Twitter network analysis

ANALYZING SOCIAL MEDIA DATA IN R

Sowmya Vivek
Data Science Coach

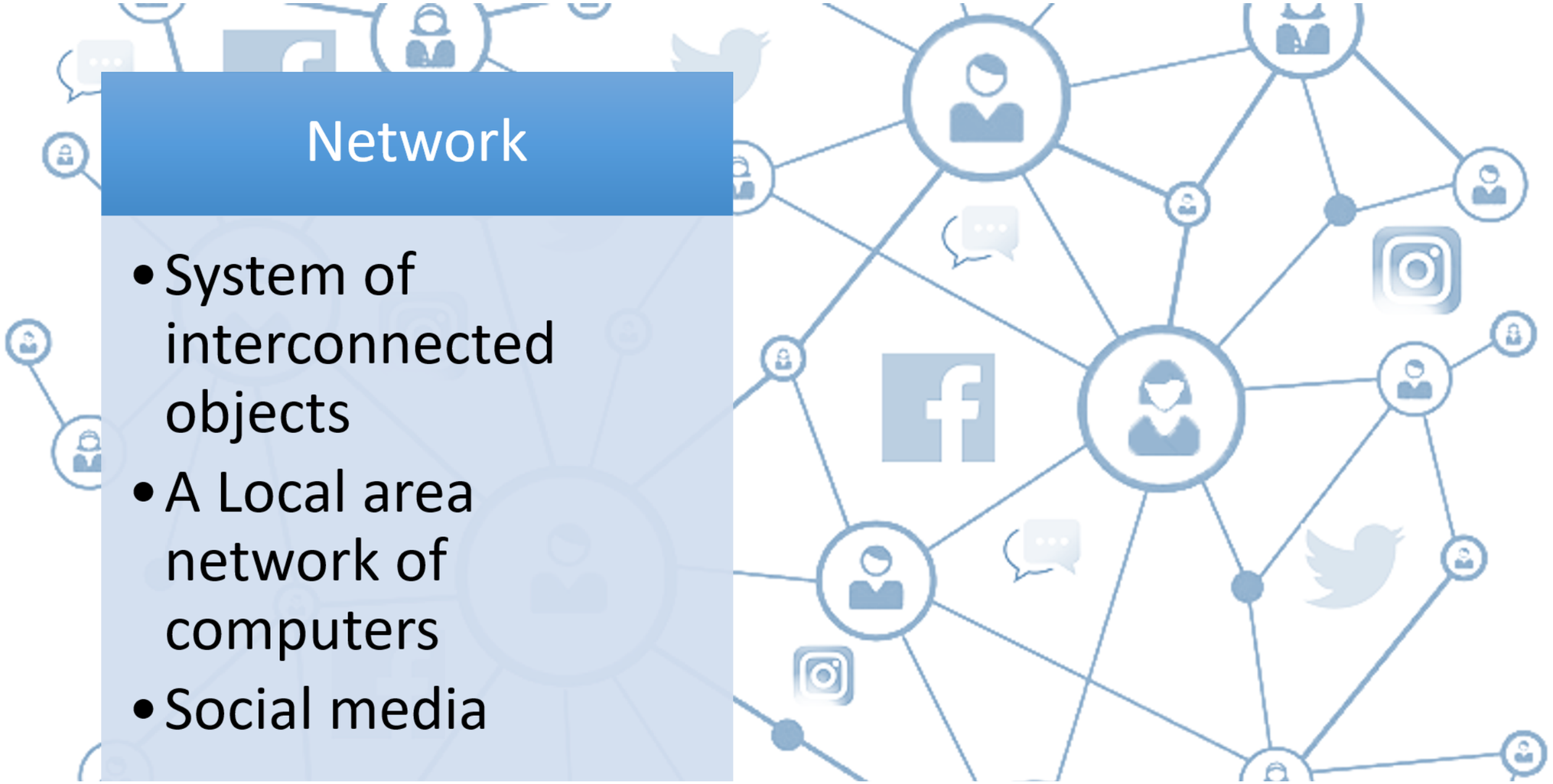
Lesson overview

- Understand the concepts of networks
- Application of network concepts to social media
- Create a retweet network for a topic

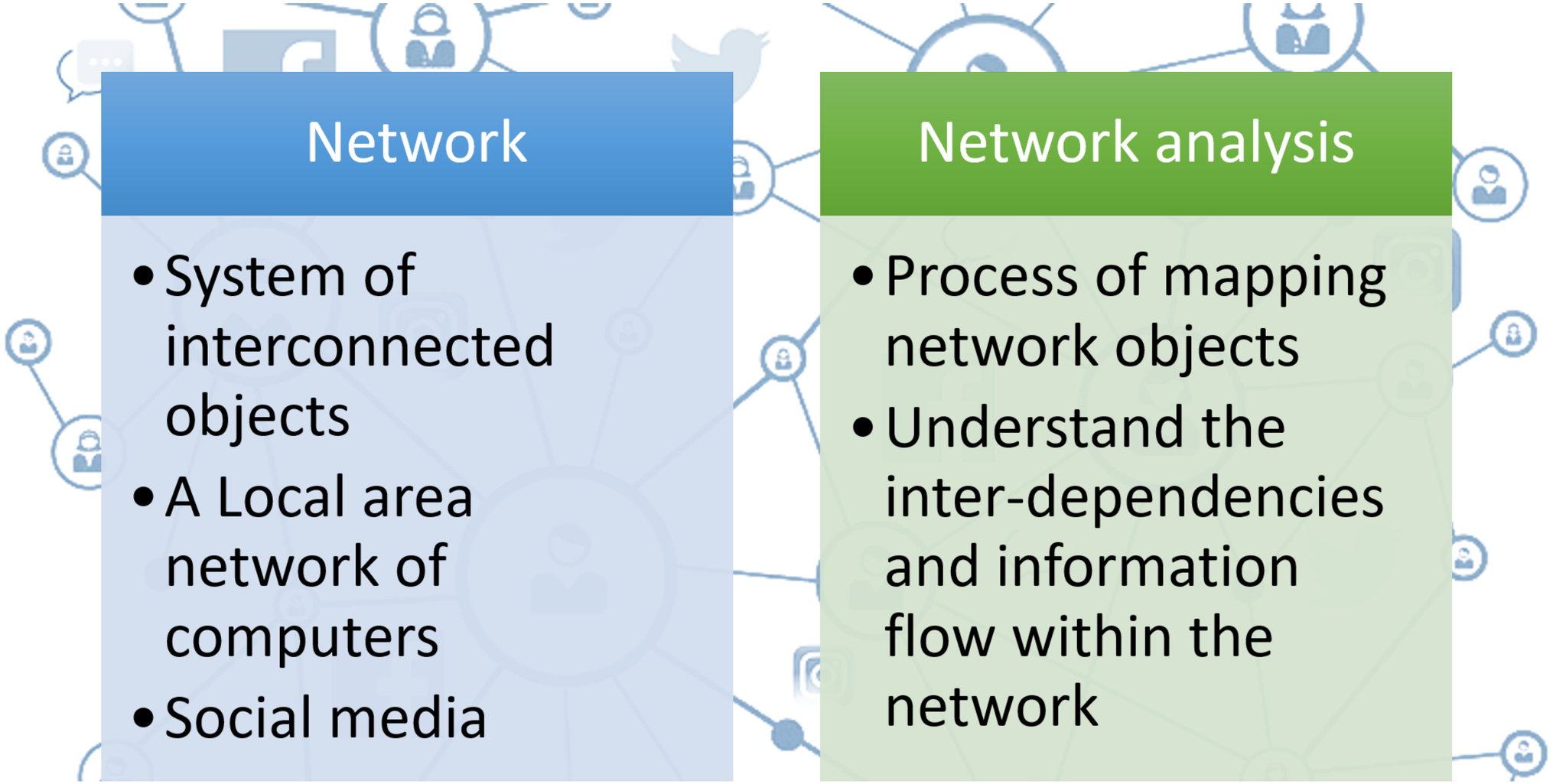
Network and network analysis

Network

- System of interconnected objects
- A Local area network of computers
- Social media



