Social Networks

Connecting with others

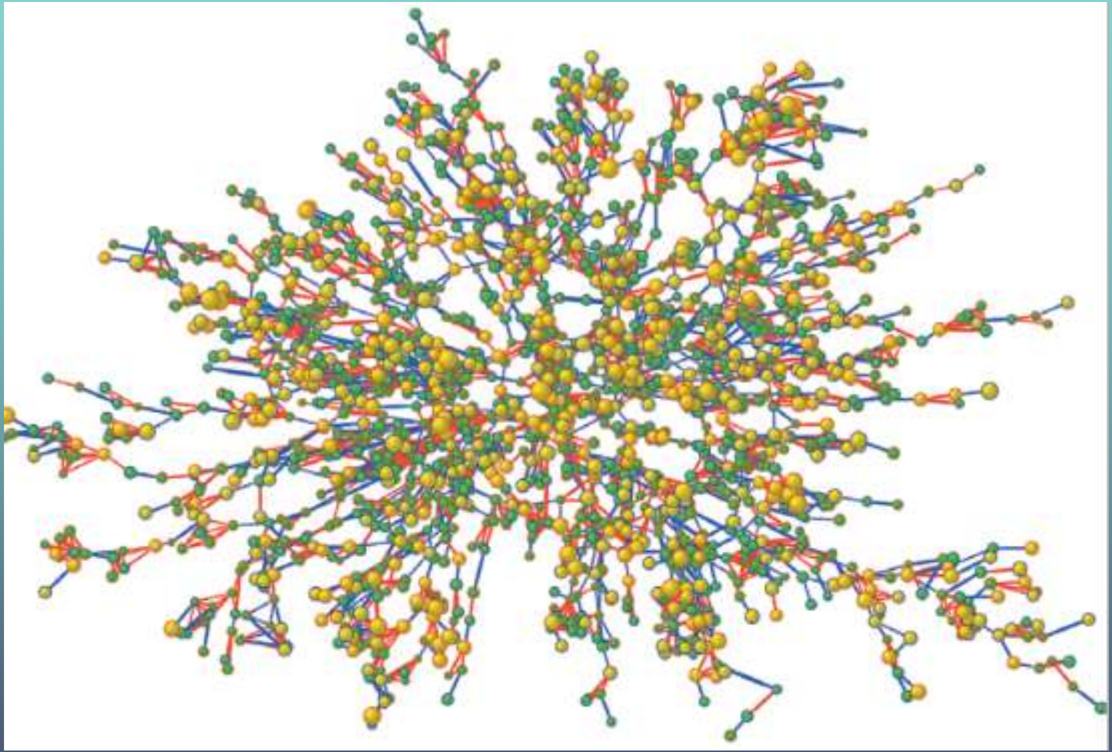
Obesity is an epidemic.

