



Head Node Detection Using Graph Mining

NETWORK SCIENCE

UNDER

PROFESSOR CHARALAMPOS CHELMIS

MAY 04, 2017

THURSDAY

Presenters

2

DISHA PATEL

M.S. in Computer
Science

dpatel3@albany.edu

MANSI PATEL

M.S. in Computer
Science

mnpatel@albany.edu

BHAVIK PATEL

M.S. in Computer
Science

bpatel2@albany.edu

Outline

3

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

Introduction

- ▶ Get to the roots of everything.
- ▶ Awareness: Social media is not about the exploitation of technology but service to community.
- ▶ Find root of any cause, issue, topic, etc.
- ▶ Eg: Fake News Detection
 - ▶ To dig out the source of any fake news.

Outline

5

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

Applications

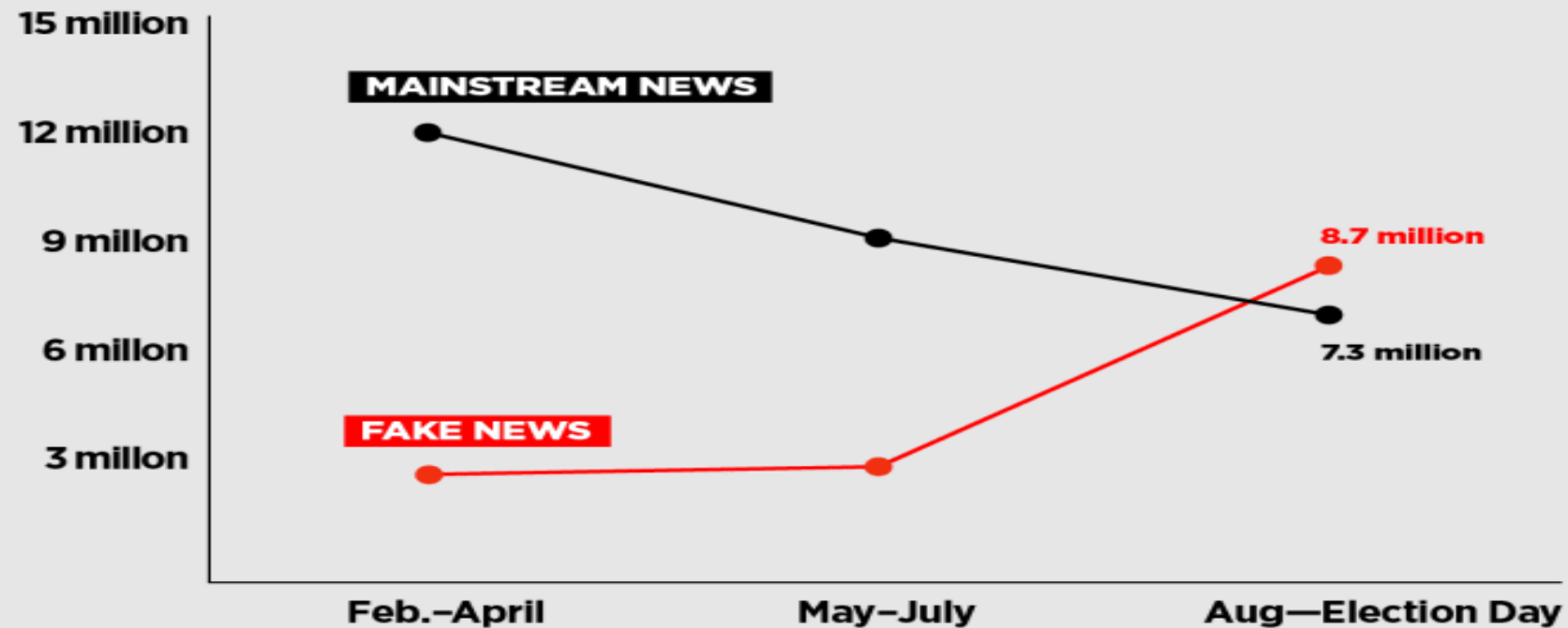
6

- ▶ Detecting root web page for networks of fraudulent/rogue websites.
- ▶ Information Diffusion/Fake News Detection
- ▶ To detect suspicious events in Telecommunication Networks.
- ▶ Detection of Terrorist Groups in Online Social Networks.
- ▶ Cancer Detection

Survey(1/2)

7

Total Facebook Engagements for Top 20 Election Stories



ENGAGEMENT REFERS TO THE TOTAL NUMBER OF SHARES, REACTIONS, AND COMMENTS FOR A PIECE OF CONTENT ON FACEBOOK SOURCE: FACEBOOK DATA VIA BUZZSUMO

<https://1.bp.blogspot.com/-T78BMYUp70/WC9VHhYE8gl/AAAAAAAAADl8/Mh1qdRqf6scib4hAEIj904pL7GwvrVvwCLcB/s1600/fake%2Bnews%2Bversus%2Breal.jpg>

Survey(2/2)

8

- ▶ A day after a black activist was kicked and punched by voters at a Donald Trump rally in Alabama, Trump tweeted an image packed with racially loaded and incorrect murder statistics.



Says crime statistics show blacks kill 81 percent of white homicide victims.

— Donald Trump on Sunday, November 22nd, 2015 in a tweet

USA CRIME STATISTICS ~ 2015



BLACKS KILLED BY WHITES ~~ 2%
BLACKS KILLED BY POLICE ~~ **1%**
WHITES KILLED BY POLICE ~~ 3%
WHITES KILLED BY WHITES ~~ 16%
WHITES KILLED BY BLACKS ~~ 81%
BLACKS KILLED BY BLACKS ~~ **97%**

