoutStrategy(strategy);

VTK_CREATE(vtkGraphMapper, mapper);
mapper->SetInputConnection(layout->GetOutputPort());
mapper->SetEdgeColorArrayName("edge category");
mapper->ColorEdgesOn();
mapper->SetVertexColorArrayName("vertex category");
mapper->ColorVerticesOn();
VTK_CREATE(vtkActor, actor);
actor->SetMapper(mapper);
  
```

This layout places vertices around a circle. It assigns points to the vertices around a circle with unit radius. [11]

When to read data from txt file, user's properties were visualized as movable circular. As the shown above,

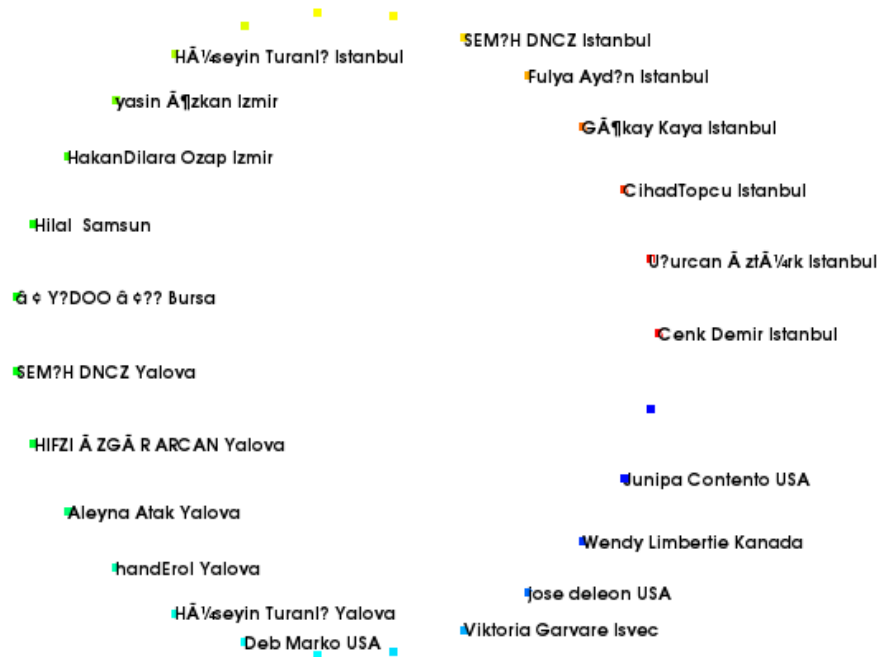
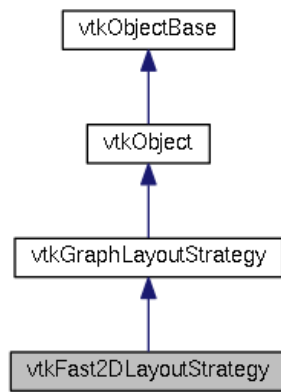


Figure 5. Twitter Data Visualization Using VTK Circular Layout

2. VTK Fast 2D Layout



```

VTK_CREATE(vtkFast2DLayoutStrategy, strategy);
VTK_CREATE(vtkGraphLayout, layout);
layout->SetInputConnection(edgeCategory->GetOutputPort());
layout->SetLayoutStrategy(strategy);