The NEW ENGLAND JOURNAL of MEDICINE

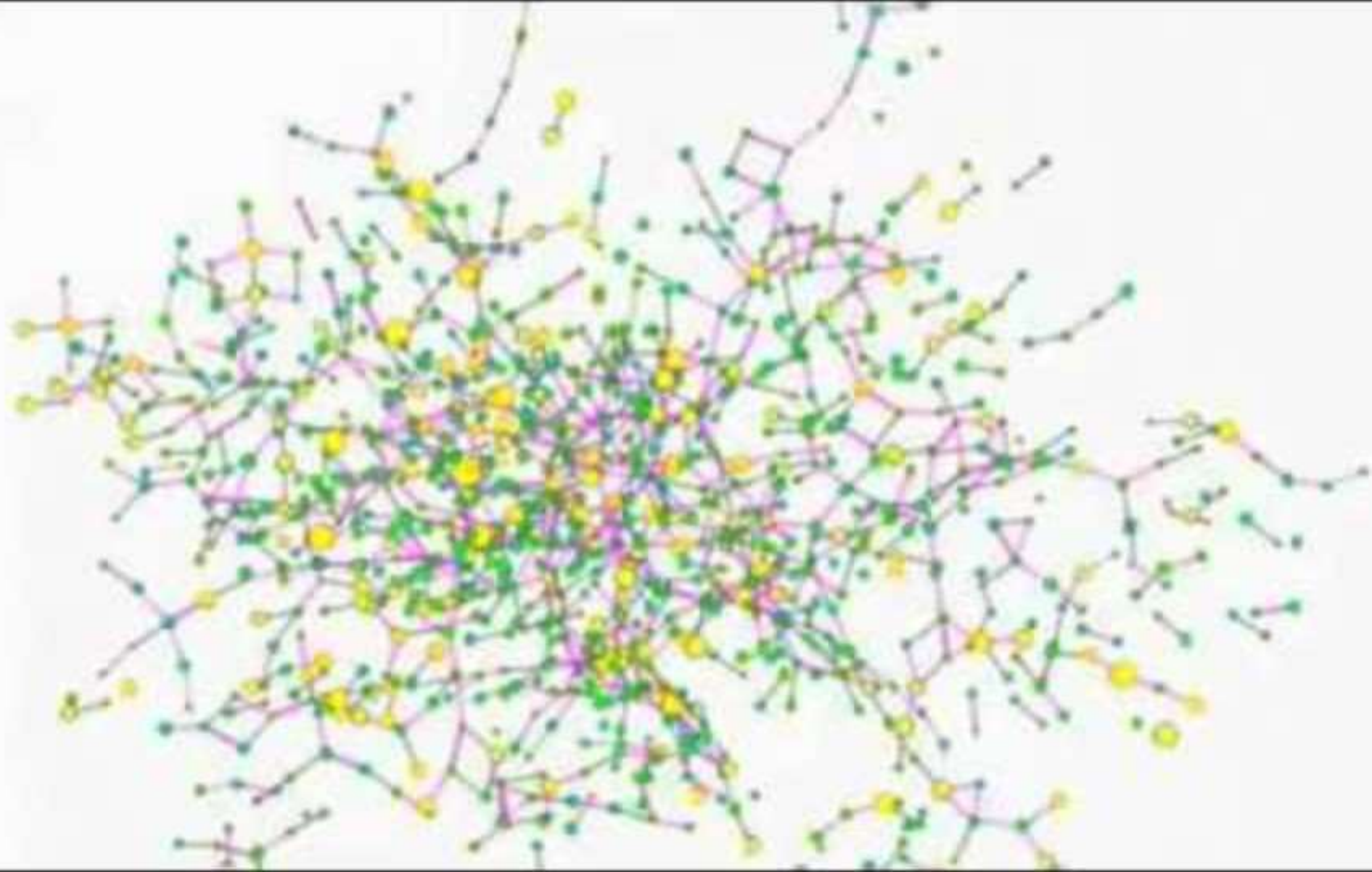
SPECIAL ARTICLE

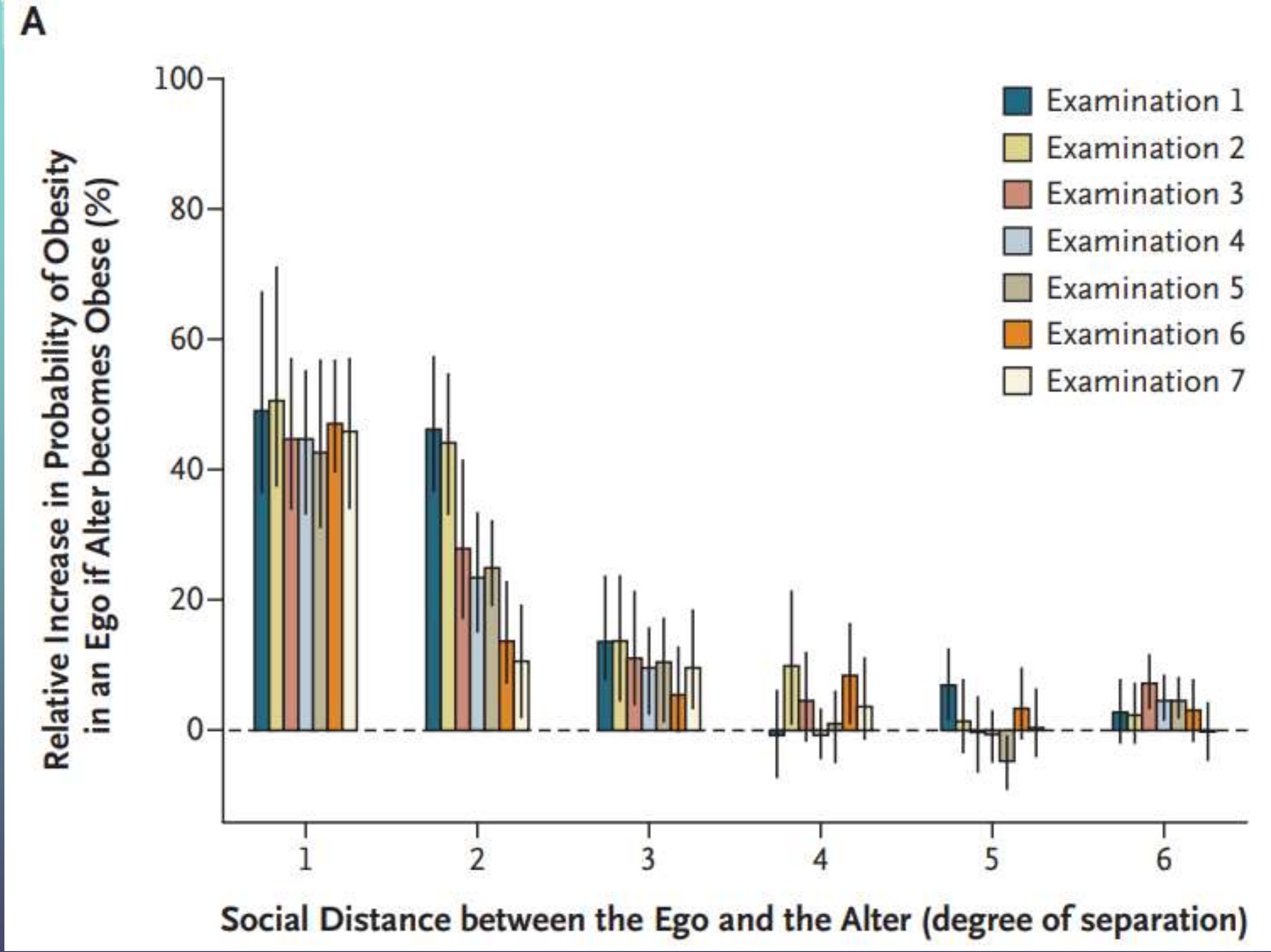
The Spread of Obesity in a Large Social Network over 32 Years

Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.



Node: individual; edge: connections; size of node: body mass index; yellow: obesity (i.e., BMI > 30)




A
Examination 1
Examination 2
Examination 3
Examination 4
Examination 5
Examination 6
Examination 7
Relative Increase in Probability of Obesity in an Ego if Alter becomes Obese (%)
100
80
60
40
20
0
1
2
3
4
5
6
Social Distance between the Ego and the Alter (degree of separation)

45%, 25%, and 10%

Q1

What are the benefits of leveraging online social networks?

Q2

How to describe the structure of social networks?

Q3

How to find important influencers in a social network?

What is the difference?