SOURCE: Crime Statistics Bureau - San Francisco

Outline

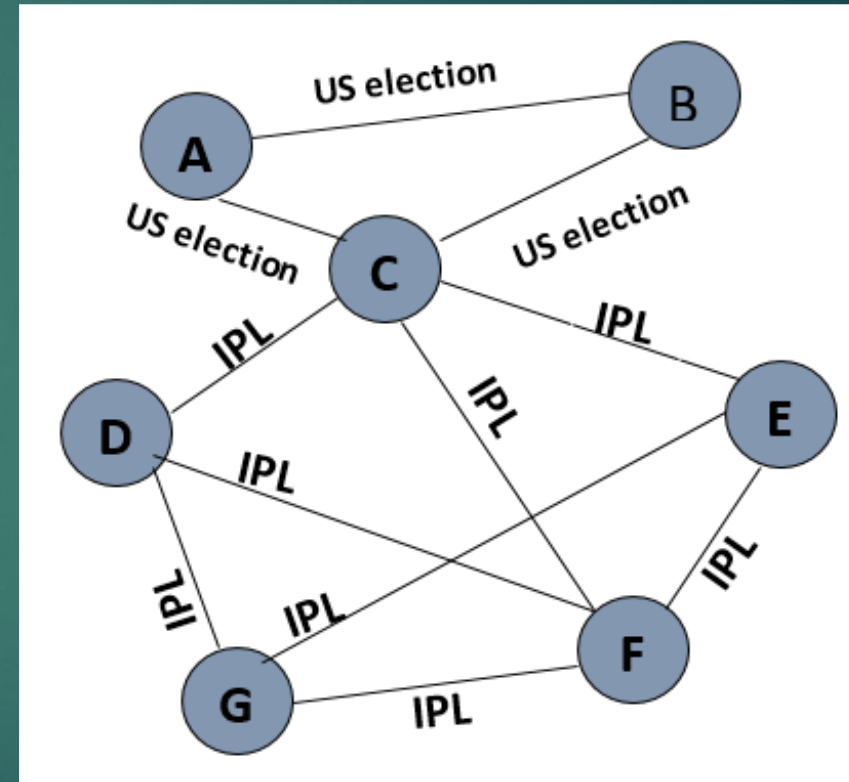
9

- ▶ Introduction
- ▶ Applications
- ▶ **Proposed Approach**
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

Proposed Approach

10

- ▶ 2 step Method to be performed on network:
 1. **Community detection** algorithm.
 2. **Back track the source** from that specific Community
- ▶ This approaches mean to first confine the search by logic and than apply the search.



Outline

11

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ **Data Set**
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

Dataset(1/2)

12

Twitter Dataset:

No.	Column Name	Format Example	Description
1.	Tweet_id	32327382	Id of tweet
2.	Tweet_time	Sat May 16 23:58:44 UTC 2009	timestamp of tweet
3.	User_name	rocketscience575	Username who tweeted
4.	Tweet	Obama is visiting istanbul today, therefore all main roads have been closed cause and effect!!!	Tweet text

Dataset(2/2)

13

- ▶ Dataset (in csv format) contains tweets with the following information:
 - ▶ Polarity of the tweet(0= negative, 2= neutral, 4= positive)
 - ▶ Id of the tweet
 - ▶ Data of the tweet
 - ▶ User Name
 - ▶ Text of the tweet.
- ▶ The tweets in the data set are from the time period between: April 6, 2009 to June 25, 2009.

Outline

14

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ **Implementation**
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

Implementation(1/)

15

- ▶ Treat each User as an individual node.
- ▶ An edge would be formed between two users if they have tweeted on a common topic.
- ▶ Approach:
 - ▶ Community Detection
 - ▶ Back tracking
- ▶ In particular community back track nodes(users) via timestamp of the tweets.

