

DeepDepth
An Approach To Visualize Twitter Tweets

AmirSaber Sharifi

October 2014

Contents

List of Figures

Chapter 1

Introduction

1.1 Social Networks

WY: I like what you wrote a lot better. But it is useful to have a formal definition of social networks that you will use throughout your project. Maybe start your intro with the following paragraphs.

And then write a paragraph about what are social networks used for. Creating online communities of common interests, to share information in a community, ...? Use Twitter an example here. How do people share information? How do people create a community.

And then write a paragraph about the importance of social networks from the organizations' perspective. Use Twitter here again as an example. Give stats on how many tweets are there. How it can be useful, etc.

WY: I want you to incorporate this information with the above section. So, combine Sections 1.1 and 1.2 together.

- Explain why Twitter is important
- What does it do?
- What can we do with these data?
- Give examples or use cases for how & why analyzing its data is important
- Briefly define social networks and their usages

A *social network* is a structure of relationships linking *social actors* (Marsden 2000:2727) or the set of actors and the ties among them (Wasserman and Faust 1994). Relationships or ties are the basic building blocks of human experience, mapping the connections that individuals have to one another (Pescosolido 1991)[?]. Social Networks are typically defined as a social structure made up of a set of social actors, who can be individuals or or-

ganizations, and a set of *dyadic ties* between these actors. Being friends, relatives or colleges are examples of possible dyadic ties.

done

- First two sentences above: They are too informal. Remove them.
- Third sentence: formally define social networks by including a reference for it.
- Fourth sentence: Before this sentence, you need to define what does “social actors” mean.

As the world becomes more interconnected through the *World Wide Web*, social networks are becoming more relevant through World Wide Web. When these relations happen through World Wide Web they are called *Social Web*. In general, Social Web is a set of relationships that link people over the World Wide Web[?]. Social Web covers the design and develop of software or web pages that encourage social interactions using Word Wide Web. Examples of some Social Web implementation can be games, education and social networking websites. Facebook, Google Plus, Tumbler and Twitter can be named as examples of social networking websites.

Done “world wide web” is first used before it is italicized. Also, keep consistent. Either the first letters are always capitalized or always lower case. Same comment for “Social Web”. And when you use “Web,” it is unclear which one you are referring to.

Depend on the social networking website, social actors can be individuals, organizations and groups and social dyadic ties between these actors also depends on the social networking website. As an example in Twitter, both individuals and organizations can have accounts and they can interact with other actors by following others, posting a message, replying to other actor messages and re-post a message that has been originally posted by other actor.

These social networking websites can get benefits to individuals and organizations. As an individual, making friends, chatting, posting photos and having social interactions in general can be expected as a benefit of a social networking website. Creating online communities so users with common interests can join and socialized around the topic or find old school friends on these websites are some other benefits of social web for individuals.

On the other hand organization can use social networking websites to reach their audiences easier. They can establish communities and pages to reach more audiences and even use smart advertisement tools to advertise their message to specific group of people. By looking at trends and hit rate of their posts they can make better decisions for their future activities or products. These are some examples of benefits for organizations while a lot

more can be imagined using social network websites and related analytic tools designed for them.

Each social network could be analyzed with different perspectives. One can be analyzing social network structures to identify patterns and examine network dynamics. The other can be analyzing social network data. These data can be the message that a user posts or a color they choose as background or even time of the day users are active on social networks.

Twitter is one of the famous social networking website. Users in Twitter can read and post 140 character short messages that are called *tweets*. Twitter provide various ways of posting a tweet like using their web site, application or text messages. It's established on March 2006 and since then it grows fast between users[?]. There are about 284 million active users on Twitter. Users on Twitter post average of 6,000 tweets per second which correspond to 350,000 tweets per minute and 500 million tweets per day[?]. Twitter also supports more than 35 languages. This is a substantial amount of data that can be used in different analyses.