Web 1.0

Expedia
Google
eBay
Amazon.com
CNN.com
WSJ.com

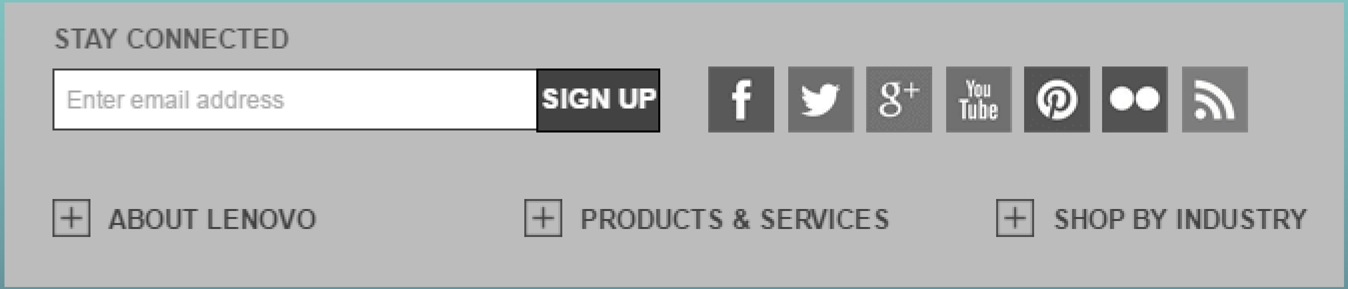
Web 2.0 and beyond

X (Twitter)
Snapchat
Instagram
Pinterest
Reddit
Wikipedia
Facebook

Customer-to-Customer Interactions

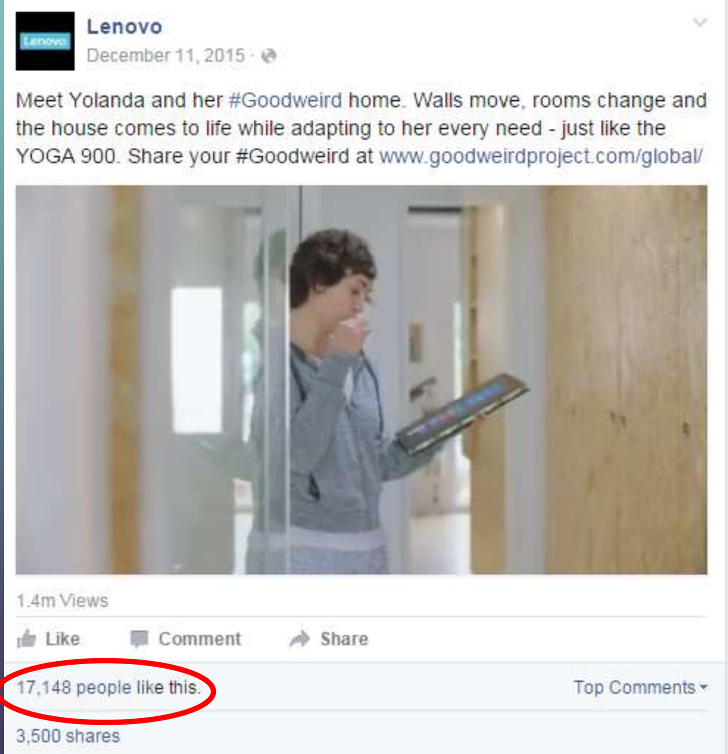
Leveraging Online Social Networks

Get Fans



The vast majority of large brands today have an active social media presence, such as FB fan page. For brands to resonate on Facebook, the first step is to accumulate your fan base.

Engage

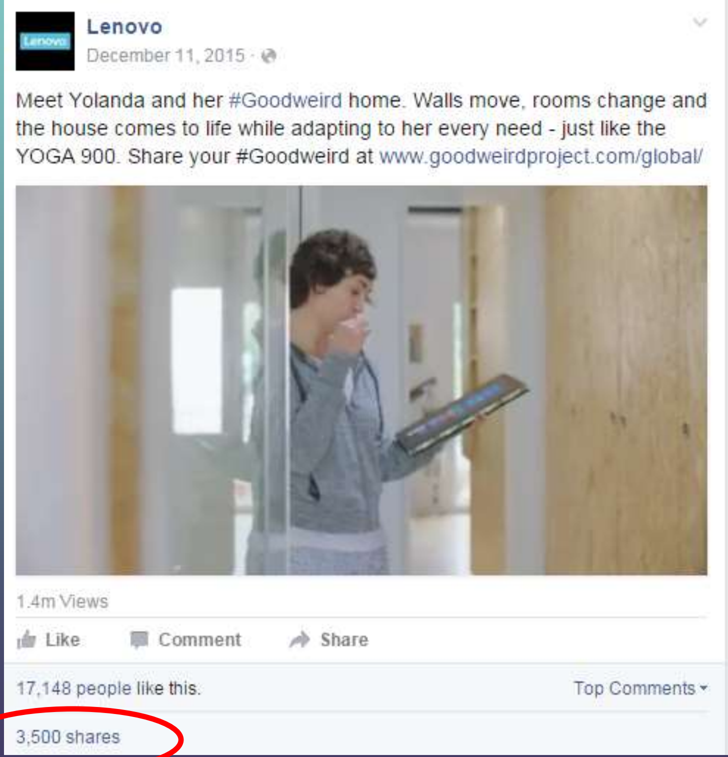


Brand messages only reach subset of fans.

Users that engage in fan page more likely to receive messages on news feed.

Users can engage by *liking, sharing, posting, commenting and checking in*.

Amplify



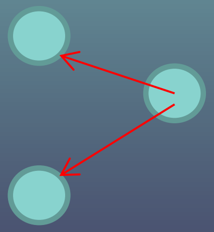
Spread brand message across social network (i.e., newsfeed).

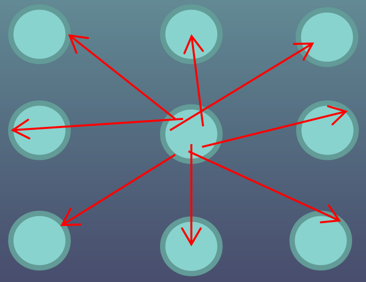
Organic word-of-mouth advertising.

Network can also be used for social advertising.