Network and network analysis



Network

- System of interconnected objects
- A Local area network of computers
- Social media

Network analysis

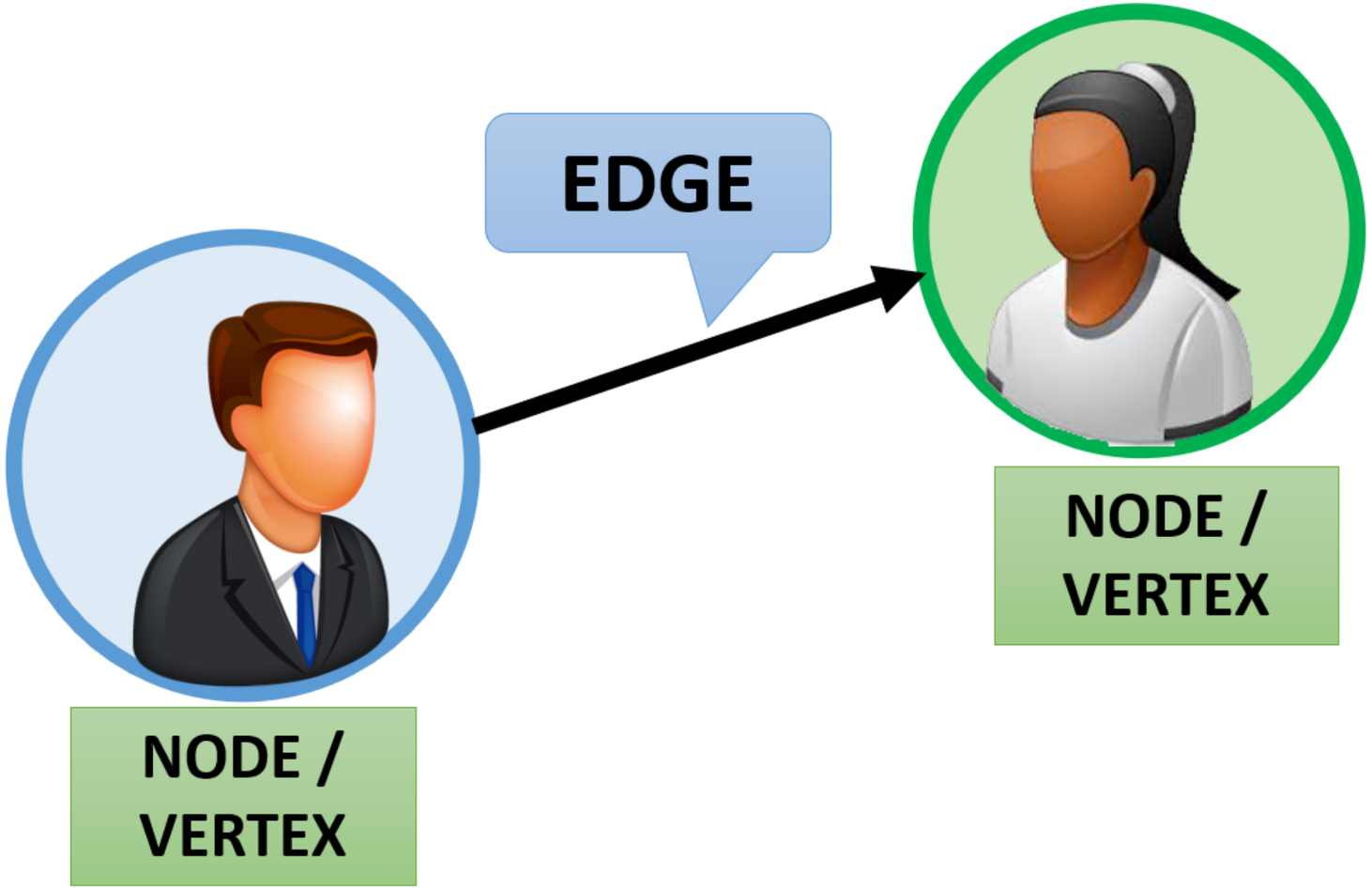
- Process of mapping network objects
- Understand the inter-dependencies and information flow within the network