Outline

16

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

DEMO

Dataset

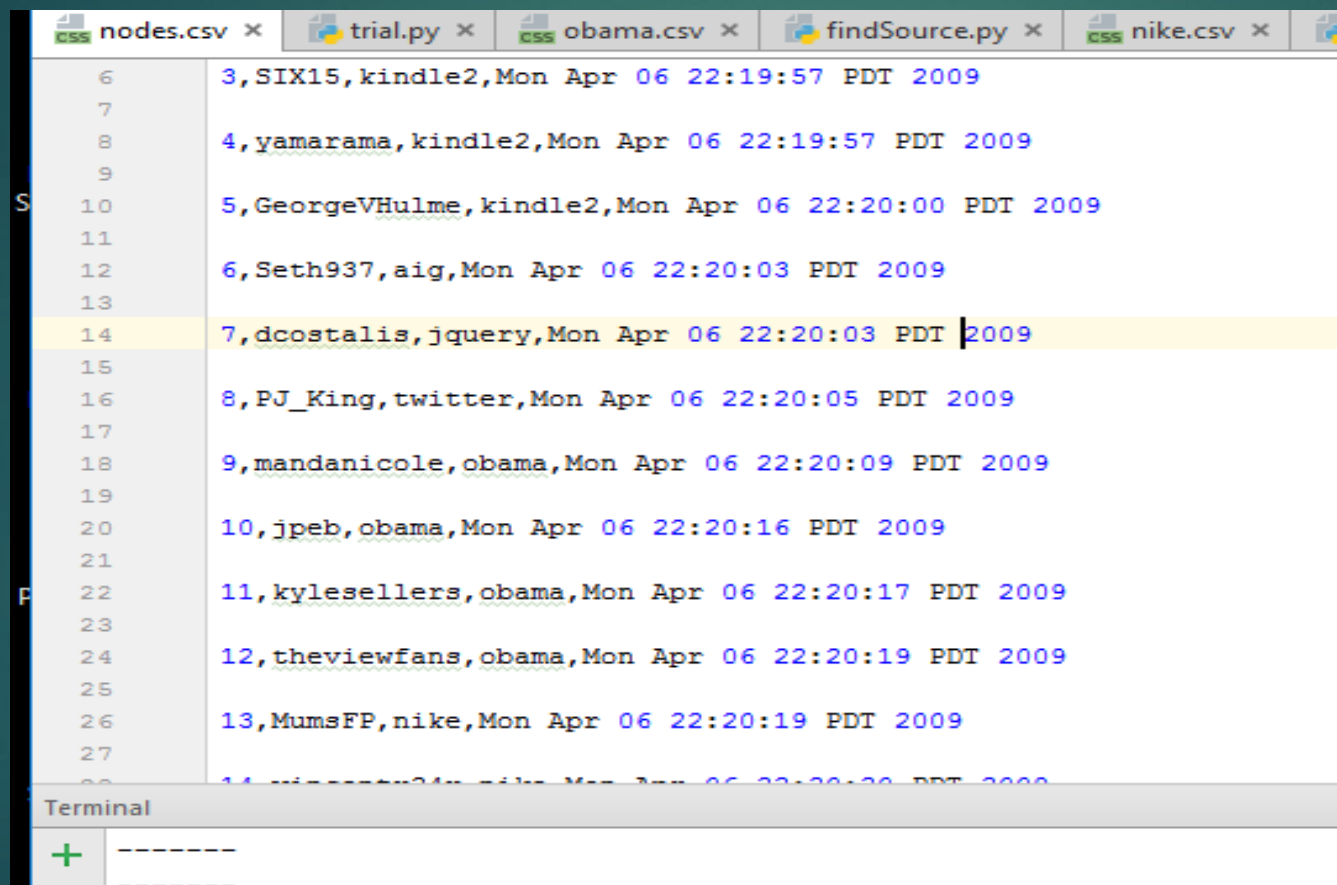
18

A	B	C	D	E	F	G
0	1467810369	Mon Apr 06 22:19:45	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda	
0	1467810672	Mon Apr 06 22:19:49	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by texting it... and might cry as a res	
0	1467810917	Mon Apr 06 22:19:53	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Managed to save 50% The rest go c	
0	1467811184	Mon Apr 06 22:19:57	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire	
0	1467811193	Mon Apr 06 22:19:57	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all. i'm mad. why am i here? because	
0	1467811372	Mon Apr 06 22:20:00	NO_QUERY	joy_wolf	@Kwesidei not the whole crew	
0	1467811592	Mon Apr 06 22:20:03	NO_QUERY	mybitch	Need a hug	
0	1467811594	Mon Apr 06 22:20:03	NO_QUERY	coZZ	@LOLTrish hey long time no see! Yes.. Rains a bit ,only a bit LOL , i'm fine than	
0	1467811795	Mon Apr 06 22:20:05	NO_QUERY	2Hood4Hollywood	@Tatiana_K nope they didn't have it	
0	1467812025	Mon Apr 06 22:20:09	NO_QUERY	mimismo	@twittera que me muera ?	
0	1467812416	Mon Apr 06 22:20:16	NO_QUERY	erinx3leannexo	spring break in plain city... it's snowing	
0	1467812579	Mon Apr 06 22:20:17	NO_QUERY	pardonlauren	I just re-pierced my ears	
0	1467812723	Mon Apr 06 22:20:19	NO_QUERY	TLeC	@caregiving I couldn't bear to watch it. And I thought the UA loss was embarra	
0	1467812771	Mon Apr 06 22:20:19	NO_QUERY	robobbierobert	@octolinz16 It it counts, idk why I did either. you never talk to me anymore	
0	1467812784	Mon Apr 06 22:20:20	NO_QUERY	bayofwolves	@smarrison i would've been the first, but i didn't have a gun. not really thoug	
0	1467812799	Mon Apr 06 22:20:20	NO_QUERY	HairByJess	@iamjazzyfizzle I wish I got to watch it with you!! I miss you and @iamlilnicki I	
0	1467812964	Mon Apr 06 22:20:22	NO_QUERY	lovesongwriter	Hollis' death scene will hurt me severely to watch on film wry is directors cut r	
0	1467813137	Mon Apr 06 22:20:25	NO_QUERY	armotley	about to file taxes	
0	1467813579	Mon Apr 06 22:20:31	NO_QUERY	starkissed	@LettyA ahh ive always wanted to see rent love the soundtrack!!	
0	1467813782	Mon Apr 06 22:20:34	NO_QUERY	gi_gi_bee	@FakerPattyPattz Oh dear. Were you drinking out of the forgotten table drink:	
0	1467813985	Mon Apr 06 22:20:37	NO_QUERY	quanvu	@alydesigns i was out most of the day so didn't get much done	
0	1467813992	Mon Apr 06 22:20:38	NO_QUERY	swinspeedx	one of my friend called me, and asked to meet with her at Mid Valley today...k	

training.1600000.processed.noem

Nodes Collection

19

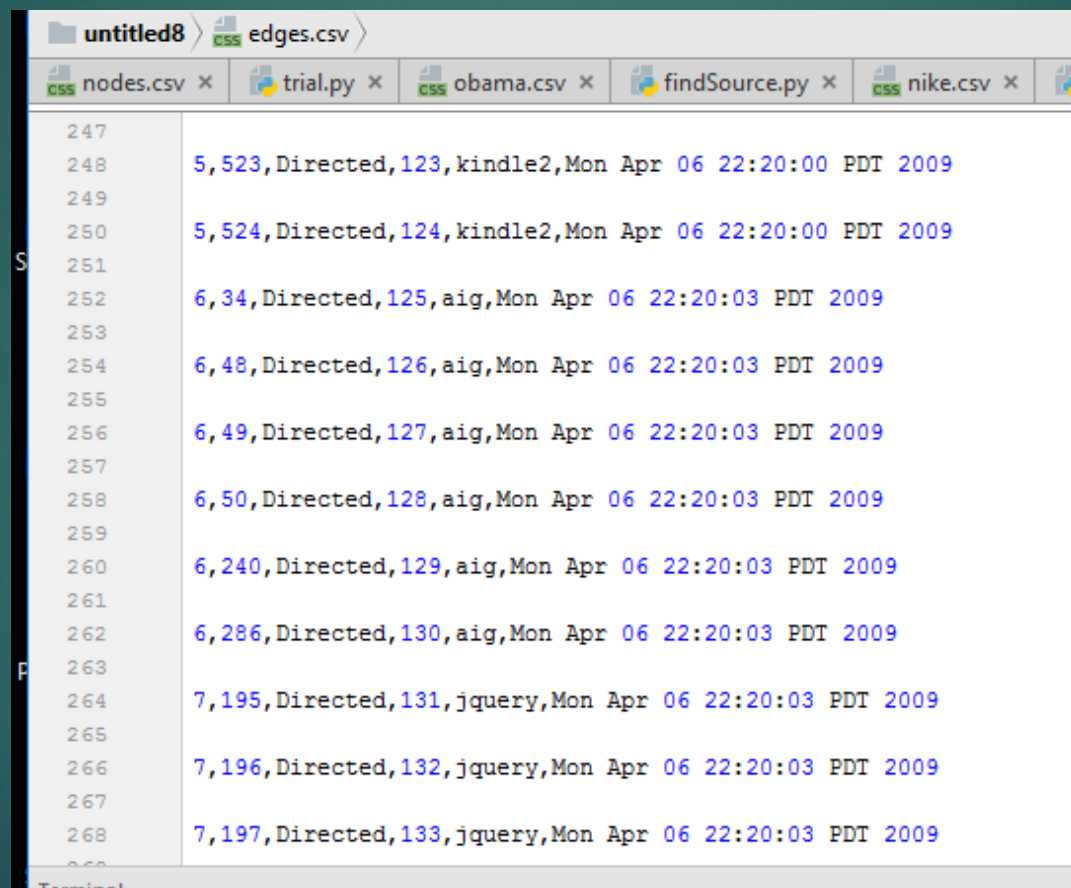


```
nodes.csv x trial.py x obama.csv x findSource.py x nike.csv x
6 3,SIX15,kindle2,Mon Apr 06 22:19:57 PDT 2009
7
8 4,yamarama,kindle2,Mon Apr 06 22:19:57 PDT 2009
9
10 5,GeorgeVHulme,kindle2,Mon Apr 06 22:20:00 PDT 2009
11
12 6,Seth937,aig,Mon Apr 06 22:20:03 PDT 2009
13
14 7,dcostalis,jquery,Mon Apr 06 22:20:03 PDT 2009
15
16 8,PJ_King,twitter,Mon Apr 06 22:20:05 PDT 2009
17
18 9,mandanicole,obama,Mon Apr 06 22:20:09 PDT 2009
19
20 10,jpeb,obama,Mon Apr 06 22:20:16 PDT 2009
21
22 11,kylesellers,obama,Mon Apr 06 22:20:17 PDT 2009
23
24 12,theviewfans,obama,Mon Apr 06 22:20:19 PDT 2009
25
26 13,MumsFP,nike,Mon Apr 06 22:20:19 PDT 2009
27
28 14,wisconsin34,nike,Mon Apr 06 22:20:20 PDT 2009

Terminal
+ -----
+ -----
```

