TASK 1 DATA MINING CAPSTONE PROJECT

1. Task 1.1

In the first task of the Data Mining Casptone Project, it was used a LDA algorithm applied to a 100,000 reviews sample provided by the python script in order to get exactly 10 topics.

I first used in python the following line on the terminal:

python py27_processYelpRestaurants.py –sample

to get the 100,000 restaurant reviews from 10 different cuisines. After that, the command

python py27_ldaTopicModeling.py -o "sample_topics10.txt" -K 10

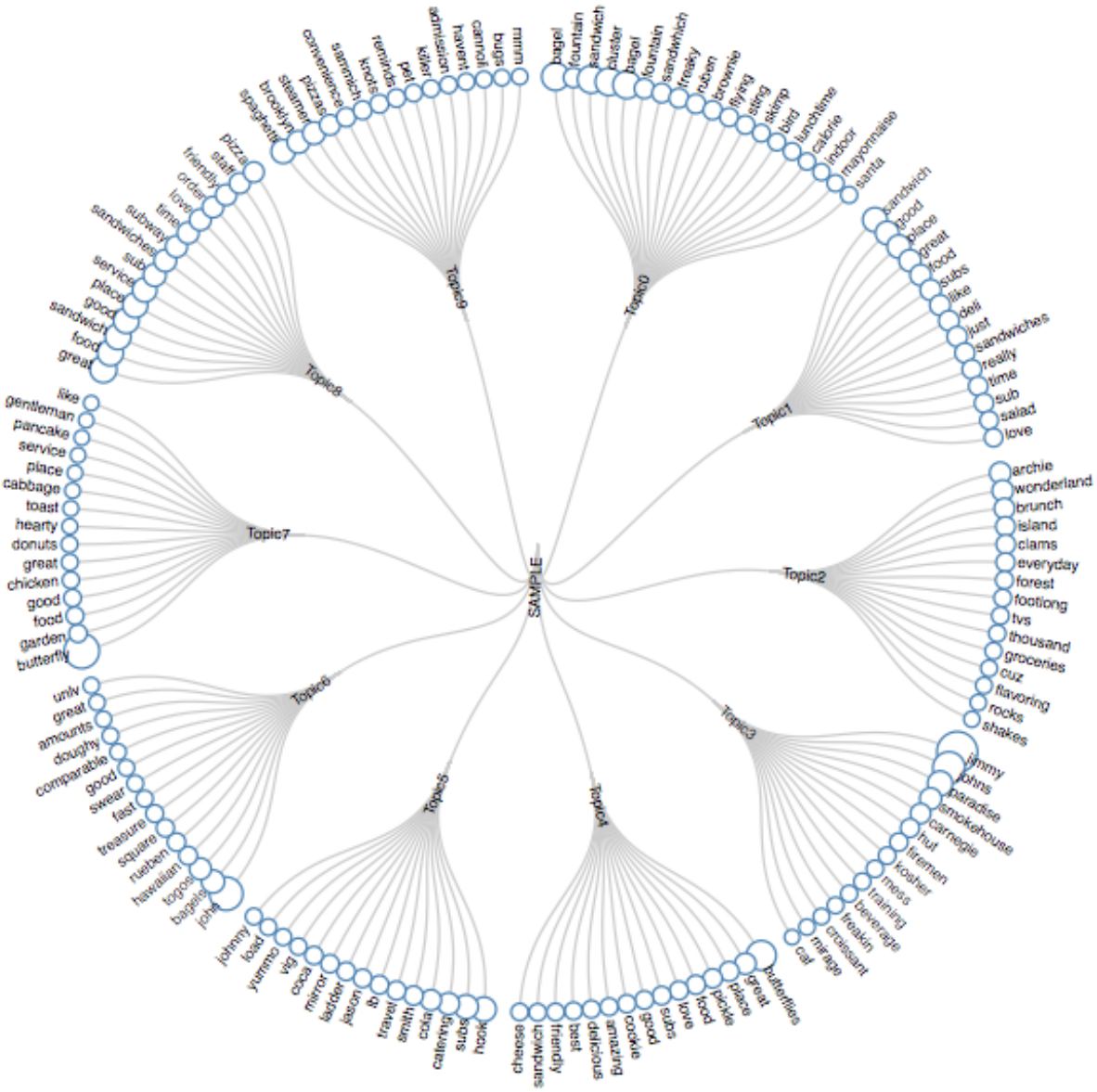
provided me with 10 different topics with its respective weights from the LDA Topic Modelling Algorithm.

In the following visualisation, I depicted with *D3.js* using the Radial Reingold-Tilford Tree (for reference see http://bl.ocks.org/mbostock/4063550) the distribution from the topics with its terms: Each size of the nodes describes the weight of the term in the topic.

The visualisation (next page) is convenient and very appropriate and shows in a direct way how is the distribution of the topics from the given sample. Although if one reviews terms, they are not entirely consistent within each topic suggesting a tune in the LDA model.

2. Task 1.2

Due to lack of time, I was not able to complete this task. Hopefully I will have more time for Task 2 :).


SAMPLE
Topic0
bagel
fountain
sandwich
bluster
bagel
fountain
sandwhich
freaky
ruben
brownie
flying
string
skimp
bird
lunchtime
calorie
indoor
mayonnaise
santa
Topic1
sandwich
good
place
great
food
subs
like
deli
just
sandwiches
really
time
sub
salad
love
Topic2
archie
wonderland
brunch
island
clams
everyday
forest
footlong
fvs
thousand
groceries
cuz
flavoring
rocks
shakes
Topic3
jimmy
johns
paradise
smokehouse
carnegie
hut
firemen
kosher
mess
training
beverage
freakin
croissant
mirage
cat
Topic4
butterflies
great
place
pickle
food
love
subs
good
cookie
amazing
delicious
best
friendly
sandwich
cheese
Topic5
mccu
subs
calloway
coke
smith
travel
lb
jason
ladder
mirror
coca
veg
yummo
load
johnny
Topic6
univ
great
amounts
doughy
comparable
good
swear
fast
treasure
square
rueben
hawaiian
togos
bagels
john
Topic7
like
gentleman
pancake
service
place
cabbage
toast
hearty
donuts
great
chicken
good
food
garden
butterfly
Topic8
great
food
sandwich
good
place
service
sub
sandwiches
subway
time
love
order
friendly
staff
pizza
spaghetti
brooklyn
steamer
pizzas
convenience
sammich
knots
reminds
pet
killer
samission
haventt
cannoli
bigs
minn
Topic9
bagel
fountain
sandwich
bluster
bagel
fountain
sandwhich
freaky
ruben
brownie
flying
string
skimp
bird
lunchtime
calorie
indoor
mayonnaise
santa