VTK_CREATE(vtkGraphMapper, mapper);
mapper->SetInputConnection(layout->GetOutputPort());
mapper->SetEdgeColorArrayName("edge category");
mapper->ColorEdgesOn();
mapper->SetVertexColorArrayName("vertex category");
mapper->ColorVerticesOn();
VTK_CREATE(vtkActor, actor);
actor->SetMapper(mapper);
  
```

VTK Fast 2D Layout is a simple fast 2D graph layout. This class is a density grid based force directed layout strategy. Also please note that 'fast' is relative to quite slow.[12]

When to read data from txt file, user's properties were visualized as movable points. As the shown,

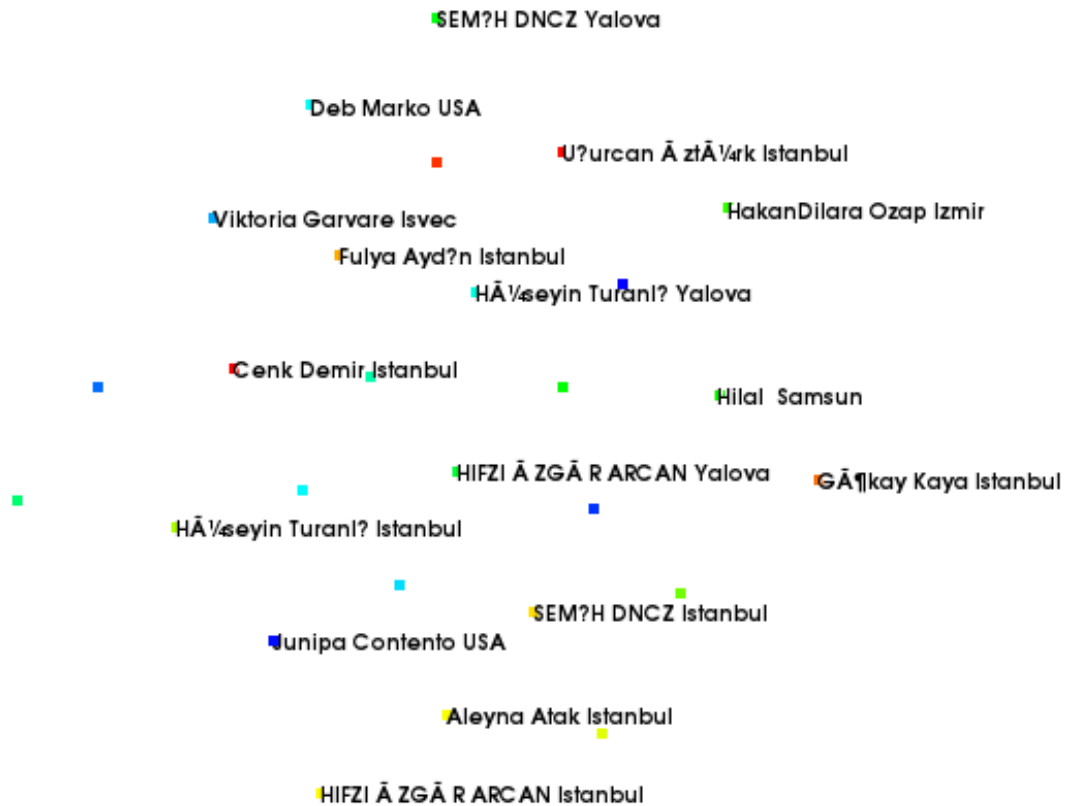
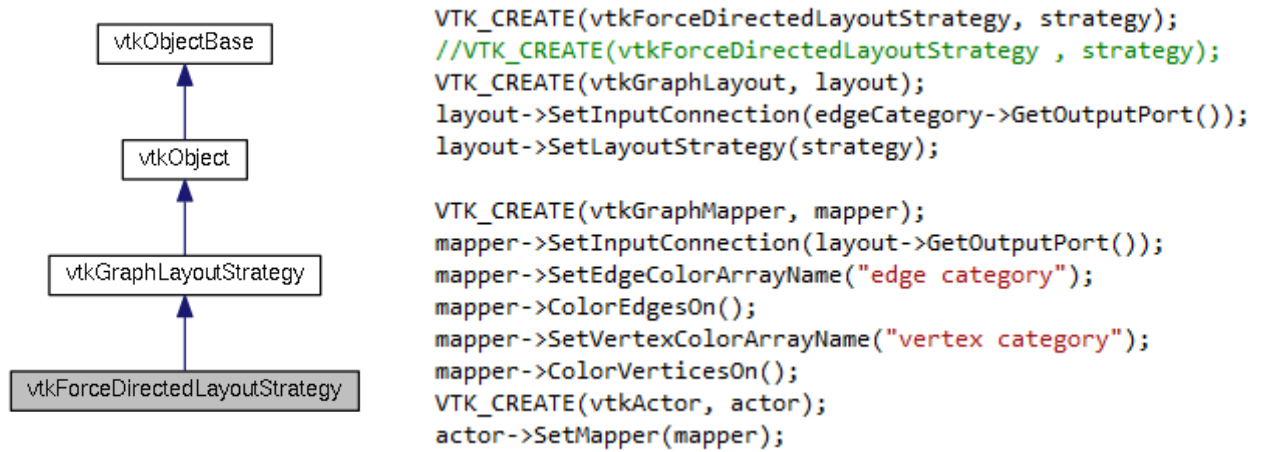


Figure 6. Twitter Data Visualization Using VTK Fast 2D Layout

3. VTK Force Directed Layout



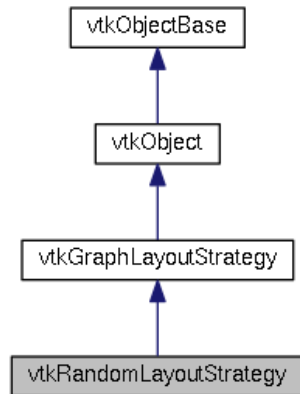
VTK Force Directed Layout is a force directed graph layout algorithm. Lays out a graph in 2D or 3D using a force-directed algorithm. The user may specify whether to layout the graph randomly initially, the bounds, the number of dimensions (2 or 3), and the cool-down rate.[13]

When to read data from txt file, user's properties were visualized as movable points. As the shown,



Figure 7. Twitter Data Visualization Using VTK Force Directed Layout

4. VTK Random Layout with vtkGlyp3D



```

VTK_CREATE(vtkRandomLayoutStrategy, strategy);

