

# Neo4j Tweet Analysis

*Sunit Jain*

*January 6, 2015*

## Setup

```
install.packages("devtools")
devtools::install_github("nicolewhite/RNeo4j")
```

## Launch Database

```
setwd("~/Neo4j/Twitter/omicsGraph")
suppressMessages(require(igraph))
suppressMessages(require(RNeo4j))
graph = startGraph("http://localhost:2794/db/data/")
graph$version
```

```
## [1] "2.1.6"
```

## Constructing the query

Users are connected if they've tweeted the same hashtag. Edges are weighted by how many times the users have tweeted the same hashtag. Undirected.



where, the hashtags name `h.name` *should not be* `<> "metagenomics"`, `"rnaseq"` and `"omics"`, since we've only downloaded these tweets.

```
query = "
MATCH (u1:User)-[:POSTS]->(:Tweet)<-[:TAGS]-(h:Hashtag)-[:TAGS]->(:Tweet)<-[:POSTS]-(u2:User)
WHERE h.name <> 'rnaseq' AND h.name <> 'metagenomics' AND (ID(u1) < ID(u2))
RETURN u1.screen_name, u2.screen_name, COUNT(*) AS weight
"
```

Execute the query and present the result

```
data = cypher(graph, query)
head(data)
```

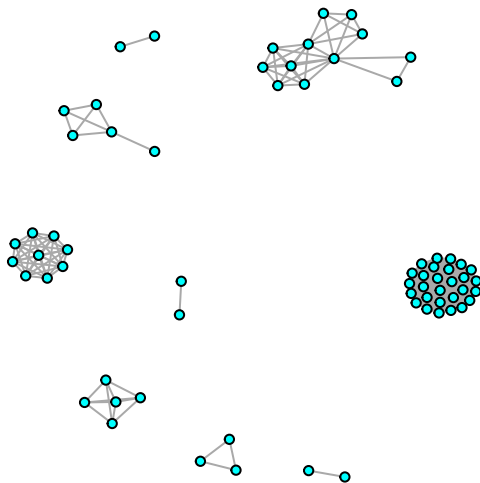
```
##   u1.screen_name u2.screen_name weight
## 1   SeqComplete   cypypasteusa      2
## 2    hellbrat FrontCellDevBio      1
## 3   Seekerotruth   adam_zander      4
## 4    KarenRegis   winterthur      4
## 5  AugustLady241   MrJohnJohn      4
## 6  mcantrellphd  blueheartFISH      3
```

Convert data frame to graph data frame

```
g = graph.data.frame(data, directed = F)
```

Remove text labels and plot.

```
V(g)$label = NA
V(g)$size = 4
V(g)$color = "cyan"
plot(g)
```



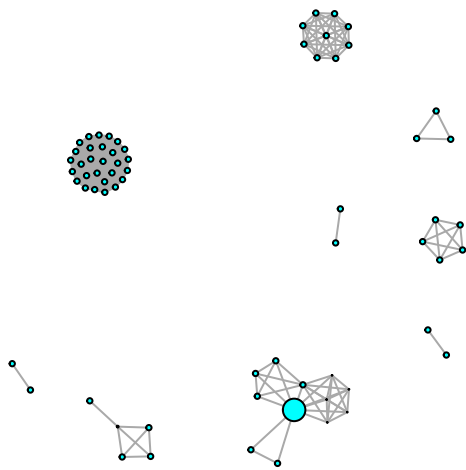
Top 5 betweenness.

```
sort(betweenness(g), decreasing = T)[1:5]
```

##	rnomics	RajanChaudhari	Lau_uribefig	bffo	RSGSpain
##	36.5	3.0	1.2	1.2	1.2

Make size of node a function of its betweenness.

```
V(g)$size = betweenness(g) / (max(betweenness(g)) * .1)  
plot(g)
```



## Clustering.

```
cluster = edge.betweenness.community(g)$membership  
  
colors = rainbow(max(cluster))  
colors = colors[sample(length(colors))]  
  
V(g)$color = colors[cluster]  
V(g)$size = 4  
plot(g)
```

