# Untitled

*Vishesh Kakarala*

*April 11, 2016*

```
library('stats')
library('networkD3')
library('igraph')
```

```
##
## Attaching package: 'igraph'

## The following objects are masked from 'package:stats':
##
##     decompose, spectrum

## The following object is masked from 'package:base':
##
##     union
```

```
library('twitteR')
library('RCurl')
```

```
## Loading required package: bitops
```

```
library('ngram')
library('js')
library('tau')
library('tm')
```

```
## Loading required package: NLP
```

```
library('stringdist')
```

## Question 1. Intro Network Analyses

a) Read in the files and visualize the network. Read the data

```
keys <- read.table("~/Assignment 4/keys.txt", sep = "\t", fill = FALSE, col.names = c("node_id","name"))
Subs <- read.table("~/Assignment 4/subs.txt", sep = "\t", fill = FALSE, col.names = c("ingredient_A","in
```

Visualizing the network using forceNetwork function form networkD3 library

```
keys$group <- 1

keys$node_id <- keys$node_id-1
Subs$ingredient_A <- Subs$ingredient_A-1
Subs$ingredient_B <- Subs$ingredient_B-1
forceNetwork(Links = Subs,Nodes = keys,Source = "ingredient_A", Target = "ingredient_B",NodeID = "name"
```

b)Calculate the degree centrality of each node

```
deg <- graph.data.frame(Subs,keys, directed = T)
keys$centrality <- degree(deg,mode = "Total")
keys$standardized_centrality <- keys$centrality/561

head(keys)
```

```
##   node_id        name group centrality standardized_centrality
## 1       0        leek     1          3             0.005347594
## 2       1       onion     1         33             0.058823529
## 3       2 white onion     1          4             0.007130125
## 4       3     shallot     1          8             0.014260250
## 5       4     hot dog     1          1             0.001782531
## 6       5     sausage     1         19             0.033868093
```

d)Which are the most "connected" node(s). To see which of the nodes are well connected we can check the degree centrality of the nodes, the nodes with the highest degree centrality are well connected.

```
keys$name[head(order(keys$centrality,decreasing= T), n =10)]
```

```
##  [1] onion      chicken    milk       seasoning  bread      pepper
##  [7] tomato     pineapple  olive oil  applesauce
## 562 Levels: allspice almond almond extract almond paste ... zucchini
```

c)Visually determine what are the furthest ingredients from cocoa powder.

```
keys$group[keys$name == "cocoa powder"] <- 2
# we visualize the network graph again with different group colours to identify the required node

forceNetwork(Links = Subs,Nodes = keys,Source = "ingredient_A", Target = "ingredient_B",NodeID = "name"
```

visually we can see that saffron and iceberg lettuce are the furthest away from cocoa powder