

# Exploring your data set

CASE STUDIES: NETWORK ANALYSIS IN R



**Edmund Hart**  
Instructor

```
library(igraph)
library(dplyr)
amzn_raw <- read.csv("datasets/amazon_purchase_no_book.csv")
head(amzn_raw)
```

```
  from to          title.from group.from categories.from salesrank.from
1  1  44 42 The NBA's 100 Greatest Plays      DVD          33124           5
2  2 179 71 Africa Screams/Jack & The Bean    DVD          53825          21

totalreviews.from totalreviews.1.from
1              13              13
2              13              13

          title.to group.to
1          Pixote      DVD
2 Jonny Quest - Bandit in Adventures Best Friend Video

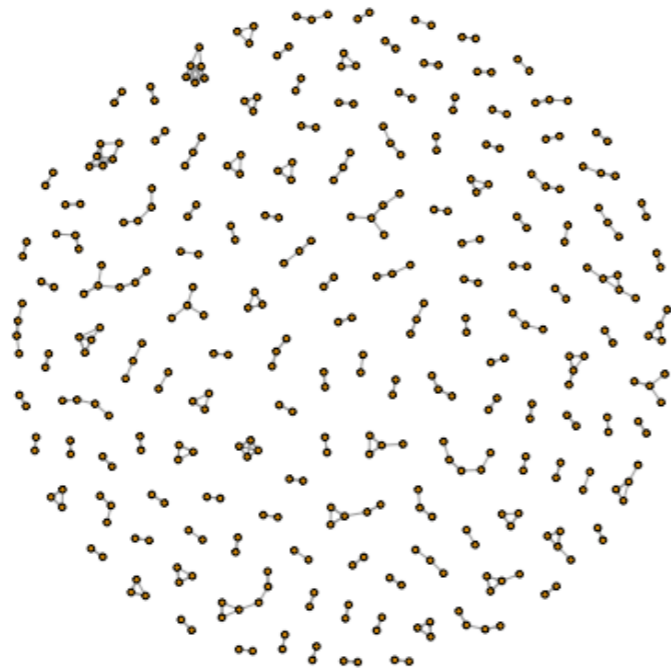
categories.to salesrank.to totalreviews.to totalreviews.1.to      date
1          19685          15          24          24 2003-03-02
2          21571           5           2           2 2003-03-02
```

# Creating the graph

```
amzn_g <- amzn_raw %>%  
  filter(date == "2003-03-02") %>%  
  select(from, to) %>%  
  graph_from_data_frame(directed = TRUE)  
gorder(amzn_g)  
gsize(amzn_g)
```

# Visualize the graph

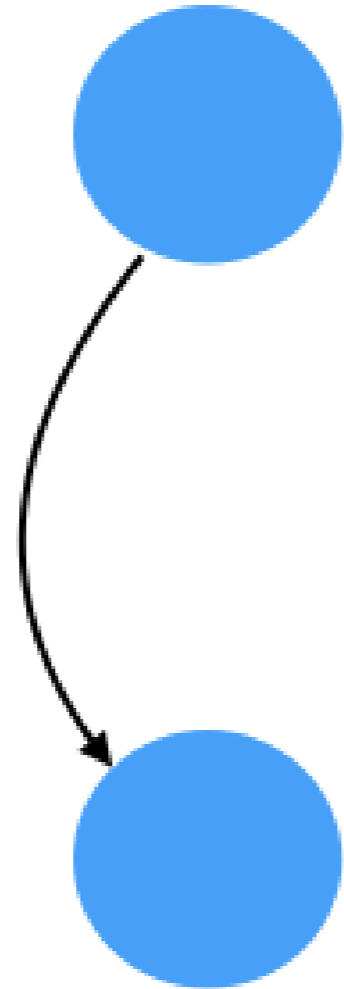
```
sg <- induced_subgraph(amzn_g, 1:500)
sg <- delete.vertices(sg, degree(sg) == 0)
plot(sg, vertex.label = NA, edge.arrow.width = 0,
     edge.arrow.size = 0, margin = 0, vertex.size = 2)
```