Amplification Ratio

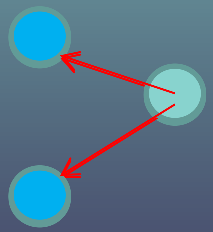
Amplification ratio
= # Friends of Fans exposed / # Fans exposed

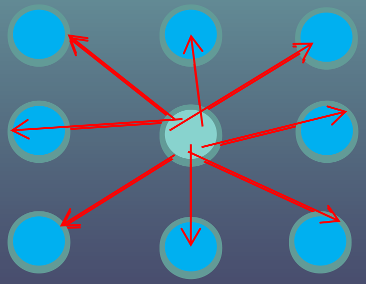




Amplification Ratio

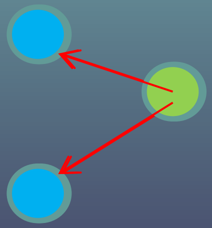
Amplification ratio
= # Friends of Fans exposed / # Fans exposed

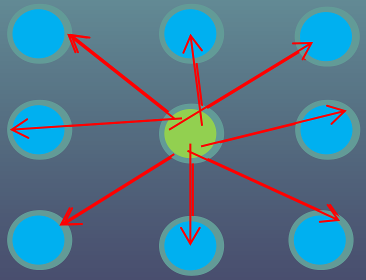




Amplification Ratio

Amplification ratio
= # Friends of Fans exposed / # Fans exposed
= 10/2 = 5.





AMPLIFICATION RATIO

Here are some facts. According to FB:

The top ten corporate brands had an average Amplification Ratio average of 1.05 (Range: 0.42 to 2.18).

The top 100 brands (excluding Celebrities & Entertainment) had an average Amplification Ratio of 0.84 (Range: 0.06 to 2.87).

CASE STUDY: Holiday Sales

Case study focused on Amazon, Best Buy, Target and Walmart.

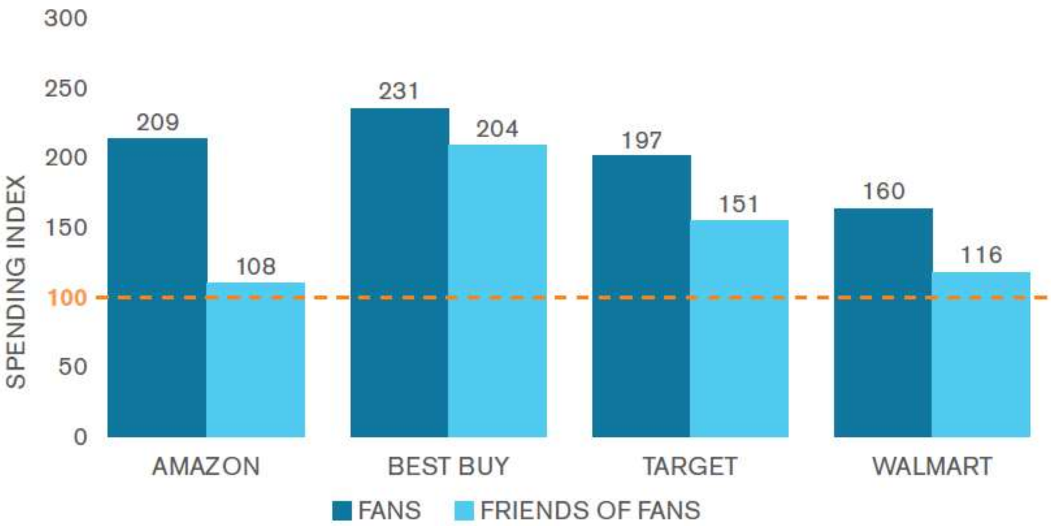
Retailers offered Facebook fans Black Friday deals.

Friends of Fans received notifications about their friends becoming fans, which lead to increased amplification.

Online and offline purchases of exposed fans and Friends of Fans compared to typical week.

CASE STUDY: Holiday Sales

Figure 3 Fans & Friends of Fans: Spending Index for Leading Retail Brands
Online & In-Store Purchase Behavior
Source: comScore Social Essentials, U.S., November-December 2011



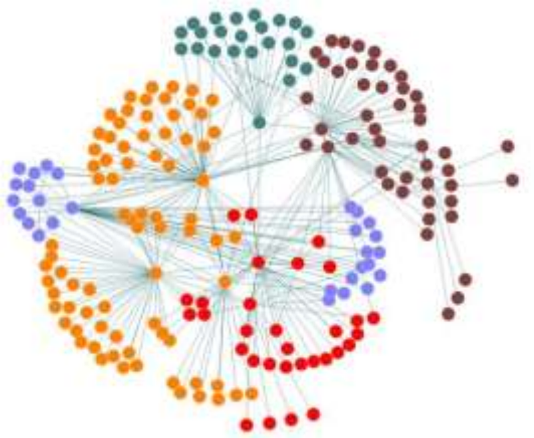
INDEX OF 100 = SEGMENT SPENT AS MUCH, ON AVERAGE, AS THE GENERAL POPULATION

Social Networks Analysis: Theory

Social Network Analysis Useful for…

Spotting influential people
- Who has a lot of linkages?
- Who is vital at linking people up?
- Why not just looking at no. of friends/followers?
  - Strength of tie

Understanding how connected the network is
- How many people are connected?
- What is the longest path between people?
- How to measure the density of a network?