Components of a network

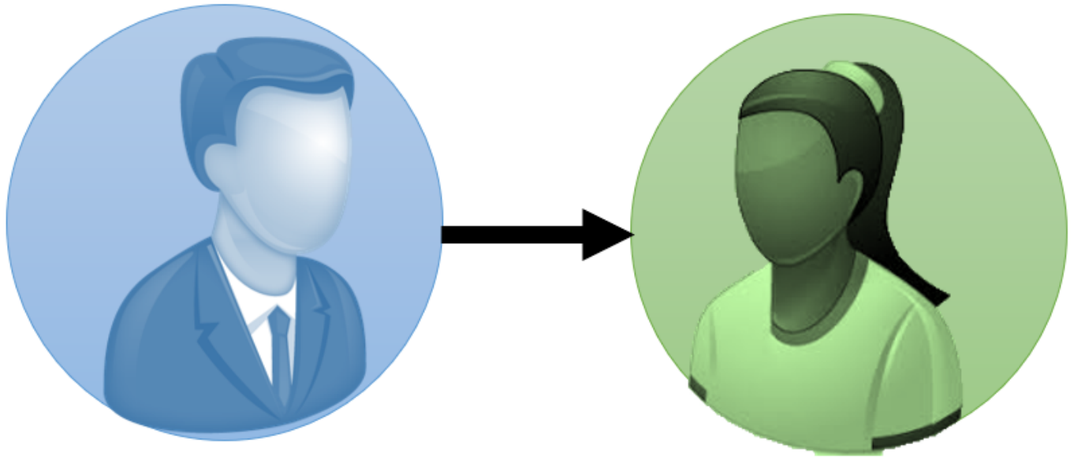
NODE / VERTEX

NODE / VERTEX

Components of a network



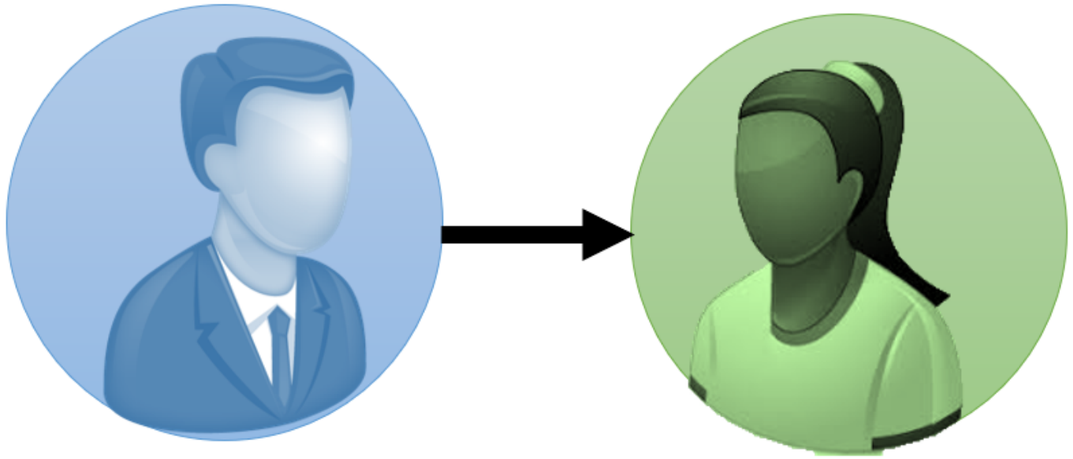
Directed vs undirected network



Directed Network

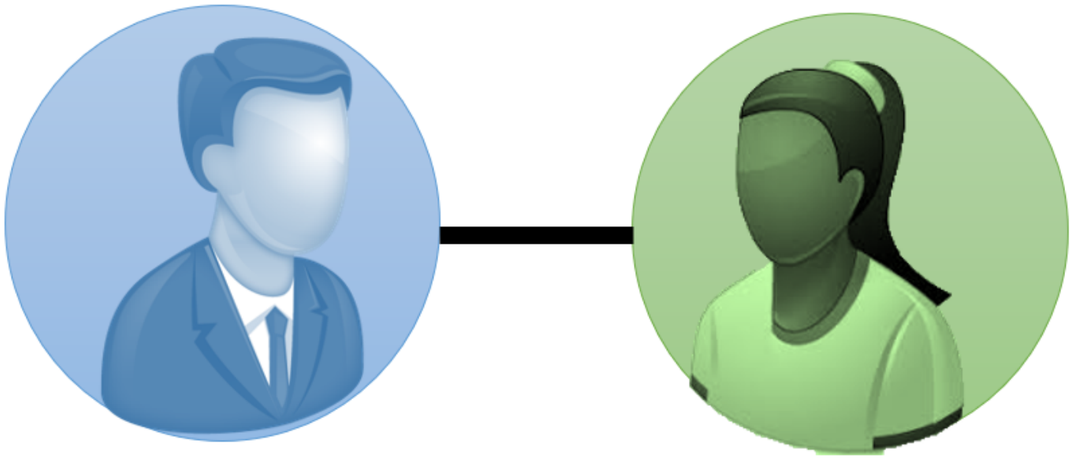
- Edges of network point to one direction

Directed vs undirected network



Directed Network

- Edges of network point to one direction



Undirected Network

- Edges have no direction

Applications in social media

- Twitter users create complex network structures
- Analyze the structure and size of the networks
- Identify key players and influencers in a network
- Pivotal to transmit information to a wide audience

Retweet network

- Network of users who retweet original tweets posted
- A directed network where the source vertex is the user who retweets
- Target vertex is the user who posted the original tweet
- Position on a retweet network helps identify key players to spread brand messaging

Retweet network of #OOTD

- Create a retweet network of users who retweet on #OOTD
- This hashtag is popular amongst users in the age group 16-24
- Can be used to grab the attention of potential customers

Create the tweet data frame

```
# Create tweet data frame for tweets on #OOTD
twts_OOTD <- search_tweets("#OOTD ", n = 18000, include_rts = TRUE)
```

Create data frame for the network

```
# Create data frame for the network
rt_df <- twts_OOTD[, c("screen_name" , "retweet_screen_name" )]
```

```
head(rt_df,10)
```

```
screen_name       retweet_screen_name
<chr>                       <chr>
ShesinfashionCc              NA
glamwearplanet               NA
lanacond0r            LiveKellyRyan
animeninjaz                  NA
zeluslondon                  NA
IonaJaneLevy                 NA
```

Include only retweets in the data frame

```
# Remove rows with missing values
rt_df_new <- rt_df[complete.cases(rt_df), ]
```

Convert data frame to a matrix

```
# Convert to matrix
matrx <- as.matrix(rt_df_new)
```

Create the retweet network

```
# Create the retweet network
library(igraph)
nw_rtweet <- graph_from_edgelist(el = matrx, directed = TRUE)
```

View the retweet network

```
# View the retweet network
print.igraph(nw_rtweet)
```

View the retweet network

```
IGRAPH 7f42937 DN-- 4100 4616 --
+ attr: name (v/c)
+ edges from 7f42937 (vertex names):
 [1] MaikielYungin  ->ZingletC        MaikielYungin  ->ZingletC
 [3] victoria_shop_1->victoria_shop_1 victoria_shop_1->victoria_shop_1
 [5] victoria_shop_1->victoria_shop_1 victoria_shop_1->victoria_shop_1
 [7] victoria_shop_1->victoria_shop_1 victoria_shop_1->victoria_shop_1
 [9] victoria_shop_1->victoria_shop_1 w3daily        ->RealFirstBuzz
[11] w3daily        ->RealFirstBuzz   w3daily        ->RealFirstBuzz
[13] w3daily        ->RealFirstBuzz   w3daily        ->RealFirstBuzz
[15] w3daily        ->RealFirstBuzz   w3daily        ->RealFirstBuzz
```

Let's practice!

ANALYZING SOCIAL MEDIA DATA IN R

Network centrality measures

ANALYZING SOCIAL MEDIA DATA IN R

Sowmya Vivek
Data Science Coach

Lesson overview

- Concept of network centrality measures
- Degree centrality and betweenness
- Identify key players in the network and their role in a promotional campaign

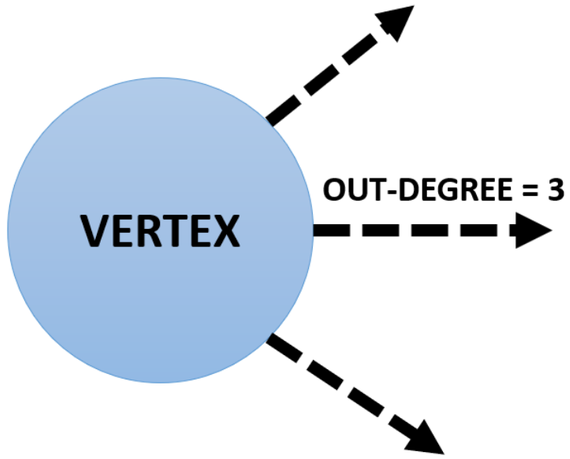
Network centrality measures

- Influence of a vertex is determined by the number of edges and its position
- Network centrality is the measure of importance of a vertex in a network
- Network centrality measures assign a numerical value to each vertex
- Value is a measure of a vertex's influence on other vertices

Degree centrality

- Simplest measure of vertex influence
- Determines the edges or connections of a vertex
- In a directed network, vertices have out-degree and in-degree scores