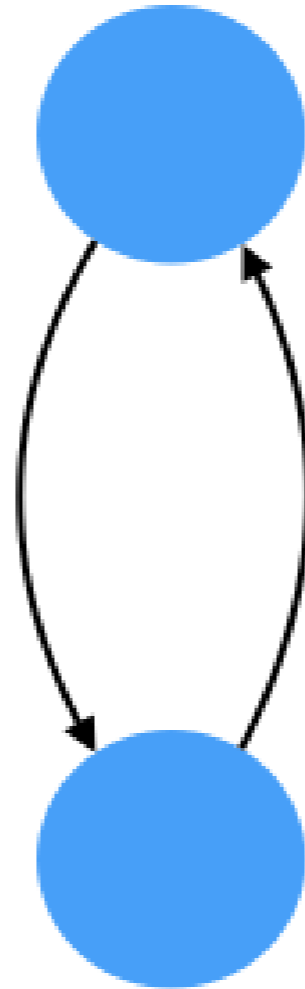
**Null**

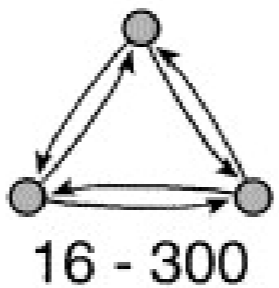
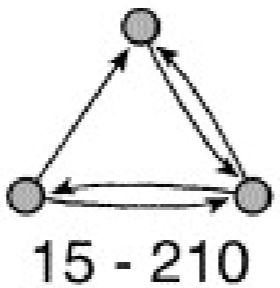
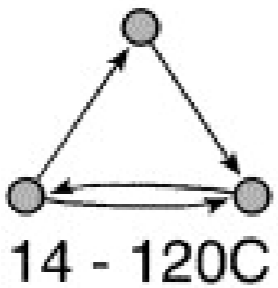
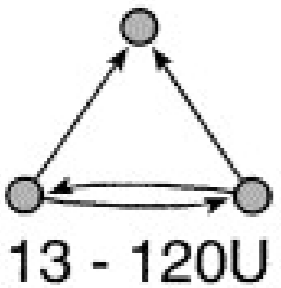
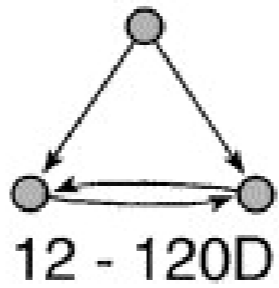
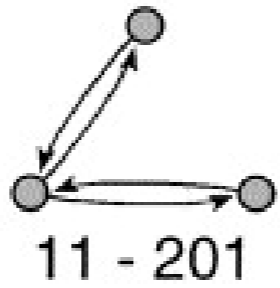
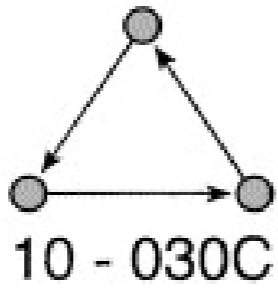
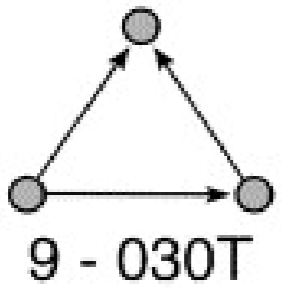
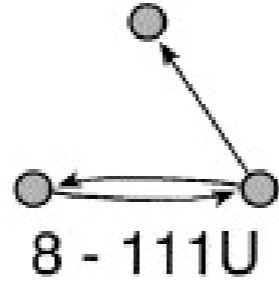
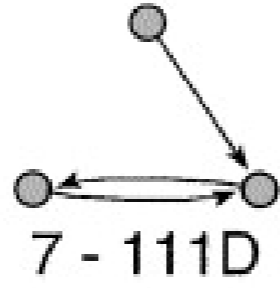
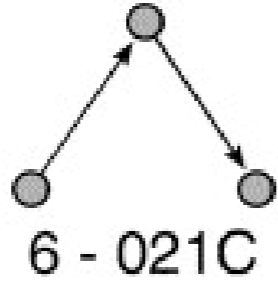
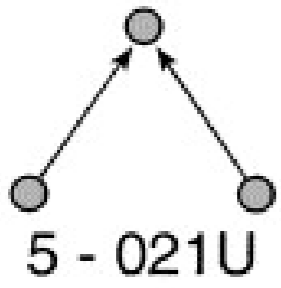
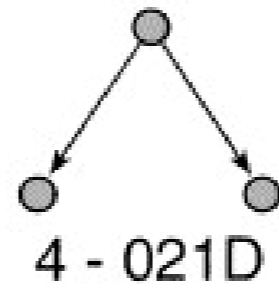
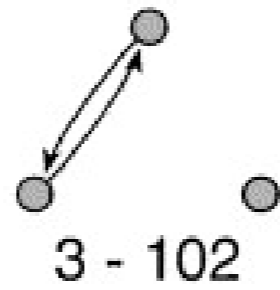
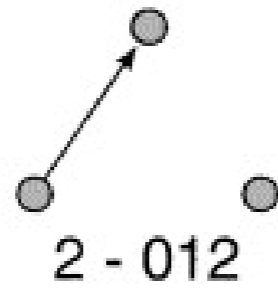
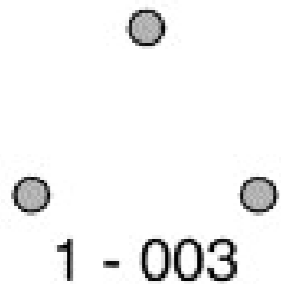