Metrics

Individual

Has meaning independently of social network

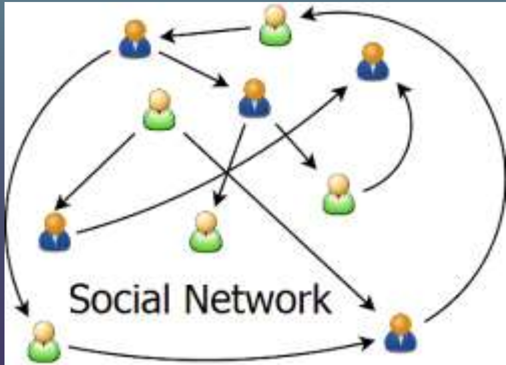
You live in Hong Kong island, HK

Connection

You are close friends with 10 people at HKU

Whole Network

On average, students know each other within 4 steps



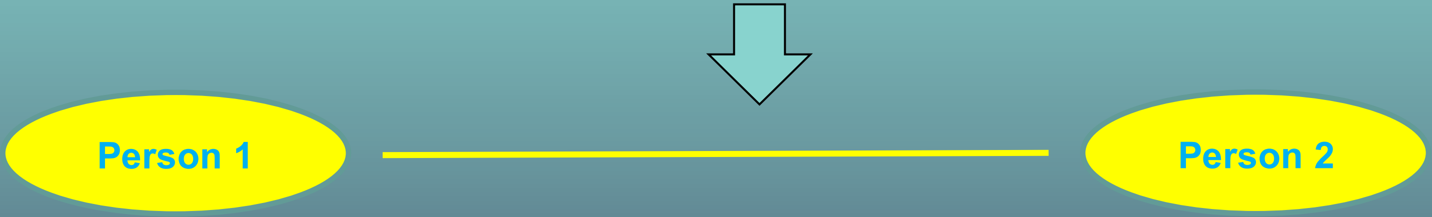
Edges

Person 1

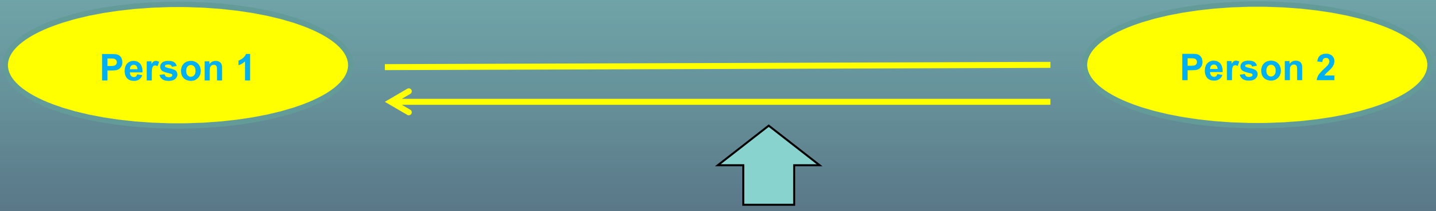
Edges

Edges

Undirected (e.g., study at HKU)

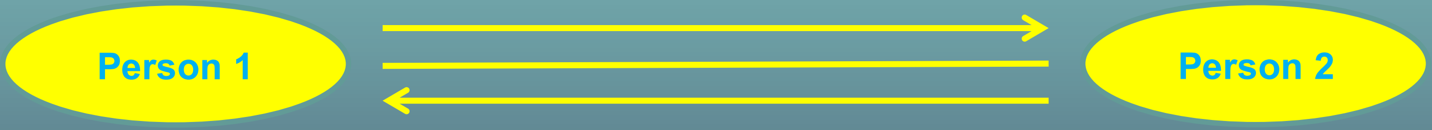


Edges



Directed (e.g., 2 follows 1)

Edges



Edges are also called links or ties.

Nodes and Edges

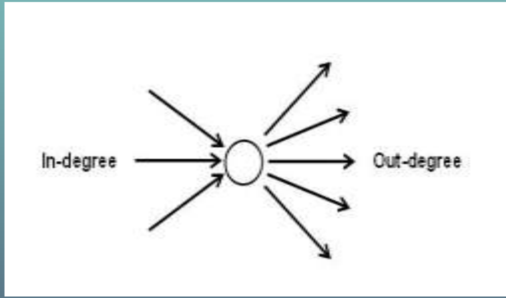
Vertex/Node: an end point
Often a person

Edge/Link: What connects up the Nodes
A relationship

Maximum number of edges in group of size $N(N-1)/2$.
Where everyone connects to everyone else
If undirected (my friends also have me as a friend)

Who is well-connected?

Degree (centrality): The number of linkages you have.
  "In-degree", e.g., someone that follows me.
  "Out-degree", e.g., I follow someone else.



Edge Weight
  Sometimes edge can also carry weight
  Can capture how deep the relationships are
  E.g., frequency of interactions between two nodes.


Closeness Centrality
& Betweeness Centrality

Closeness Centrality

Only applies to a fully connected network (i.e., a path exists between any pair of nodes).

$$\text{Closeness Centrality}(x) = \frac{N - 1}{\sum_y d(x, y)}$$

N: number of nodes in the network
d(x, y): the shortest distance between nodes x and y.

Betweenness Centrality

Applies to disconnected networks as well.