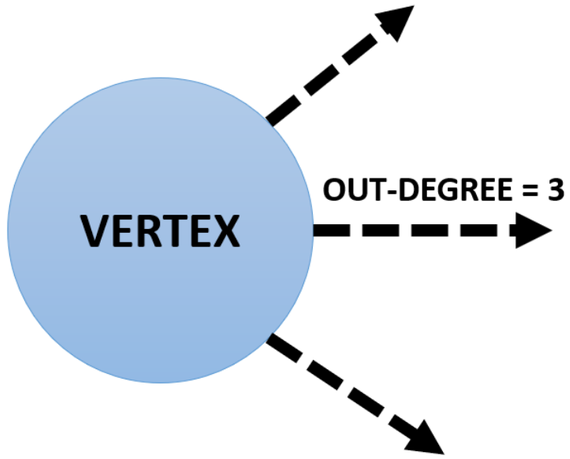
Out-degree



Out-degree

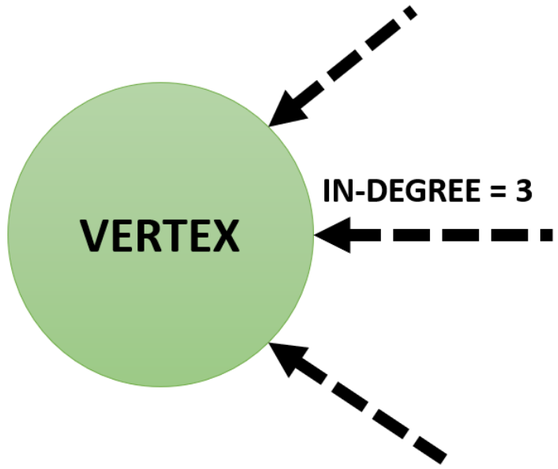
- Number of outgoing edges from a vertex
- Measure of number of times user retweets

In-degree



Out-degree

- Number of outgoing edges from a vertex
- Measure of number of times user retweets



In-degree

- Number of incoming edges to a vertex
- Measure of number of times a user's posts are retweeted

Degree centrality of a user

```
library(igraph)
# Calculate out-degree
out_deg <- degree(nw_rtweet,
                  "OutfitAww",
                   mode = c("out"))
out_deg
```

```
OutfitAww
    20
```

```
library(igraph)
# Calculate in degree
in_deg <- degree(nw_rtweet,
                   "OutfitAww",
                    mode = c("in"))
in_deg
```

```
OutfitAww
   23
```

Users who retweeted most

```
# Calculate the out-degree scores
out_degree <- degree(nw_rtweet, mode = c("out"))
```

```
# Sort the users in descending order of out-degree scores
out_degree_sort <- sort(out_degree, decreasing = TRUE)
```

Users who retweeted most

```
# View the top 3 users
out_degree_sort[1:3]
```

```
VanesEtim   RedNileShop     w3daily
  209           147           62
```

Users whose posts were retweeted most

```
# Calculate the in-degree scores
in_degree <- degree(nw_rtweet, mode = c("in"))
```

```
# Sort the users in descending order of in-degree scores
in_degree_sort <- sort(in_degree, decreasing = TRUE)
```

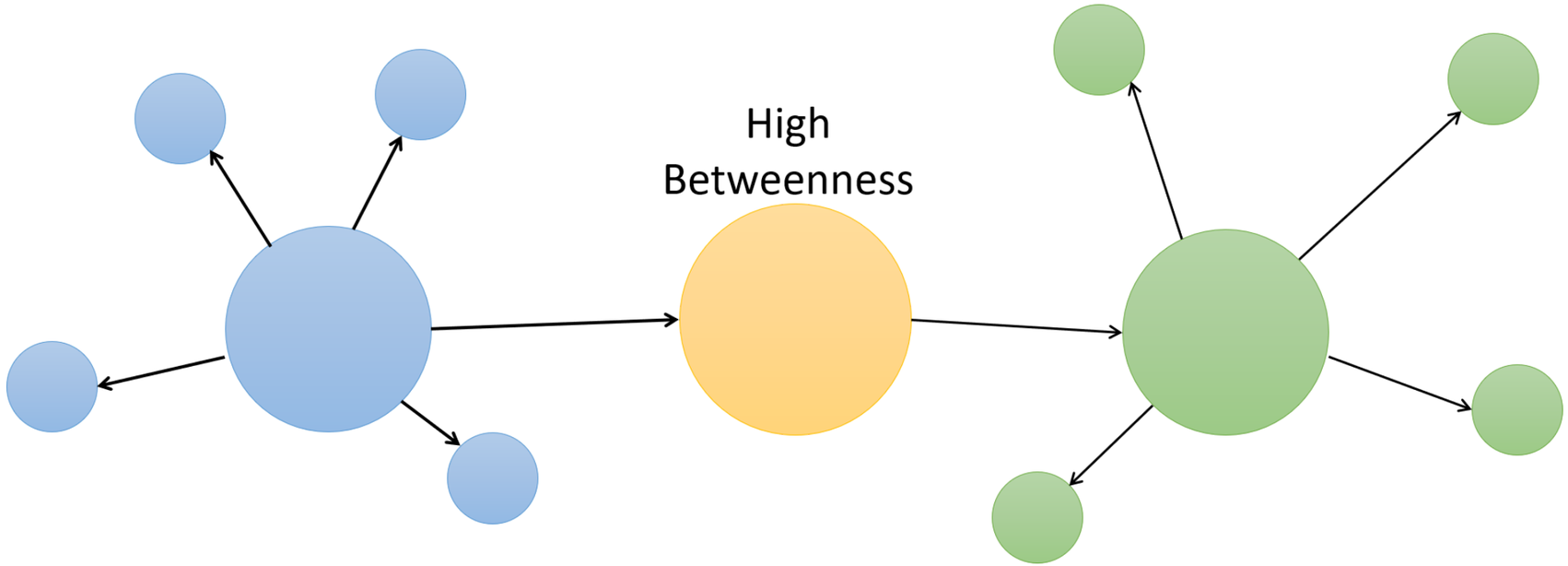
Users whose posts were retweeted most

```
# View the top 3 users
in_degree_sort[1:3]
```

```
XyC_129    SocialBflyMag       jisoupy
  171              167             142
```

Betweenness

- Degree to which nodes stand between each other
- Captures user role in allowing information to pass through network
- Node with higher betweenness has more control over the network



Identifying users with high betweenness

```
# Calculate the betweenness scores of the network
betwn_nw <- betweenness(nw_rtweet, directed = TRUE)
```

```
# Sort the users in descending order of betweenness scores
betwn_nw_sort <- betwn_nw %>%
                   sort(decreasing = TRUE) %>%
                   round()
```

Identifying users with high betweenness

```
# View the top 3 users
betwn_nw_sort[1:3]
```

```
GuruOfficial   Home_and_Loving    SimplyTasheena
   65                  54                 40
```

Let's practice!

ANALYZING SOCIAL MEDIA DATA IN R

Visualizing twitter networks

ANALYZING SOCIAL MEDIA DATA IN R

Sowmya Vivek
Data Science Coach

Lesson overview

- Plot a network with default parameters
- Apply formatting attributes to improve the readability
- Use network centrality and attributes to enhance the plot