**Asymmetric**



**Mutual**





# Let's practice

CASE STUDIES: NETWORK ANALYSIS IN R

# Exploring temporal structure

CASE STUDIES: NETWORK ANALYSIS IN R



**Edmund Hart**  
Instructor



# Are important products always important?

```
# Get unique Dates
d <- sort(unique(amzn_raw$date))
# Create graph from first date
amzn_g <- graph_from_data_frame(
  amzn_raw %>%
  filter(date == d[1]) %>%
  select(from, to), directed = TRUE
)
```

# Are important products always important?

```
# Find products that are "important"
high_out_degree <- degree(amzn_g, mode = "out") > 2

low_in_degree <- degree(amzn_g, mode = "in") < 1

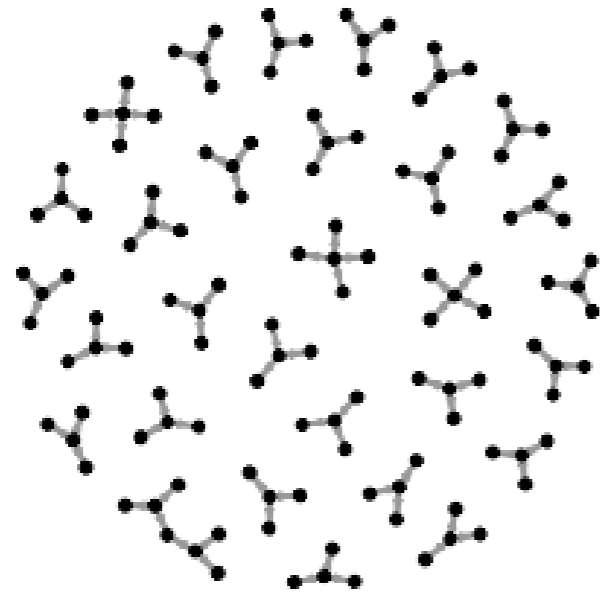
important_nodes <- high_out_degree & low_in_degree

imp_prod <- V(amzn_g)[important_nodes]

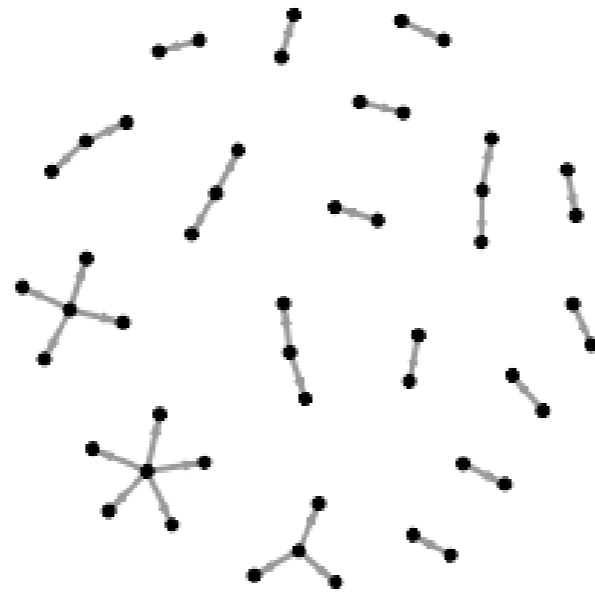
# Store as a data frame to later join on
tmp_df <- data.frame(imp_prod = as.numeric(names(imp_prod)))
```

```
## Create list to hold output
time_graph <- list()
## Create a 2x2 layout for plots and increase margins
par(mfrow = c(2, 2), mar = c(1.1, 1.1, 1.1, 1.1))
## Loop over the data to build
for(i in 1:length(d)){
  ## Create a data frame at each time stamp
  ip_df <- amzn_raw %>%
    filter(date == d[i]) %>%
    right_join(tmp_df, by = c("from" = "imp_prod")) %>%
    na.omit()
  ## Create an igraph object from that data frame
  time_graph[[i]] <- ip_df %>%
    select(from, to) %>%
    graph_from_data_frame(directed = TRUE)
  ## See what important vertices look like by date
  plot(time_graph[[i]], main = d[i]) }
}
```