Edges collection

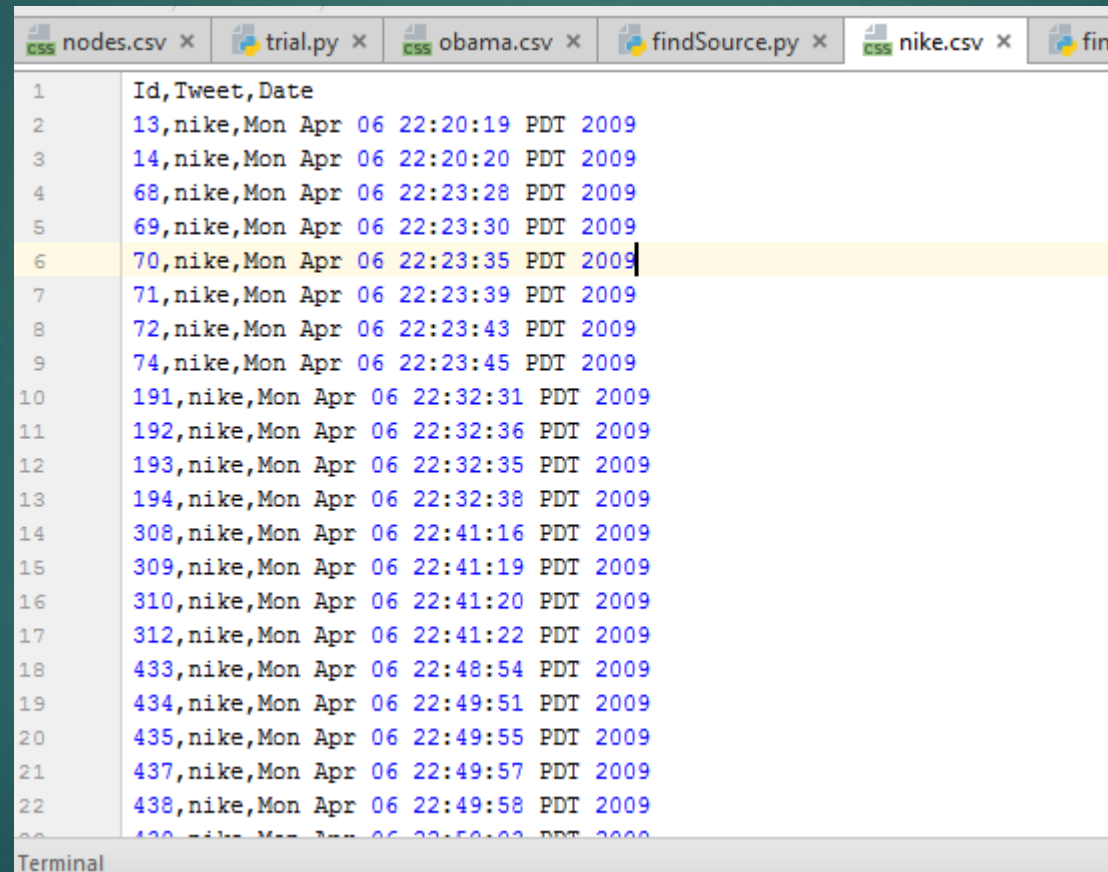
20



```
untitled8 > edges.csv
nodes.csv x trial.py x obama.csv x findSource.py x nike.csv x
247
248 5,523,Directed,123,kindle2,Mon Apr 06 22:20:00 PDT 2009
249
250 5,524,Directed,124,kindle2,Mon Apr 06 22:20:00 PDT 2009
251
252 6,34,Directed,125,aig,Mon Apr 06 22:20:03 PDT 2009
253
254 6,48,Directed,126,aig,Mon Apr 06 22:20:03 PDT 2009
255
256 6,49,Directed,127,aig,Mon Apr 06 22:20:03 PDT 2009
257
258 6,50,Directed,128,aig,Mon Apr 06 22:20:03 PDT 2009
259
260 6,240,Directed,129,aig,Mon Apr 06 22:20:03 PDT 2009
261
262 6,286,Directed,130,aig,Mon Apr 06 22:20:03 PDT 2009
263
264 7,195,Directed,131,jquery,Mon Apr 06 22:20:03 PDT 2009
265
266 7,196,Directed,132,jquery,Mon Apr 06 22:20:03 PDT 2009
267
268 7,197,Directed,133,jquery,Mon Apr 06 22:20:03 PDT 2009
```

Community Network Detected for a user input(tweet tag)

21

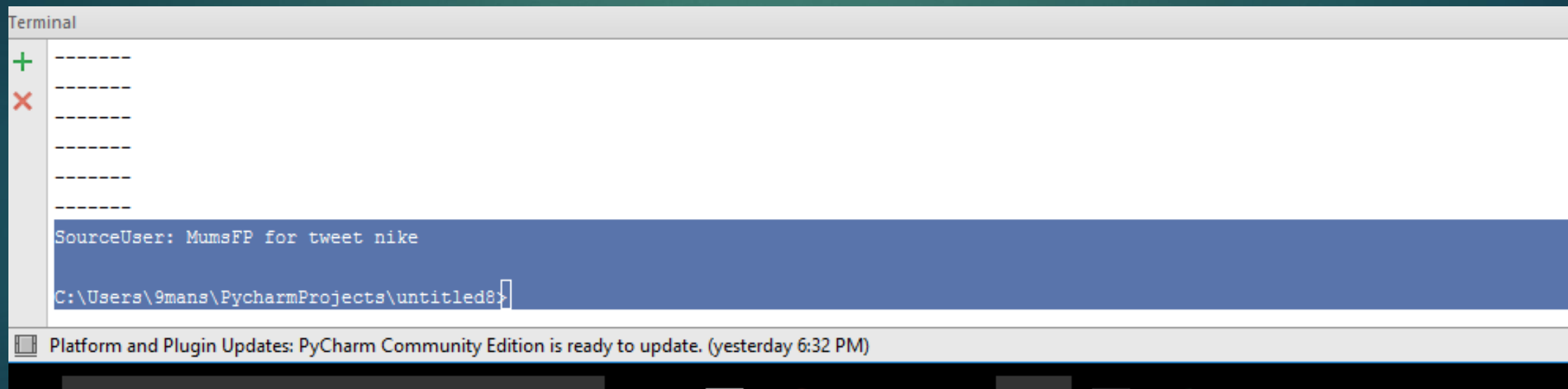


The screenshot shows a web browser window with several tabs open: 'nodes.csv', 'trial.py', 'obama.csv', 'findSource.py', 'nike.csv', and 'find'. The 'nike.csv' tab is active, displaying a table of tweet data. The table has three columns: 'Id', 'Tweet', and 'Date'. The data is sorted by date, showing tweets from April 6, 2009. The row with ID 70 is highlighted in yellow.