$$\text{Between Centrality(x)} = \sum_{y,z} \frac{\sigma_{yz}(x)}{\sigma_{yz}}$$

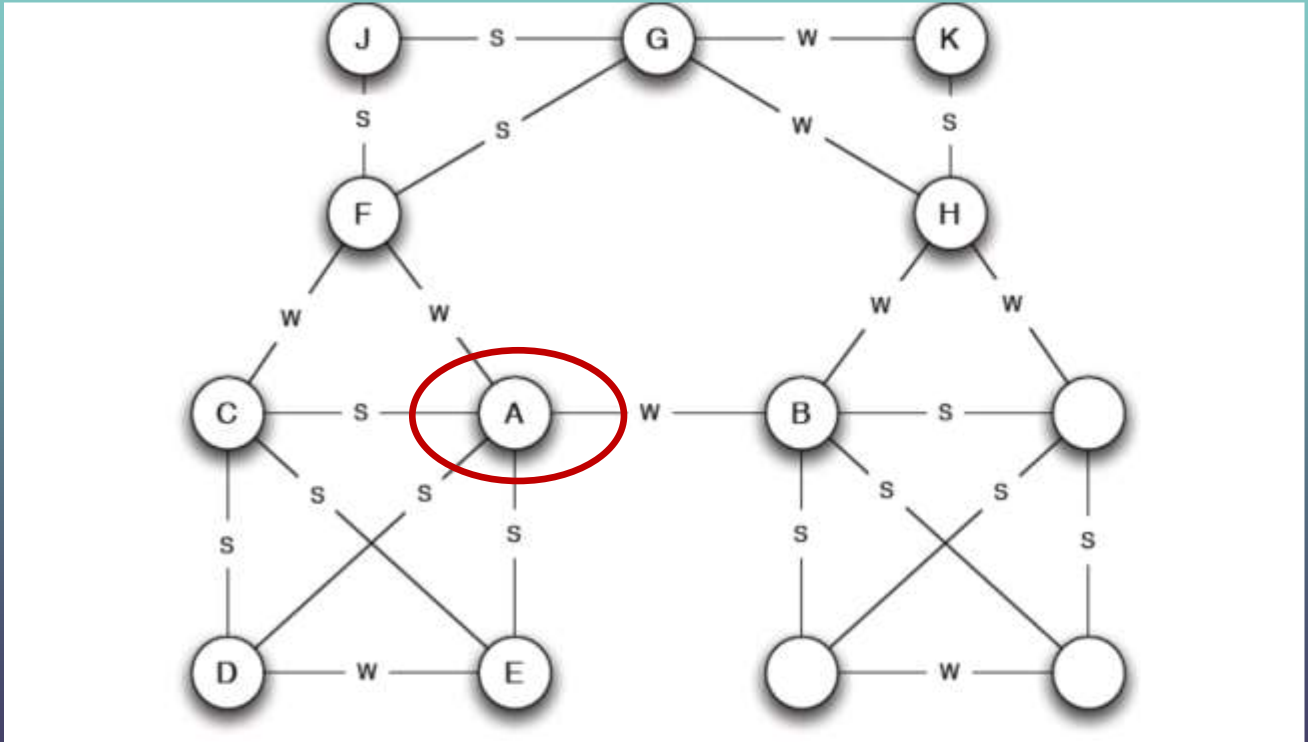
σ_{yz} is the total number of shortest paths from y to z.
$\sigma_{yz}(x)$ is the number of shortest paths from y to z that go through x.

Strong vs. Weak Ties

Suppose that two individuals are connected in a social network (i.e., they know each other).

However, the strength of their connection may differ: It may be a strong tie (i.e., they are friends) or a weak tie (they are acquaintances).

Strong vs. Weak Ties



Strong vs. Weak Ties

Although strong ties generally exert more normative influence, weak ties often have more informational influence.

Why?

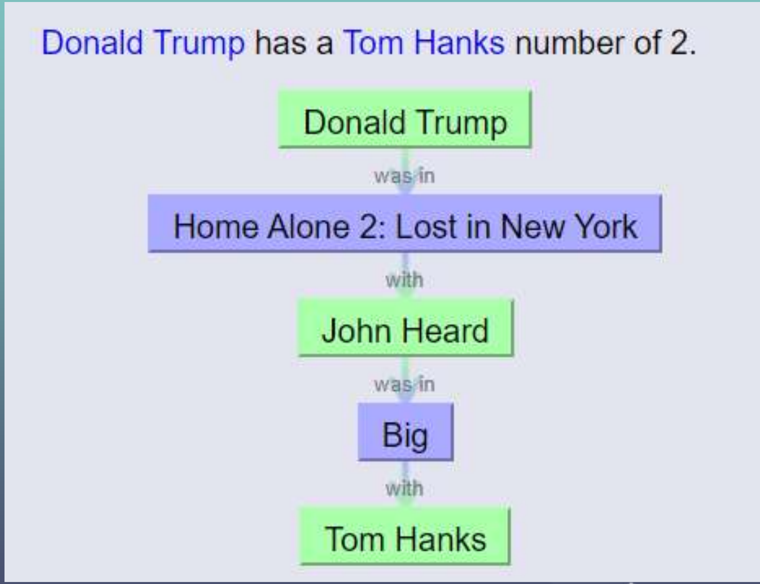
Because different social circles have different info, i.e., you probably know what your good friends know. Most jobs are found through weak connections.

Degrees of Separation

Path of how many people are needed to connect people up
Technical name: Geodesic distance

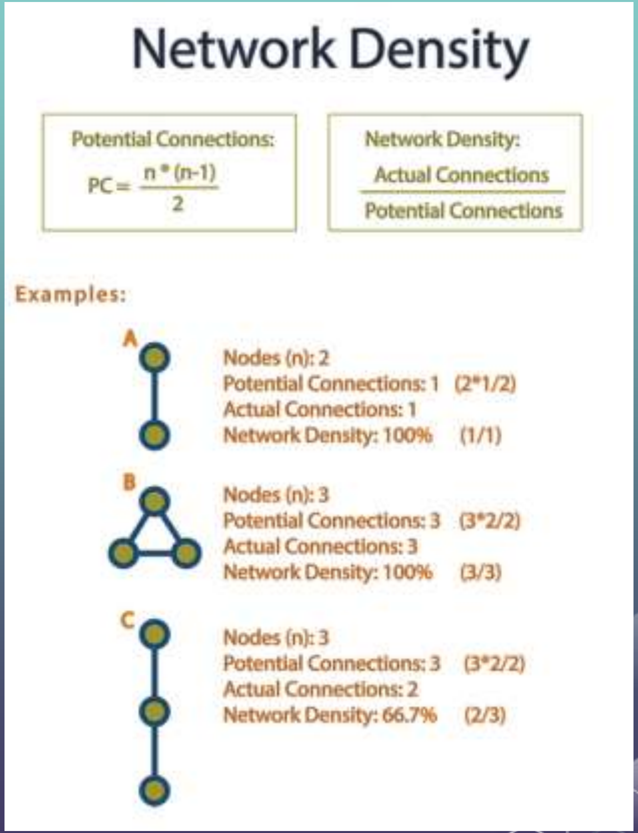
6 is the magical number: Kevin Bacon game (Link)

Don't fixate on 6! It does not apply to all networks!



Is a Network Well-Connected?

Graph/network density



Networks Analysis in R

Preparing Packages

```
library(igraph)
library(readr)
```

The "igragh" package provides you tools for network analysis while the "readr" facilitates reading data.