View a retweet network

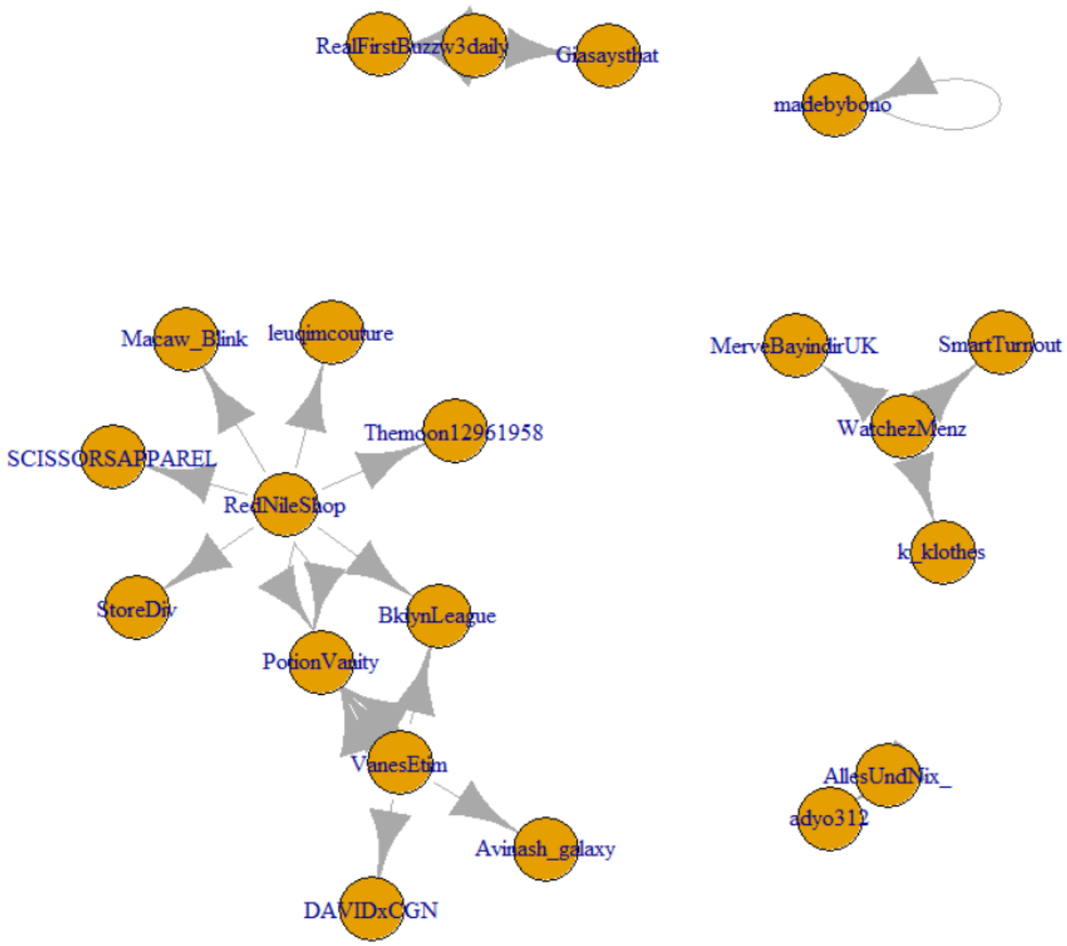
```
# View the retweet network
print.igraph(nw_rtweet)
```

```
IGRAPH e7e618c DN-- 21 39 --
+ attr: name (v/c), followers (v/c)
+ edges from e7e618c (vertex names):
 [1] w3daily    ->RealFirstBuzz   w3daily    ->RealFirstBuzz
 [3] w3daily    ->Giasaysthat     w3daily    ->RealFirstBuzz
 [5] VanesEtim  ->PotionVanity    VanesEtim  ->DAVIDxCGN
 [7] VanesEtim  ->PotionVanity    VanesEtim  ->Avinash_galaxy
 [9] VanesEtim  ->PotionVanity    VanesEtim  ->BklynLeague
[11] RedNileShop->Macaw_Blink     RedNileShop->leuqimcouture
```

Create the base network plot

```
# Create the base network plot
set.seed(1234)
plot.igraph(nw_rtweet)
```

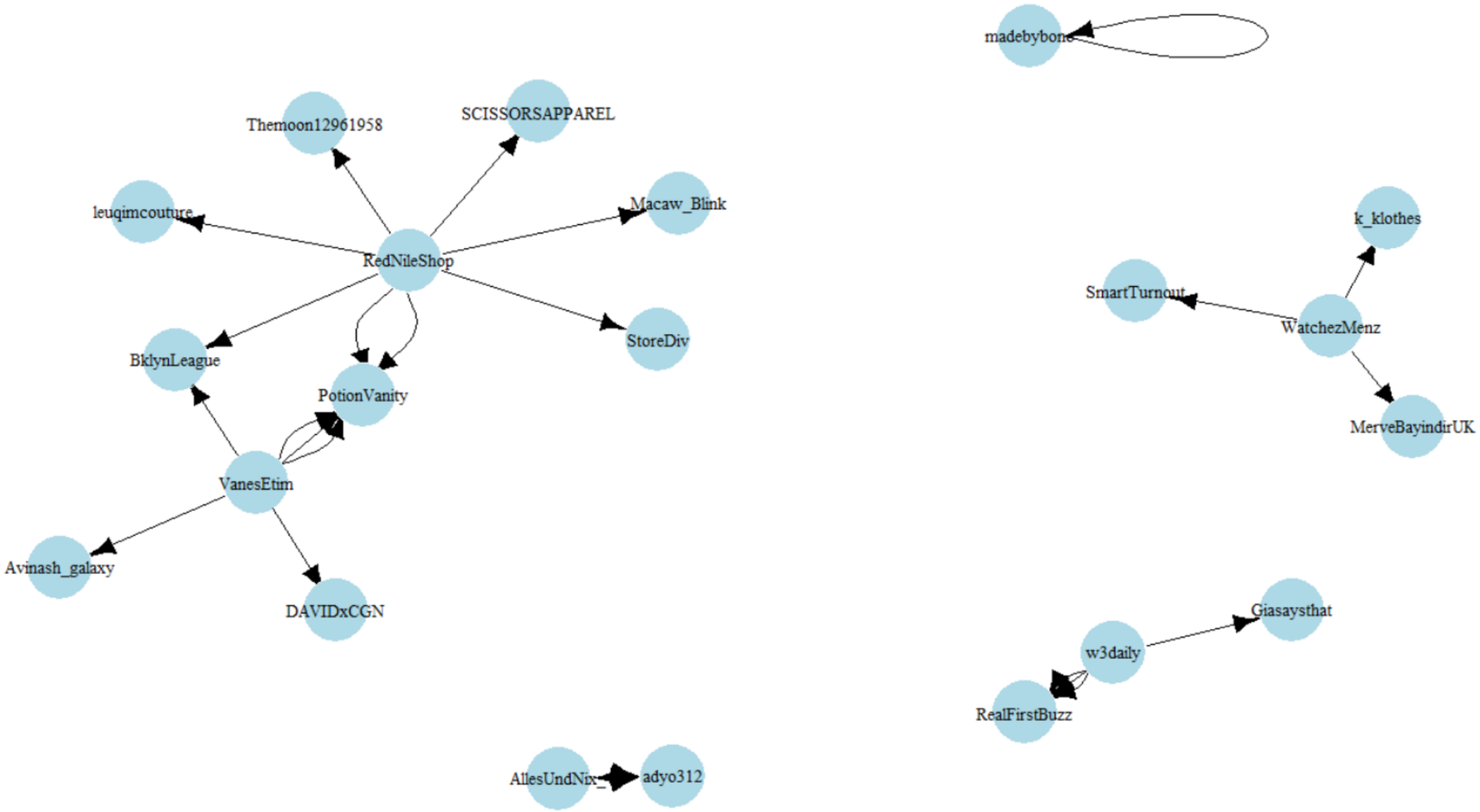
View the base network plot



Format the plot

```
# Format the network plot with attributes
set.seed(1234)
plot(nw_rtweet, asp = 9/16,
     vertex.size = 10,
     vertex.color = "lightblue",
     edge.arrow.size = 0.5,
     edge.color = "black",
     vertex.label.cex = 0.9,
     vertex.label.color = "black")
```