Id	Tweet	Date
13	nike	Mon Apr 06 22:20:19 PDT 2009
14	nike	Mon Apr 06 22:20:20 PDT 2009
68	nike	Mon Apr 06 22:23:28 PDT 2009
69	nike	Mon Apr 06 22:23:30 PDT 2009
70	nike	Mon Apr 06 22:23:35 PDT 2009
71	nike	Mon Apr 06 22:23:39 PDT 2009
72	nike	Mon Apr 06 22:23:43 PDT 2009
74	nike	Mon Apr 06 22:23:45 PDT 2009
191	nike	Mon Apr 06 22:32:31 PDT 2009
192	nike	Mon Apr 06 22:32:36 PDT 2009
193	nike	Mon Apr 06 22:32:35 PDT 2009
194	nike	Mon Apr 06 22:32:38 PDT 2009
308	nike	Mon Apr 06 22:41:16 PDT 2009
309	nike	Mon Apr 06 22:41:19 PDT 2009
310	nike	Mon Apr 06 22:41:20 PDT 2009
312	nike	Mon Apr 06 22:41:22 PDT 2009
433	nike	Mon Apr 06 22:48:54 PDT 2009
434	nike	Mon Apr 06 22:49:51 PDT 2009
435	nike	Mon Apr 06 22:49:55 PDT 2009
437	nike	Mon Apr 06 22:49:57 PDT 2009
438	nike	Mon Apr 06 22:49:58 PDT 2009
439	nike	Mon Apr 06 22:50:02 PDT 2009

Found Source Node

22



Nodes lists are common in communities

23

Mansi Patel
3/9/2017

```
Run MainCommonSource
C:\Users\9mans\Anaconda2\python.exe C:/Users/9mans/PycharmProjects/untitled8/MainCommonSource.py
('PJ_King', 9)
('LovelyMiska', 3)
('zed01', 2)
('irishyanks', 1)
('jamesmakeseyes', 1)
('CaerusMe', 1)
('StrategisAdv', 1)
```