Reading Data

```
actors <- read_csv("https://ximarketing.github.io/class/DM//Actors.csv")
movies <- read_csv("https://ximarketing.github.io/class/DM/Movies.csv")
```

Here, the first file contains the nodes information, whereas the second file contains the edge information. Each actor/actress is a node, and if two actors/actresses appear in a same movie, there is an edge between them.

Reading Data

Actor Information (nodes):

```
> head(actors)
# A tibble: 6 x 3
  Actor        Gender BestActorActress
  <chr>        <chr>  <chr>
1 Tom Hanks    Male   Winner
2 Gary Sinise  Male   None
3 Robin Wright Female None
4 Bill Paxton  Male   None
5 Kevin Bacon  Male   None
6 Ed Harris    Male   Nominated
```

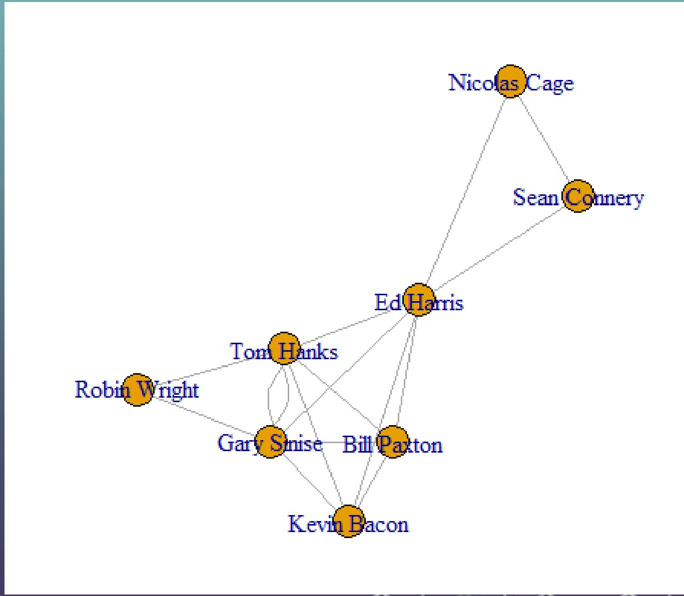
Reading Data

Movie Information (edges):

```
> head(movies)
# A tibble: 6 x 3
  `Actor 1`    `Actor 2`    Movie
  <chr>        <chr>        <chr>
1 Tom Hanks    Gary Sinise  Forest Gump
2 Tom Hanks    Robin Wright Forest Gump
3 Gary Sinise  Robin Wright Forest Gump
4 Tom Hanks    Gary Sinise  Apollo 13
5 Tom Hanks    Bill Paxton  Apollo 13
6 Tom Hanks    Kevin Bacon  Apollo 13
```

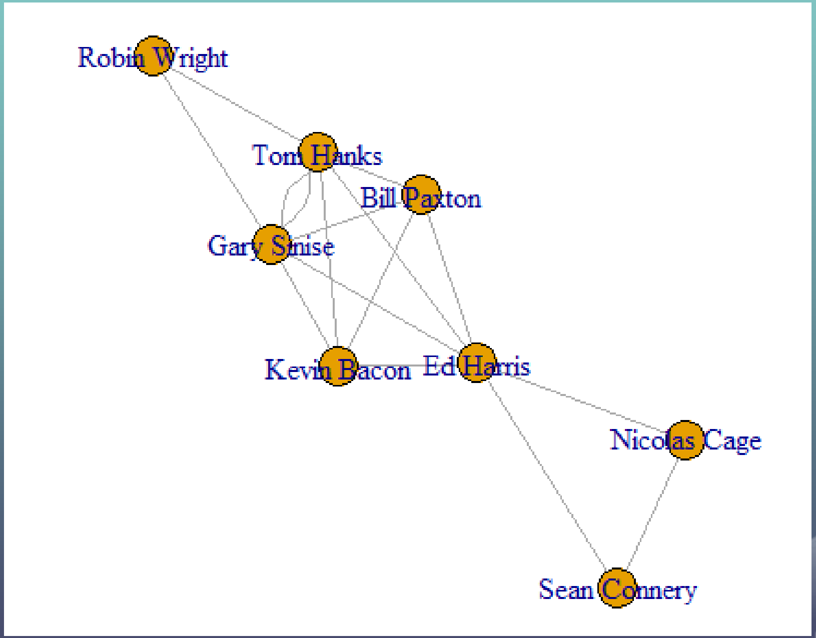
Visualize the Network

```
actorNetwork <-
graph_from_data_frame(d=movies,
vertices=actors, directed=F)
plot(actorNetwork)
```