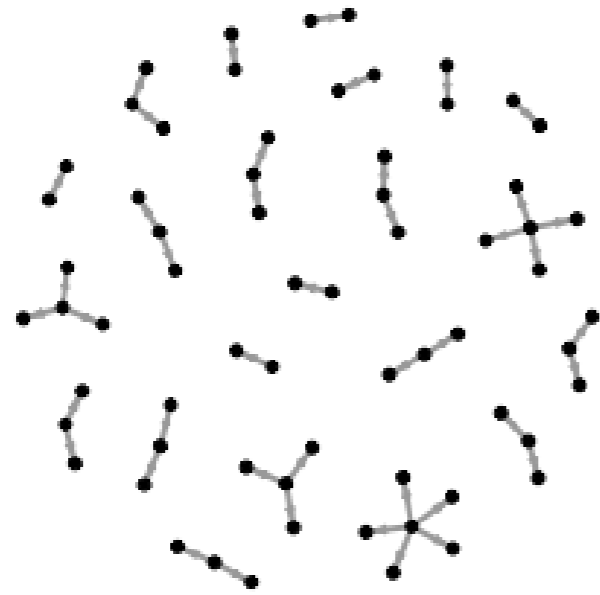
2003-03-02



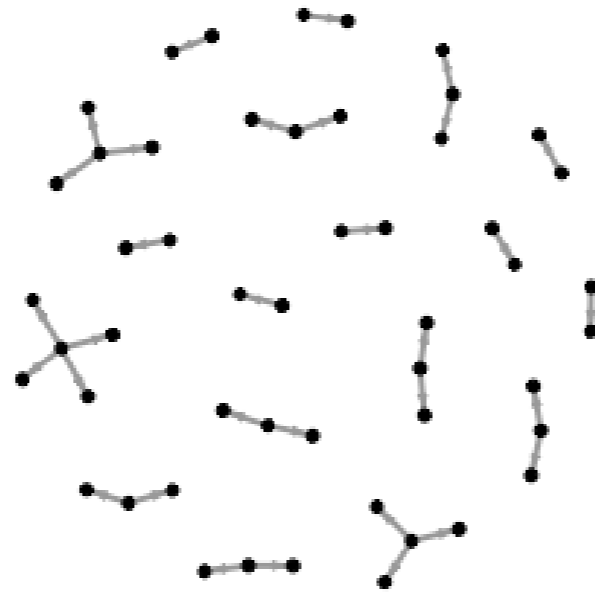
2003-03-12



2003-05-05



2003-06-01



# Let's practice!

CASE STUDIES: NETWORK ANALYSIS IN R