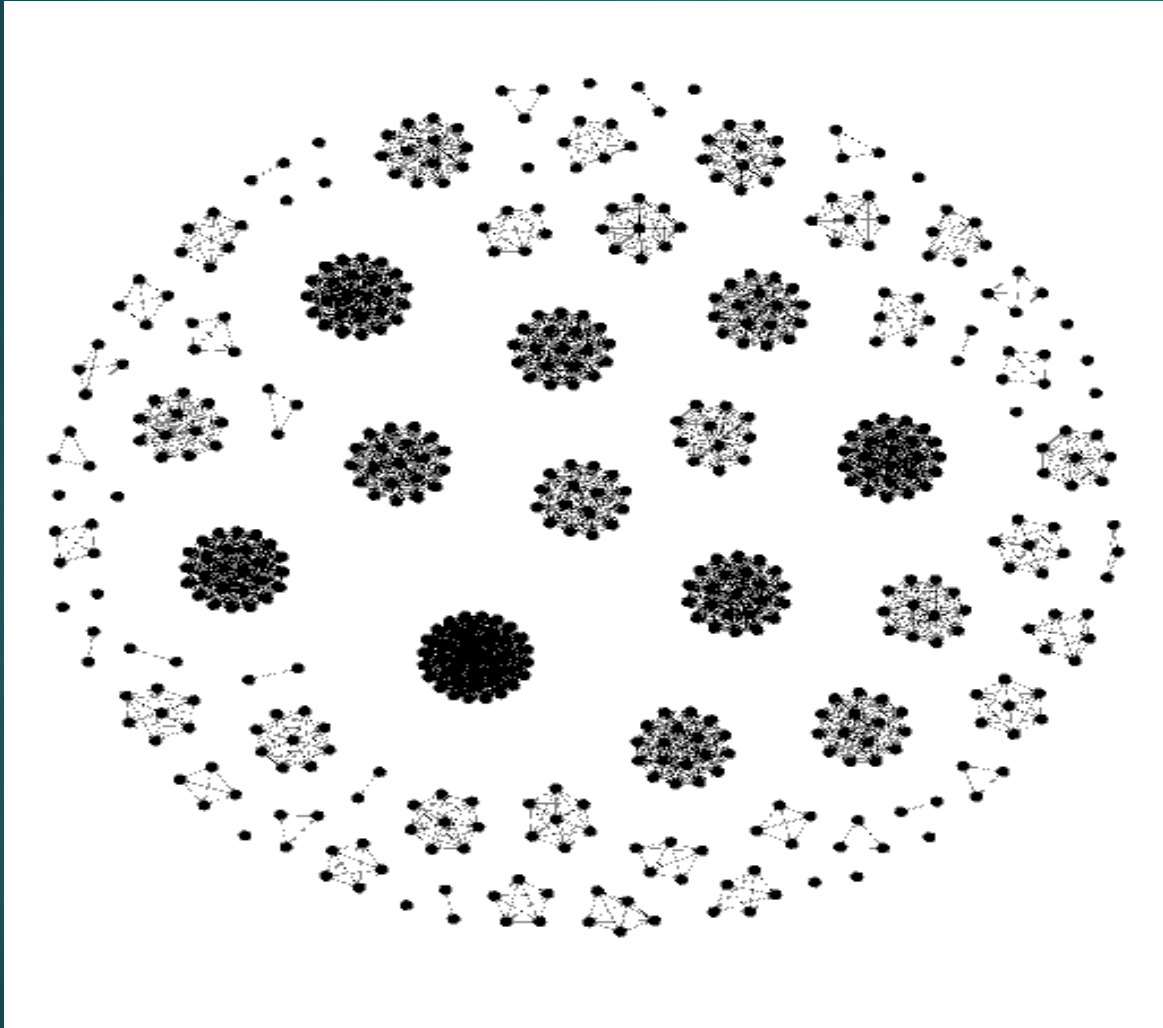
Outline

24

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

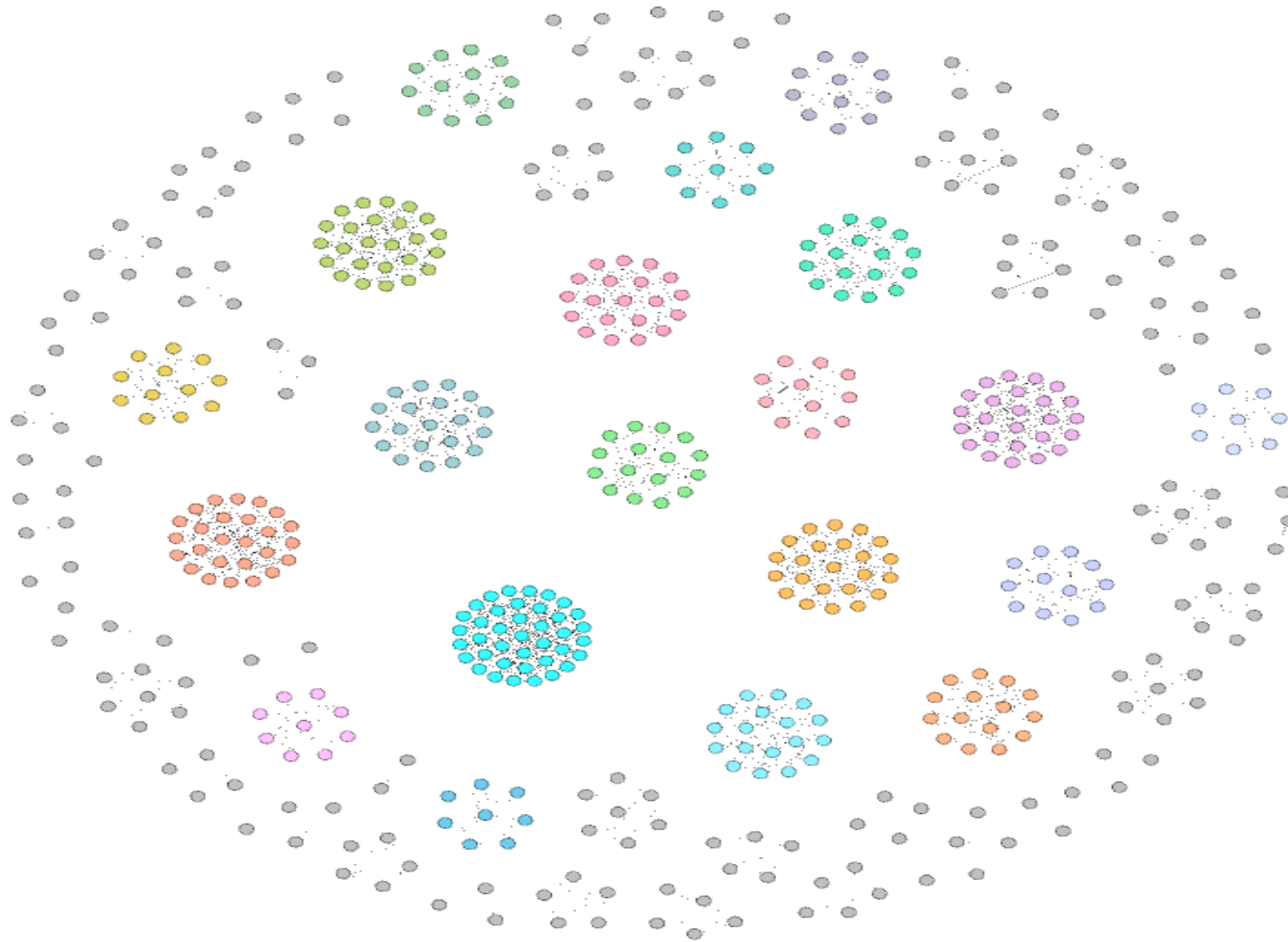
Network Analysis(1/7)

25



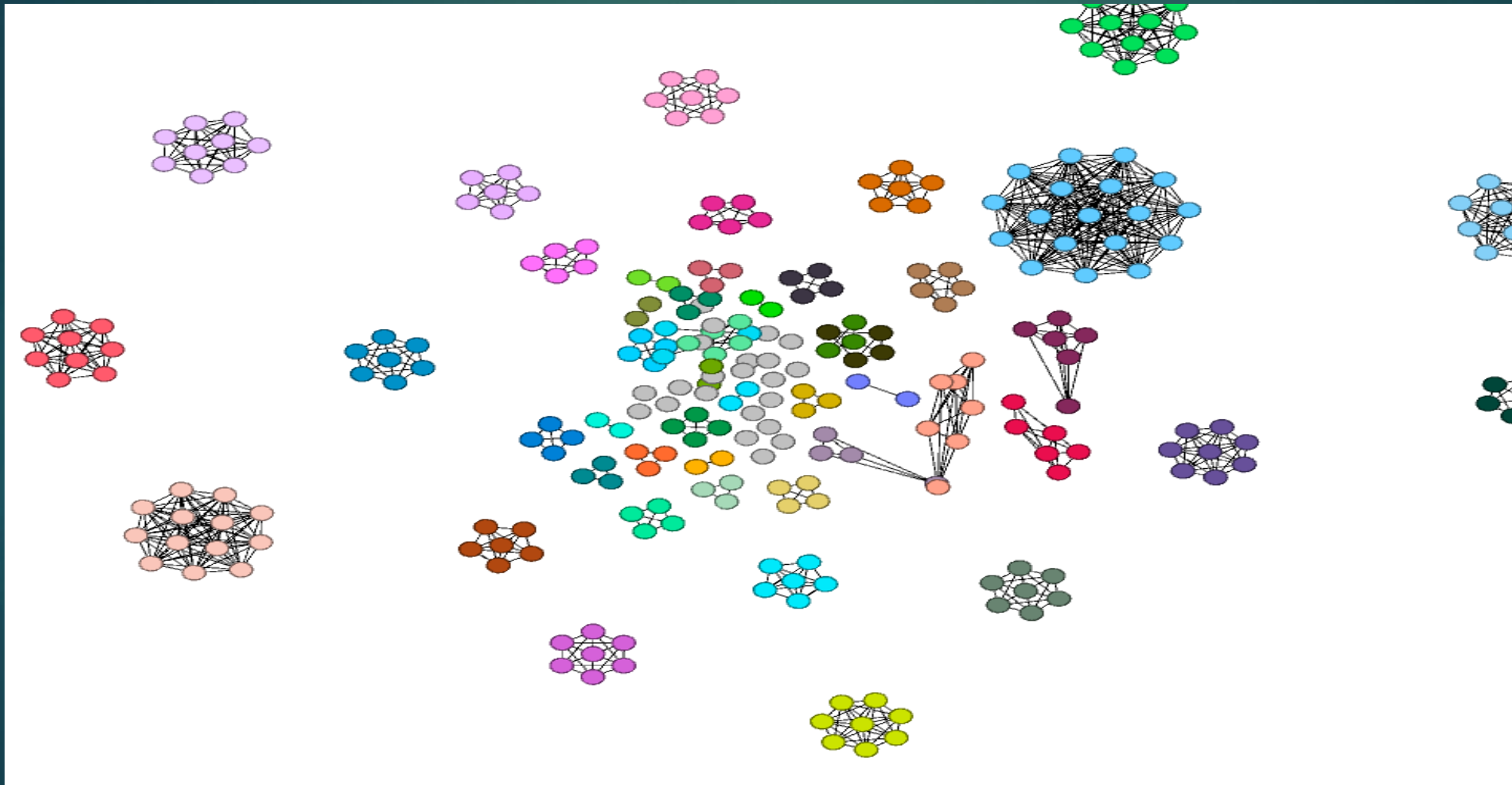
Network Analysis(2/7)