Done Add reference to sentences with stats.. like 284 million active suers, 6000 tweets per second, etc. References should not be footnotes, but they should be proper references. Talk to me if you are unclear.

Twitter provides a substantial amount of information of each tweet. Information like user number of tweets, user profile color, language and etc. These information can be useful also. One example can be looking at relation between profile color of users and their tweets about committing suicides. By implementing an appropriate tool one can analyze Twitter tweets, or any other social networking websites, in different perspectives and get useful and meaningful results. For instance by watching and monitoring tweets which contains symptoms of a disease over time, origin of a disease could be found or even prevent the next possible epidemic.

1.2 DeepDepth

- Explain DeepDepth goals
- Why this project is important?

Our goal in this master project is to design a system called DeepDepth to help users analyze social network data without having vast knowledge of computer science. Users should be able to login, request their process and get the result. Since using social network websites is growing, various entities such as researchers, laboratories and industries found social network websites a good resource of information. Noticing the importance of social network websites on our life and benefits that a good analytic tool could give us, idea of DeepDepth shaped.

Not sure!! Formalize say that your goal in this MS project is to design a system, called DeepDepth, to blablabla...

DeepDepth is an application that built to cover a part of social network websites analytic. It's goal is to provide a platform for developers, administrators and users to expand usage of social network data. By using DeepDepth users, without having a specific knowledge of computer, can initiate tailored queries and get visualized meaningful results. DeepDepth is modular, expandable and manageable by administrator which means an administrator or a developer can add query types, graphs or even new social network source into the platform. In the first iteration of the platform, Twitter is considered as the main source of data, although other sources can be added to project later.

1.3 Related works

- Introduce other similar systems
- How this project differs from them

Here are some other similar systems available on web that do analytic on social network data. Some of those systems are as follow:

Orgnet.com SNA Social Network Analysis (SNA) is a tool that does mapping and measuring of relationships and flows between actors. It provides mathematical and visual analysis of social network and calculate different degrees for each node such as betweenness and closeness.[?]

Twazzup A monitoring tool to show users each time their keyword is mentioned in a tweet.[?]

Twitscoop was a real-time visualization tool that could tank tweet words based on how frequently they are used.[?]

TweetPsych is a web application that uses linguistic analysis algorithms to create a profile for a person based on what they have been tweeted on Twitter.[?]

Tweeps analyzes the content of user's tweets and make statistic for that specific user. This service is currently down.[?]

Twitonomy analyzes activity of a user in twitter. It can gives some general numbers like average number of tweets per day or users most mentioned and active hours and days. [?]

Most twitter Monitoring and Analytics tool are not working any more because of Twitter change in their API policy which doesn't let third party apps to have access to all tweets using their API. Companies which want to have access to all tweets should go into a contract with Twitter[?].

DeapDepth is different from these examples in essence that DeepDepth is more general. It's a platform that administrator of it can add graphs, query types and etc so users will be able to use different kinds of analytic and visualizations. It's more dynamic that means DeepDepth can also get connected to different data sources and do analyzes on them. These data sources can be Facebook, Google+ or any other social network in future.

1.4 Challenges

- What are the challenges that the project faced
- How the project tackled those challenges

To implement DeepDepth there were some challenges. This challenges happened in different aspect of project. For instance in gathering data, storing data, processing data and visualizing the result.

The first step in analyzing a data is the data and gather it. Since DeepDepth is a modular application, it can use any Hive database as it source of analyzes. For Twitter scenario gathering data has been done by using Twitter Official client for developers which is called HBC. It's a client for a stream API that receive a sample of all tweets. Since HBC should be an always running process it's running in a PaaS named Heroku[?]. Heroku is also the service that help storing the data using its Treasure Data[?] api.

Processing data was a challenge that could be handled in various ways. One can suggest using Apache Storm[?] which results in a real time analysis. Because of required resources to do real time processing, DeepDepth is using an on demand processing. MapReduce is the technique that is used to process a Hive database entries.