View the formatted plot



Set vertex size based on the out-degree

```
# Create a variable for out-degree
deg_out <- degree(nw_rtweet, mode = c("out"))
deg_out
```

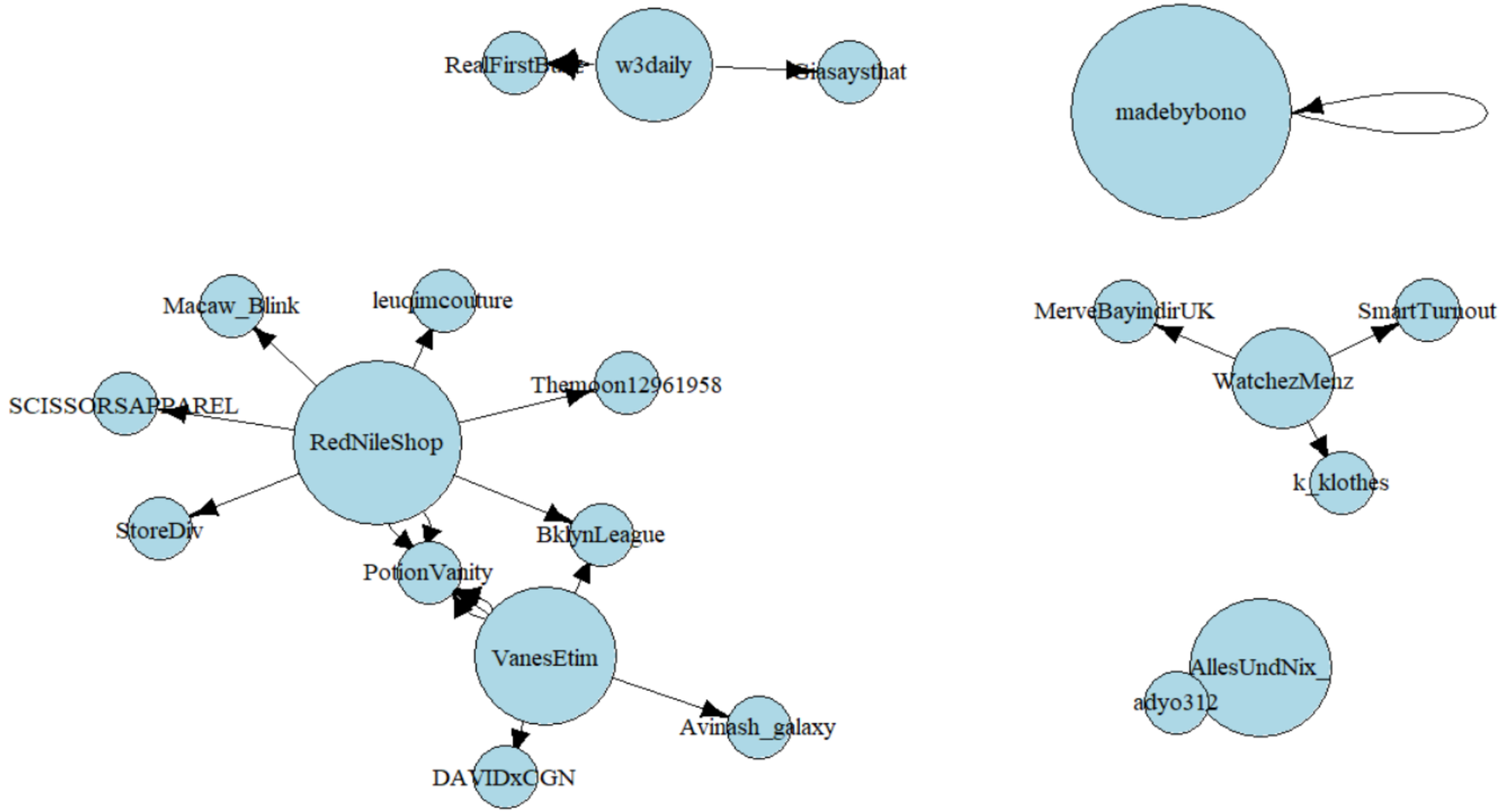
```
          w3daily     RealFirstBuzz      Giasaysthat         VanesEtim     PotionVanity
                4                 0                0                 6                0
       DAVIDxCGN   Avinash_galaxy      BklynLeague      RedNileShop      Macaw_Blink
                0                 0                0                 8                0
    leuqimcouture        StoreDiv SCISSORSAPPAREL Themoon12961958      WatchezMenz
                0                 0                0                 0                3
    MerveBayindirUK     k_klothes     SmartTurnout     AllesUndNix_         adyo312
                0                 0                0                 6                0
       madebybono
               12
```

```
vert_size <- (deg_out * 2) + 10
```

Assign vert_size to the vertex size attribute

```
# Assign vert_size to vertex size attribute and plot network
set.seed(1234)
plot(nw_rtweet, asp = 9/16,
     vertex.size = vert_size,
     vertex.color = "lightblue",
     edge.arrow.size = 0.5,
     edge.color = "black",
          vertex.label.cex = 1.2,
     vertex.label.color = "black")
```

View plot with new attributes



Adding network attributes

- Users who retweet most and have a high follower count add more value
- Network plot of users who retweet more and have a high follower count
- Add follower count as a network attribute

Follower count of network users

```
# Import the followers count data frame
followers <- readRDS("follower_count.rds")
```

View the followers data frame

```
# View the follower count
head(followers)
```

```
screen_name    followers_count
<fctr>             <dbl>
adyo312            58
AllesUndNix_       18
Avinash_galaxy     1536
BklynLeague        40
DAVIDxCGN          267
Giasaysthat        9139
```

Follower count of network users

```
# Categorize high and low follower count
followers$follow <- ifelse(followers$followers_count > 500, "1", "0")
```

View the followers data frame

```
# View the data frame with the new column
head(followers)
```

```
screen_name    followers_count    follow
<fctr>               <dbl>          <chr>
adyo312               58              0
AllesUndNix_          18              0
Avinash_galaxy      1536              1
BklynLeague           40              0
DAVIDxCGN            267              0
Giasaysthat         9139              1
```

Assign network attributes

```
# Assign external network attributes to retweet network
V(nw_rtweet)$followers <- followers$follow
```

View vertex attributes

```
# View the vertex attributes
vertex_attr(nw_rtweet)
```