VTK_CREATE(vtkGraphLayout, layout);
layout->SetInputConnection(edgeCategory->GetOutputPort());
layout->SetLayoutStrategy(strategy);
  
```

VTK Random Layout is randomly places vertices in 2 or 3 dimensions. Assigns points to the vertices of a graph randomly within a bounded range. [14]

When to read data from txt file, user's properties were visualized as movable stuctured. As the shown,

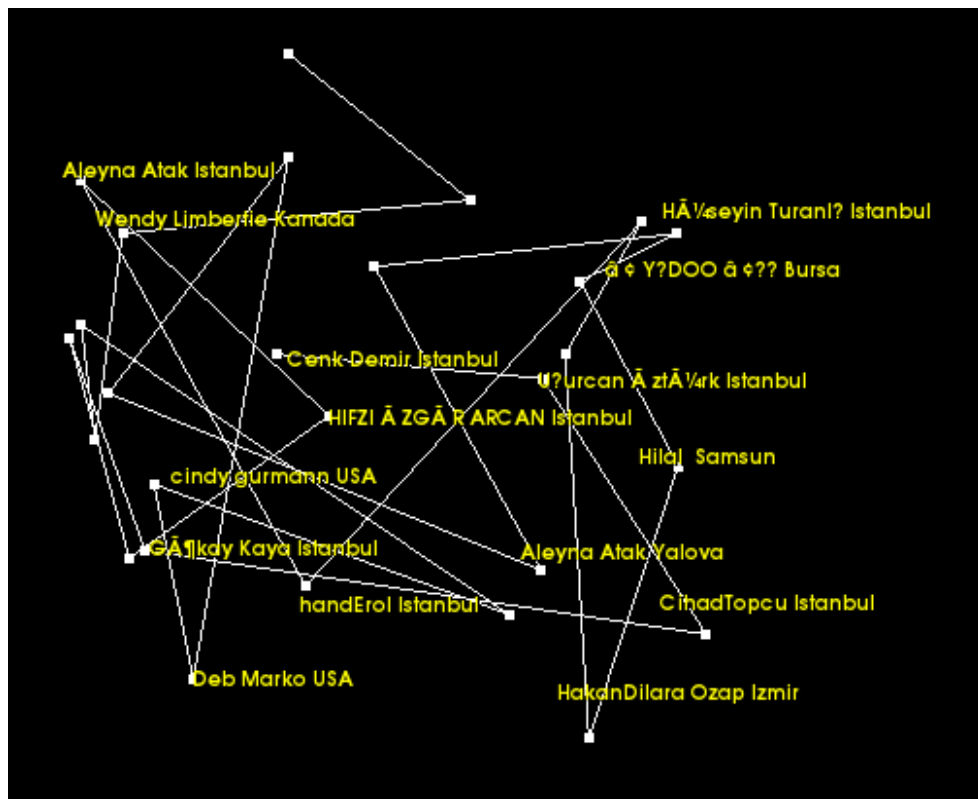


Figure 8. Twitter Data Visualization Using VTK Random Layout

E. Results

As a conclusion VTK provides a different data representations including unorganized point sets, polygonal data, images, volumes, and structured, rectilinear, and unstructured grids. And we used that the model used for 3D polygonal rendering (lights, cameras, actors) is analogous to the model used for volume and other types of rendering (lights, cameras, volumes). [15]

4. DISCUSSION AND CONCLUSIONS

Twitter is one of the most popular social media platform for the people to comment on the very popular topics or the common interests. Most of the tweets posted are reflecting the mental status of a person. As a result, many tools have been developed to observe the activity of the persons acting on twitter. Finally, we visualised these data using Visualization ToolKit(VTK).

References:

- [1] Alexander Pak, Patrick Paroubek. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. Universit'e de Paris-Sud, France.
- [2] Tamara A. Small. WHAT THE HASHTAG? A content analysis of Canadian politics on Twitter.
- [3] Creating a Tweet Visualization in 3-D, <https://blog.twitter.com/2014/creating-a-realtime-tweet-visualization-in-3-d>
- [4] Data Visualization Tools Resources, <http://www.hongkiat.com/blog/data-visualization-tools-resources/>
- [5] Election2014: The Web of Influence on Twitter, <https://interactive.twitter.com/election2014/>
- [6] Agile Modeling, http://en.wikipedia.org/wiki/Agile_modeling
- [7] Boehm, B. Get ready for agile methods, with care. Univ. of Southern California, Los Angeles, CA, USA
- [8] JFreeChart, <http://en.wikipedia.org/wiki/JFreeChart>
- [9] Visualization ToolKit, <http://vtk.org/>
- [10] Christian Bachmaier, Franz J. Brandenburg, Michael Forster, Paul Holleis, Marcus Raitner, Gravisto: Graph Visualization Toolkit
- [11] vtkCircularLayoutStrategy Class Reference, <http://www.vtk.org/doc/release/5.2/html/a00192.html>
- [12] vtkFast2DLayoutStrategy Class Reference, <http://www.vtk.org/doc/nightly/html/classvtkFast2DLayoutStrategy.html#details>
- [13] vtkForceDirectedLayoutStrategy Class Reference, <http://www.vtk.org/doc/nightly/html/classvtkForceDirectedLayoutStrategy.html>
- [14] vtkRandomLayoutStrategy Class Reference <http://www.vtk.org/doc/nightly/html/classvtkRandomLayoutStrategy.html>
- [15] William J. Schroeder, Lisa S. Avila, and William Hoffman, Visualizing with VTK: A Tutorial