After the process phase results need to be visualized. To be able to have a modular application a framework of drawing charts and graphs should be considered so administrator and developers can later add graphs to the project easily. So in DeepDepth Visualizing data part is implemented using D3.JS[?] library which is a graph framework that is able to draw specific charts and graphs and also can be programmed to draw custom graphs and charts. By using D3.Js a lot of ready to use charts are available to use in DeepDepth and developer or administrator can also add other graph types to it.

Chapter 2

DeepDepth

- Implementation of server and client sides using Test Driven Development

2.1 High-Level Description

2.1.1 Components

WY: How about combining Chapters 2 and 3 and call it DeepDepth. Then, you have a section called "High-Level Description", which is your Section 2.1 now, and then you have another section called "Implementation Details" which has the subsections on Server and Client. Then, within these subsections, you briefly describe the technology that you use, why you chose to use it, and how you use it.

- What are the programmer work space and environment provided for this project.
- What are the components in this project and briefly describe them
- How those components are related and how they interact
- Work flow Diagram
- Data flow Diagram
- Use Cases, what are user types and how different users see the project and how they use this software

2.1.2 Technologies

- Describe the technology that have been used in this project for each component
- More detail into each component and how they interact with each other

2.2 Implementation

2.2.1 Server

- What is the server side and what it is doing
- How components in server side have been implemented

2.2.1.1 Environments

- What are the development environments (Production, Development and Test)
- How they have been implemented in this project
- Examples of codes

2.2.1.2 Models

- What are Models in this project
- how they are related
- examples of implementation

2.2.1.3 Controllers

- what are the controllers
- how they have been implemented

2.2.1.4 Routes

- What are server routes
- How my code handle routes

2.2.1.5 Tasks

- how tasks are implemented in deepdepth

2.2.1.6 Views

- implementation of views

2.2.1.7 Tests

- Test types implemented for server side
- how they are implemented

2.2.2 Client

- Client side implementation

2.2.2.1 Modules

- what are the modules on client side

2.2.2.2 Config

- how configuration of modules programmed

2.2.2.3 Services

- describe client side services implemented for models

2.2.2.4 Views

- implementation of CRUD views for different modules

2.2.2.5 Tests

- Describe Tests implemented for client side

Chapter 3

Summary and Future Works

WY: This part is definitely needed. It basically summarizes what your system is about, why you think it is cool, and you can add future work here. Also, List of Figures is usually before Chapter 1 right in front of the document.

- Talking about summary of project and it's achievements(I am not sure if this part is needed or what should I put in here.)
- How this projects can be improved in future
- nice to have features

Bibliography

- [1] BERNICE A. PESCOLIDIO http://www.uk.sagepub.com/leonguerrero4e/study/materials/reference/05434_socnet.pdf
THE SOCIOLOGY OF SOCIAL NETWORKS
- [2] Halpin, Harry; Tuffield, Mischa. "A Standards-based, Open and Privacy-aware Social Web" W3C Social Web Incubator Group Report 6th December 2010 Report. W3C Incubator Group Report. Retrieved 6/8/2011
- [3] <http://www.internetlivestats.com/twitter-statistics/>
- [4] <https://about.twitter.com/company>
- [5] <http://www.orgnet.com/sna.html>
- [6] <http://www.twazzup.com/>
- [7] <http://twitscoop.com/>
- [8] <http://tweetpsych.com/>
- [9] <http://www.tweeps.com/>
- [10] <http://www.twitonomy.com/>
- [11] <http://www.brightplanet.com/2013/06/twitter-firehose-vs-twitter-api-whats-the-difference-and-why-should-you-care/>
- [12] <https://www.heroku.com/>
- [13] <http://www.treasuredata.com/>
- [14] <https://storm.apache.org/>
- [15] <http://d3js.org/>