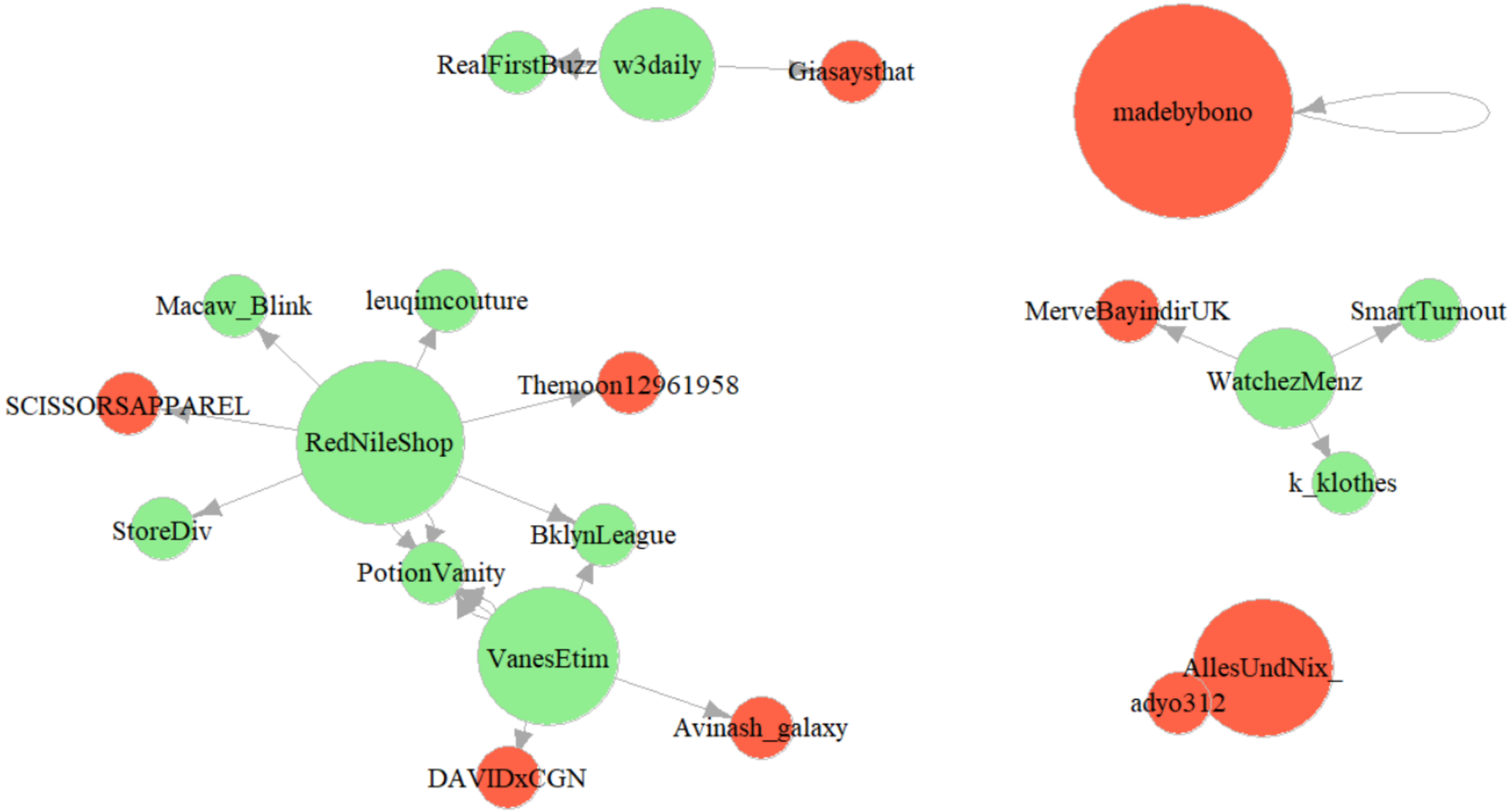
```
$name
 [1] "w3daily"          "RealFirstBuzz"   "Giasaysthat"     "VanesEtim"
 [5] "PotionVanity"     "DAVIDxCGN"       "Avinash_galaxy"  "BklynLeague"
 [9] "RedNileShop"      "Macaw_Blink"     "leuqimcouture"   "StoreDiv"
[13] "SCISSORSAPPAREL" "Themoon12961958" "WatchezMenz"     "MerveBayindirUK"
[17] "k_klothes"        "SmartTurnout"    "AllesUndNix_"    "adyo312"
[21] "madebybono"

$followers
 [1] "0" "0" "1" "0" "0" "1" "1" "0" "0" "0" "0" "0" "1" "1" "0" "1" "0" "0" "1"
[20] "1" "1"
```

Changing vertex colors

```
# Set the vertex colors for the plot
sub_color <- c("lightgreen", "tomato")
set.seed(1234)
plot(nw_rtweet, asp = 9/16,
     vertex.size = vert_size,
     edge.arrow.size = 0.5,
     vertex.label.cex = 1.3,
     vertex.color = sub_color[as.factor(vertex_attr(nw_rtweet, "followers"))],
     vertex.label.color = "black",
     vertex.frame.color = "grey")
```

View plot formatted with vertex attributes



Let's practice!

ANALYZING SOCIAL MEDIA DATA IN R

Putting twitter data on the map

ANALYZING SOCIAL MEDIA DATA IN R

Sowmya Vivek
Data Science Coach

Lesson overview

- Types of geolocation data available in tweets
- Sources of geolocation information
- Extract location details from tweets
- Plot the tweet location data on maps

Why put twitter data on the map

- Mapping locations help understand where tweets are concentrated
- Influence people in those locations with targeted marketing
- Understand reactions to planned or unplanned events

Include geographic metadata

- Twitter users can geo-tag a tweet when it is posted
- Two types of geolocation metadata
  - Place
  - Precise location

Place

- "Place" location is selected from a predefined list
- Includes a bounding box with latitude and longitude coordinates
- Not necessarily issued from the location of the tweet

Precise location

- Specific longitude and latitude "Point" coordinate from GPS-enabled devices
- Represents the exact GPS location
- Only 1-2% of tweets are geo-tagged

Sources of geolocation information

- The tweet text
- User account profile
- Twitter Place added by the user
- Precise location point coordinates

Extract tweets

```
library(rtweet)
```

```
# Extract 18000 tweets on "#politics"
pol <- search_tweets("#politics", n = 18000)
```

Extract geolocation data

```
# Extract geolocation data and append new columns
pol_coord <- lat_lng(pol)
```

- The coordinates are extracted from the columns, `coords_coords` or `bbox_coords`

View lat and lng columns

```
View(pol_coord)
```

lat	lng
NA	*NA*
38.79944	-90.32932550
NA	*NA*
NA	*NA*
NA	*NA*

Omit rows with missing lat and lng values

```
# Omit rows with missing lat and lng values
pol_geo <- na.omit(pol_coord[, c("lat", "lng")])
```