26

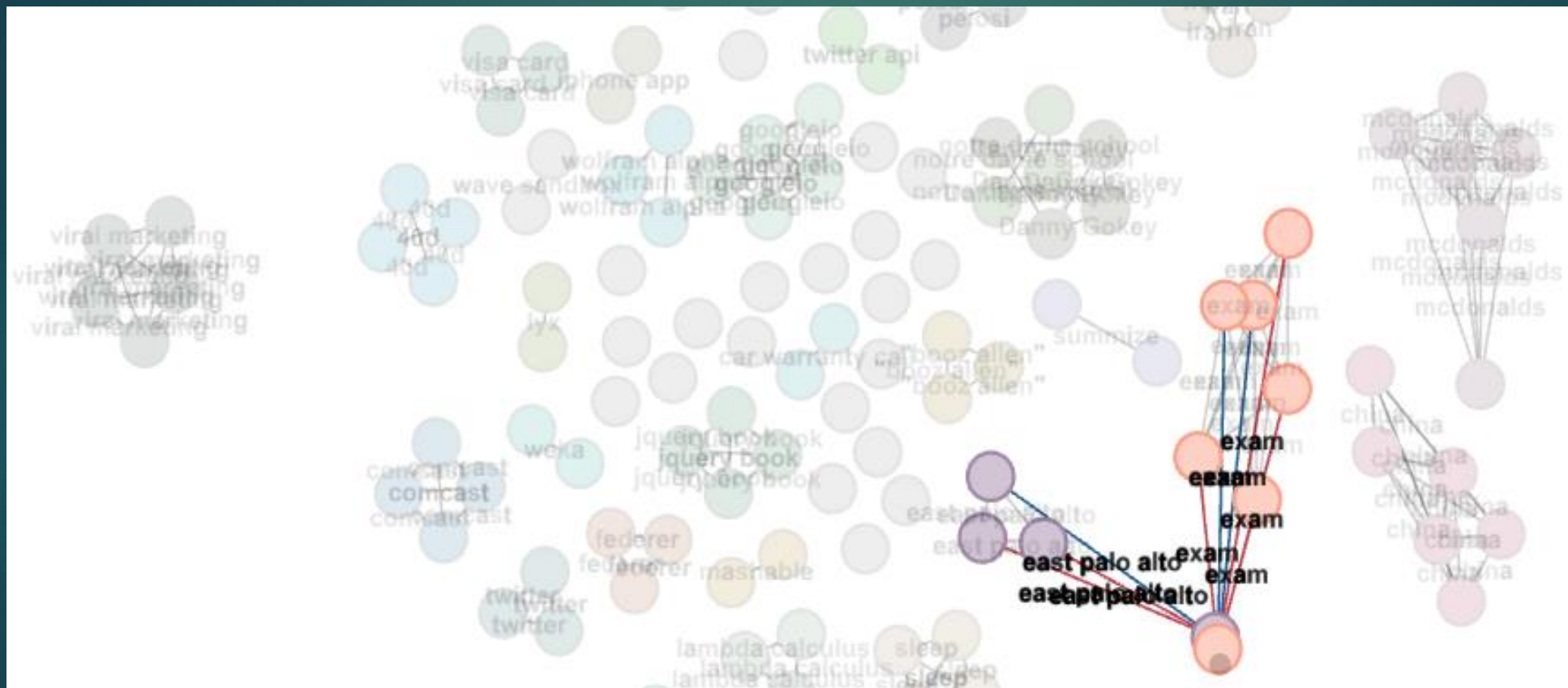


Network Analysis(3/7)