Visualize the Network

```
plot(actorNetwork)
```

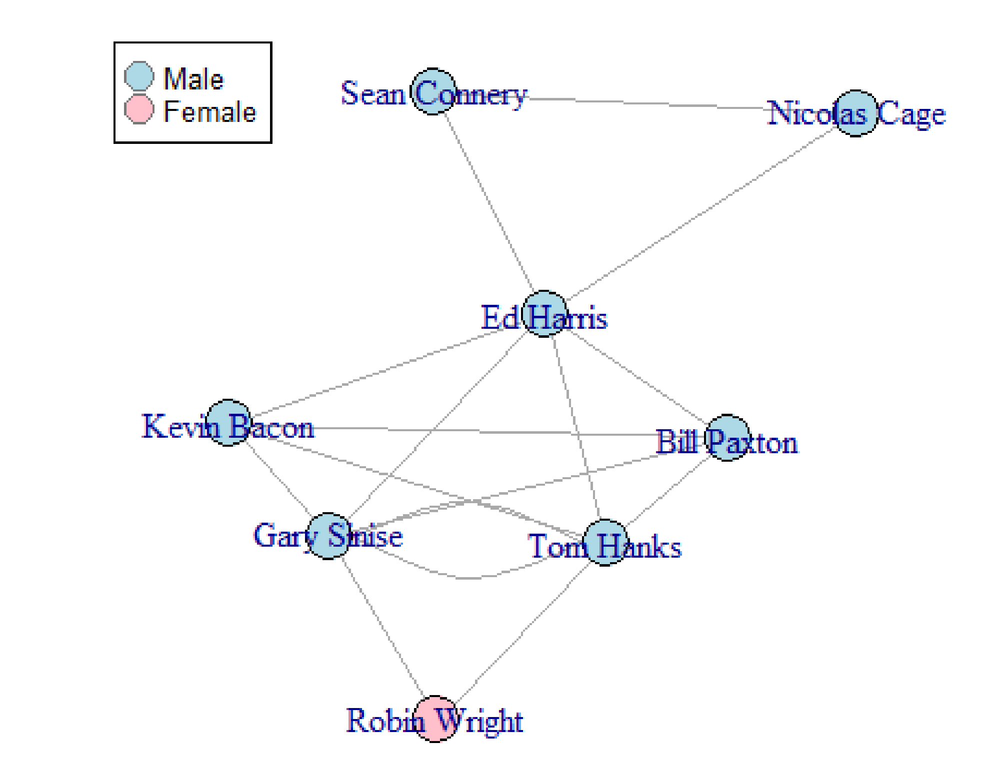


Visualize the Network

You can also add colors to your nodes:

```
V(actorNetwork)$color <-
ifelse(V(actorNetwork)$Gender == "Male",
"lightblue", "pink")
plot(actorNetwork)
legend("topleft", c("Male","Female"),
pch=21,
  col="#777777",
pt.bg=c("lightblue","pink"), pt.cex=2,
cex=.8)
```

Visualize the Network



Degree of the nodes

To check the degree of nodes in the network:

```
degree(actorNetwork, mode="all")
```

```
   Tom Hanks   Gary Sinise Robin Wright   Bill Paxton   Kevin Bacon     Ed Harris
           6             6            2             4             4             6
Sean Connery Nicolas Cage
           2            2
```

Closeness/Betweenness Centrality

```
closeness(actorNetwork, mode="all",
weights=NA, normalized=T)
```

```
   Tom Hanks  Gary Sinise Robin Wright  Bill Paxton  Kevin Bacon     Ed Harris
   0.7777778    0.7777778    0.5000000    0.7000000    0.7000000    0.8750000
Sean Connery Nicolas Cage
   0.5384615    0.5384615
```

```
betweenness(actorNetwork, directed=F,
weights=NA, normalized = T)
```

```
   Tom Hanks  Gary Sinise Robin Wright  Bill Paxton  Kevin Bacon     Ed Harris
   0.1190476    0.1190476    0.0000000    0.0000000    0.0000000    0.4761905
Sean Connery Nicolas Cage
   0.0000000    0.0000000
```

Network Density

```
edge_density(actorNetwork)
```

Exercise

There are another two files containing social networks of movie actors and actress. Play with these files yourselves! The files are downloadable here:

```
actors <-
read_csv("https://ximarketing.github.io/class/DM//ActorsExercise.csv")
movies <-
read_csv("https://ximarketing.github.io/class/DM/MoviesExercise.csv")
```

Directed Network

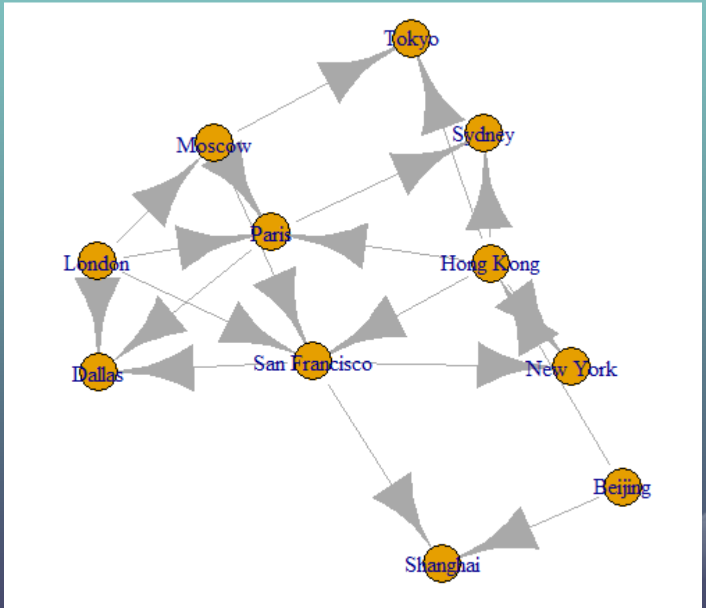
In the following exercise, we play with directed network. This is not much difference.

```
cities <-
read_csv("https://ximarketing.github.io/class/DM/Dir
ectedNodes.csv")
routes <-
read_csv("https://ximarketing.github.io/class/DM/Dir
ectedEdges.csv")
flightNetwork <- graph_from_data_frame(d=routes,
vertices=cities, directed=T)
```

Directed Network

Plot the directed network:

```
plot(flightNetwork)
```



Directed Network

We can distinguish between in-degrees and out-degrees:

```
degree(flightNetwork, mode="in")
```

```
      Beijing      Shanghai     Hong Kong         Tokyo      New York        London
            0             2             1             2             2             0
       Sydney San Francisco         Paris        Moscow        Dallas
            2             3             4             1             3
```

```
degree(flightNetwork, mode="out")
```

```
      Beijing      Shanghai     Hong Kong         Tokyo      New York        London
            2             0             5             0             0             5
       Sydney San Francisco         Paris        Moscow        Dallas
            0             3             2             3             0
```

Mobile

Is mobile different?

Mobile is the closest a brand can get to its customer.

APP Marketing

APP Pricing

Almost 94 percent of all apps are free to the consumer.

Some do charge for the initial download, but there is significant price resistance with apps, even when they are sold for less than a dollar.

APP Pricing

Getting people to pay for apps is not an easy task since consumers are used to getting apps for no charge. Therefore, there are also freemium apps. A freemium app is free to download, but users may need to pay a la carte for enhancements to the app.

In other cases, such as with the game Angry Birds, the app is free to download but comes with advertising included. If the user wants to eliminate the advertisements that come up between levels of the game, he needs to pay for the upgrade.