View geocoordinates

```
head(pol_geo)
```

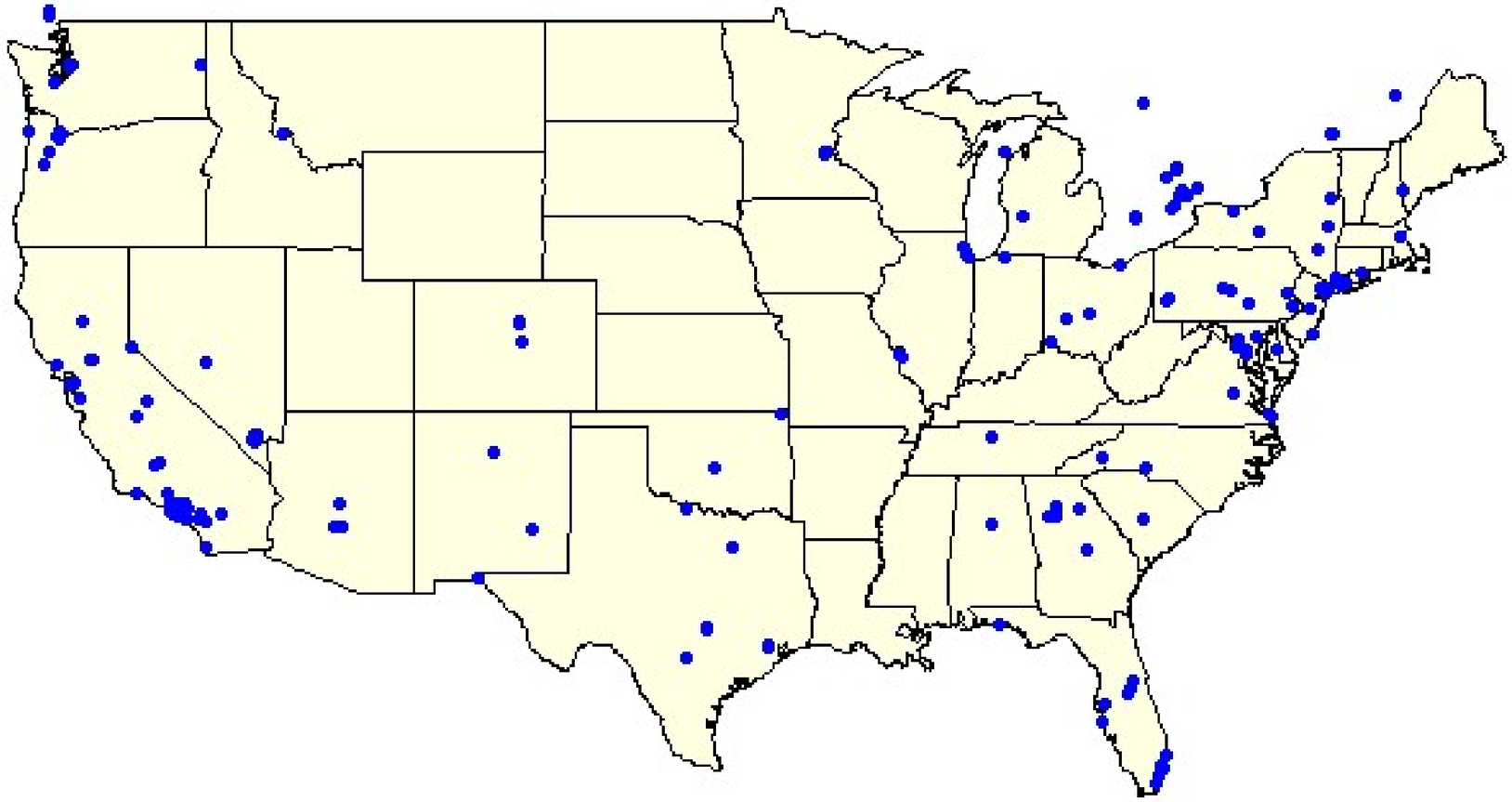
```
lat             lng
<dbl>         <dbl>
19.17414    72.874244
53.35490    -6.247621
53.27350    -6.399521
53.67989     9.372680
12.92311    77.558448
54.59940    -5.836670
```

Plot geo-coordinates on the US state map

```
# Plot longitude and latitude values of tweets on US state map
map(database = "state", fill = TRUE, col = "light yellow")

with(pol_geo, points(lng, lat, pch = 20, cex = 1, col = 'blue'))
```

View the locations on the US state map

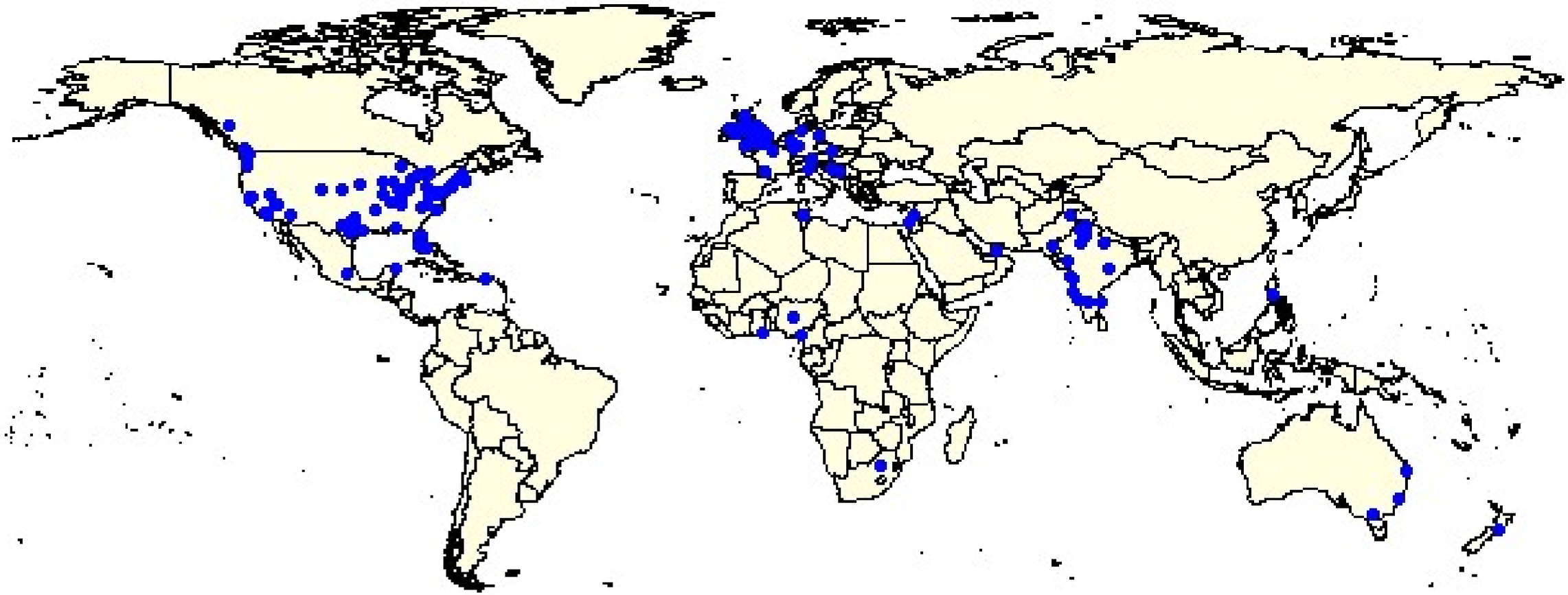


Plot geocoordinates on the world map

```
# Plot longitude and latitude values of tweets on the world map
map(database = "world", fill = TRUE, col = "light yellow")

with(pol_geo, points(lng, lat, pch = 20, cex = 1, col = 'blue'))
```

View the locations on the world map



Let's practice!

ANALYZING SOCIAL MEDIA DATA IN R

Course wrap-up

ANALYZING SOCIAL MEDIA DATA IN R

Sowmya Vivek
Data Science Coach

Our learning journey

Chapter 1

- Power of twitter data
- Tweet components

Our learning journey

Chapter 1

- Power of twitter data
- Tweet components

Chapter 2

- Filter tweets
- Golden ratio
- Twitter lists
- Trends
- Plot data over time

Our learning journey

Chapter 1

- Power of twitter data
- Tweet components

Chapter 2

- Filter tweets
- Golden ratio
- Twitter lists
- Trends
- Plot data over time

Chapter 3

- Process & visualize tweet text
- Topic modeling
- Sentiment analysis

Our learning journey

Chapter 1

- Power of twitter data
- Tweet components

Chapter 2

- Filter tweets
- Golden ratio
- Twitter lists
- Trends
- Plot data over time

Chapter 3

- Process & visualize tweet text
- Topic modeling
- Sentiment analysis

Chapter 4

- Network analysis
- Visualize locations on the map

Next steps

- Reinforce the concepts learned:
  - Collect twitter data around brands, topics, and events
  - Apply the concepts learned
- Enroll for DataCamp courses on important topics in social media analysis
  - Text mining in R
  - Networking analysis in R

Congratulations

Thank you!

ANALYZING SOCIAL MEDIA DATA IN R