27

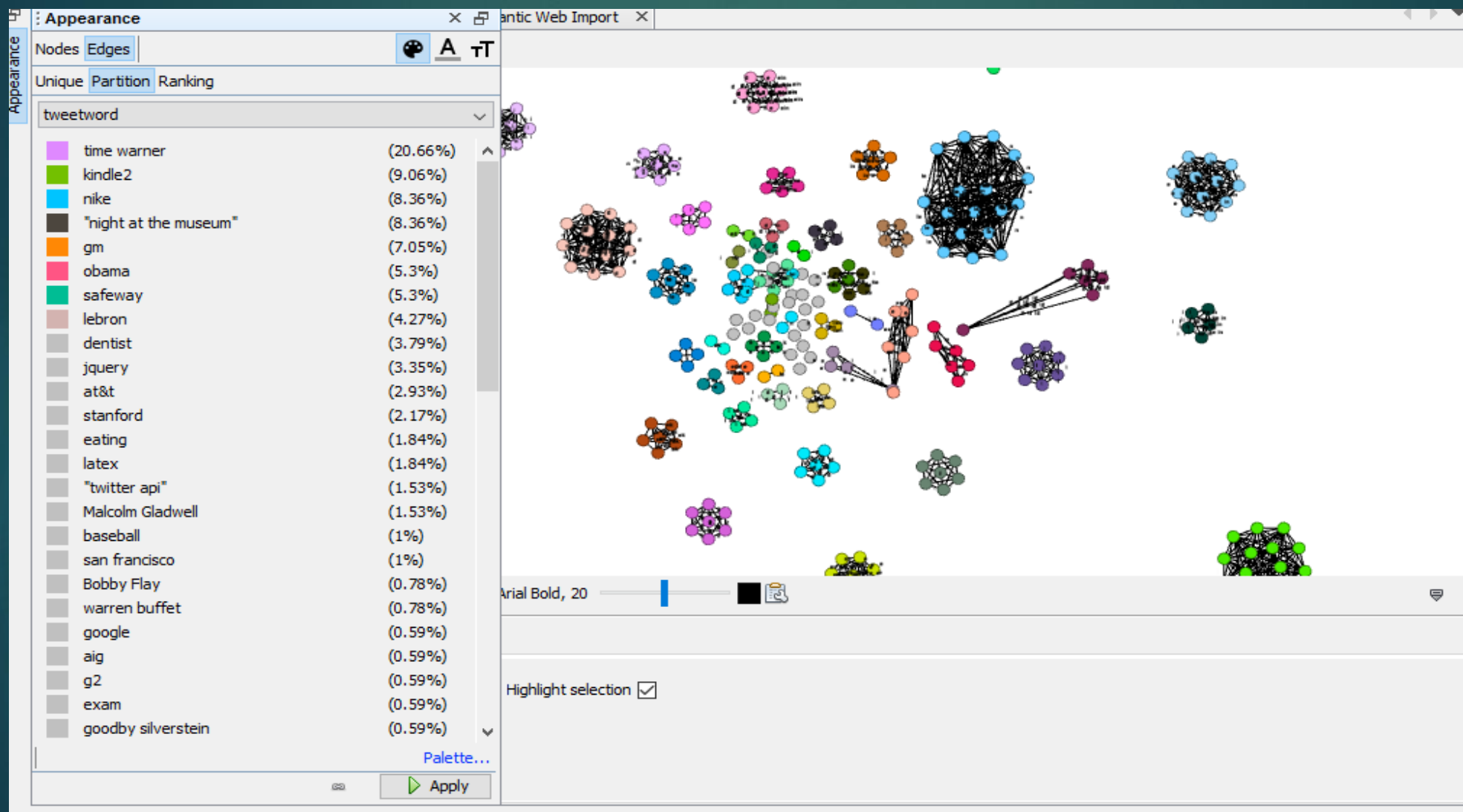


29



Network Analysis(6/7)

30



Network Analysis(6/7)

31

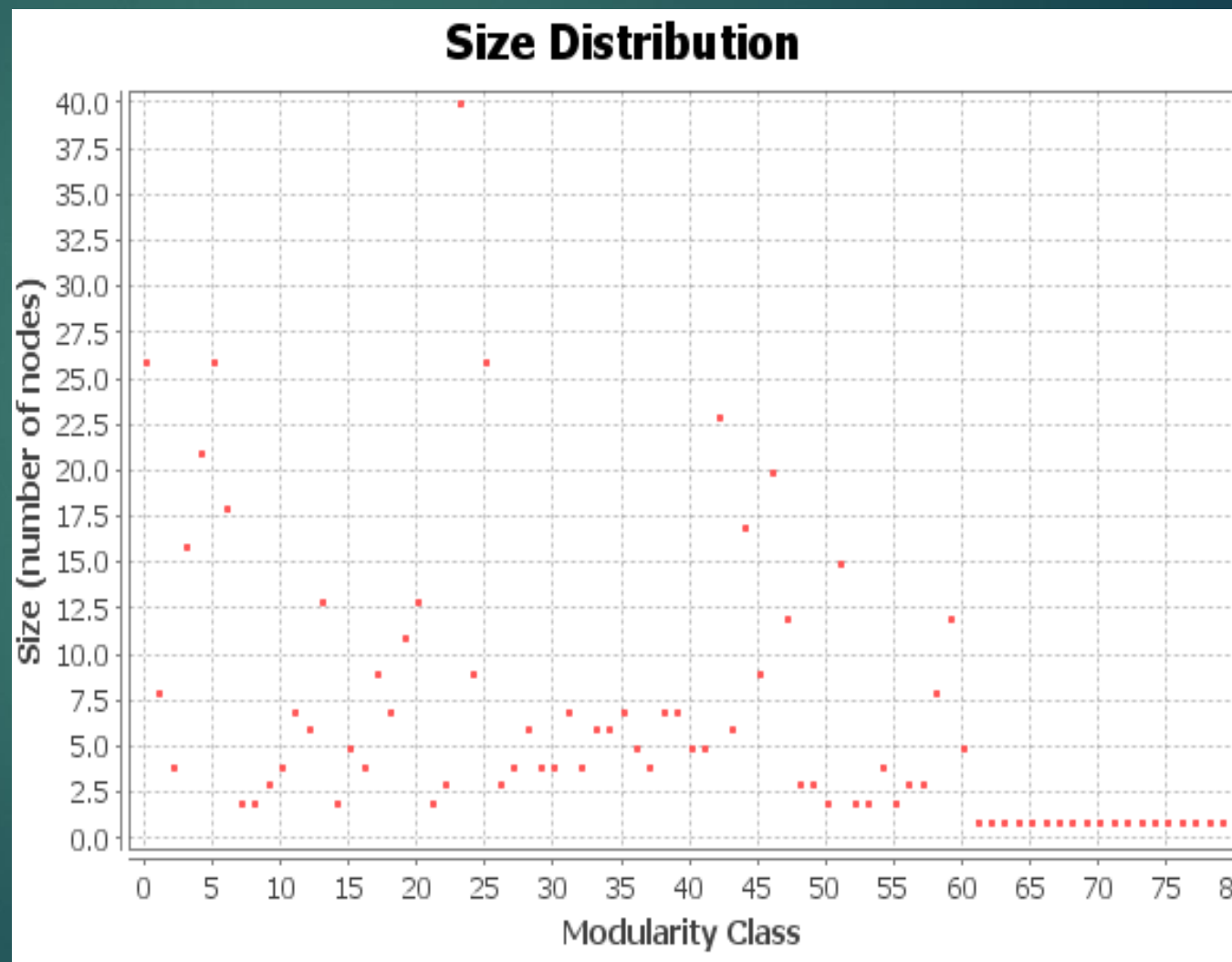
Modularity Report

Parameters:

Randomize: On
Use edge weights: On
Resolution: 1.0

Results:

Modularity: 0.917
Modularity with resolution: 0.917
Number of Communities: 80



Outline

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ **Conclusions**
- ▶ References and Work Distribution

Conclusion

33

- ▶ Source user of the particular tweeted topic was found.
- ▶ Identified communities related to that tweeted topic.
- ▶ Figure out overlapping communities, from the network of the user who tweeted in more than one topic.

Outline

34

- ▶ Introduction
- ▶ Applications
- ▶ Proposed Approach
- ▶ Data Set
- ▶ Implementation
- ▶ Results
 - ▶ Demo
 - ▶ Network Analysis
- ▶ Conclusions
- ▶ References and Work Distribution

References

35

- ▶ Yang, Jaewon, Julian McAuley, and Jure Leskovec. "Community detection in networks with node attributes." Data Mining (ICDM), 2013 IEEE 13th international conference on. IEEE, 2013.
- ▶ Pan, Y., Li, D. H., Liu, J. G., & Liang, J. Z. (2010). Detecting community structure in complex networks via node similarity. Physica A: Statistical Mechanics and its Applications, 389(14), 2849-2857.
- ▶ Start image: <https://www.carma.com/>
- ▶ Fortunato, Santo. "Community detection in graphs." Physics reports 486.3 (2010): 75-174.
- ▶ Twitter Dataset(2009 to 2014): <http://cs.stanford.edu/people/alecmgo/trainingandtestdata.zip>

Work Distribution

36

Mansi Patel
3/9/2017

- ▶ **Mansi Patel:** Implementation, Documentation of Reports, Presentation slides, Network Analysis
- ▶ **Disha Patel:** Dataset Analysis, Literature Survey, Documentation
- ▶ **Bhavik Patel:** Proposed Approach, Literature Survey, Presentation slides, Network Analysis

