Social Network Analysis: Introduction

Donglei Du
(ddu@unb.ca)

Faculty of Business Administration, University of New Brunswick, NB Canada Fredericton E3B 9Y2

Table of contents

FYI: Social Media Video 2013: SocialNomics

- http://www.youtube.com/watch?v=TXD-Uqx6_Wk&feature=player_embedded

What is social network analysis?

- Study the structure and function of complex/emergent (unexpected/unpredictable) social network via various dynamical processes occurring on top of them.
  - Many alternative definitions exist, e.g. one is here: http://lrs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/
- SNA is a branch of **Network Science**, which is an attempt to understand networks emerging in nature, technology and society using a unified set of tools and principles.
  - Different networks emerge and evolve, driven by a fundamental set of laws and mechanisms.

Further readings

- Borgatti et al. (2009)
- Butts (2009)
- Watts (2007)
- Barabási (2012)
- Scott and Carrington (2011); Wasserman (1994)

Why social network analysis?

- Networks are everywhere
- Networks exhibit interesting phenomenon
- Networks analysis are useful
- ....

Networks are everywhere: Facebook friend network

- http://www.youtube.com/watch?v=9n9irapdON4&feature=player_detailpage

Networks are everywhere: Twitter: retweet network

- http://www.youtube.com/watch?feature=player_embedded&v=2guKJfvq4uI

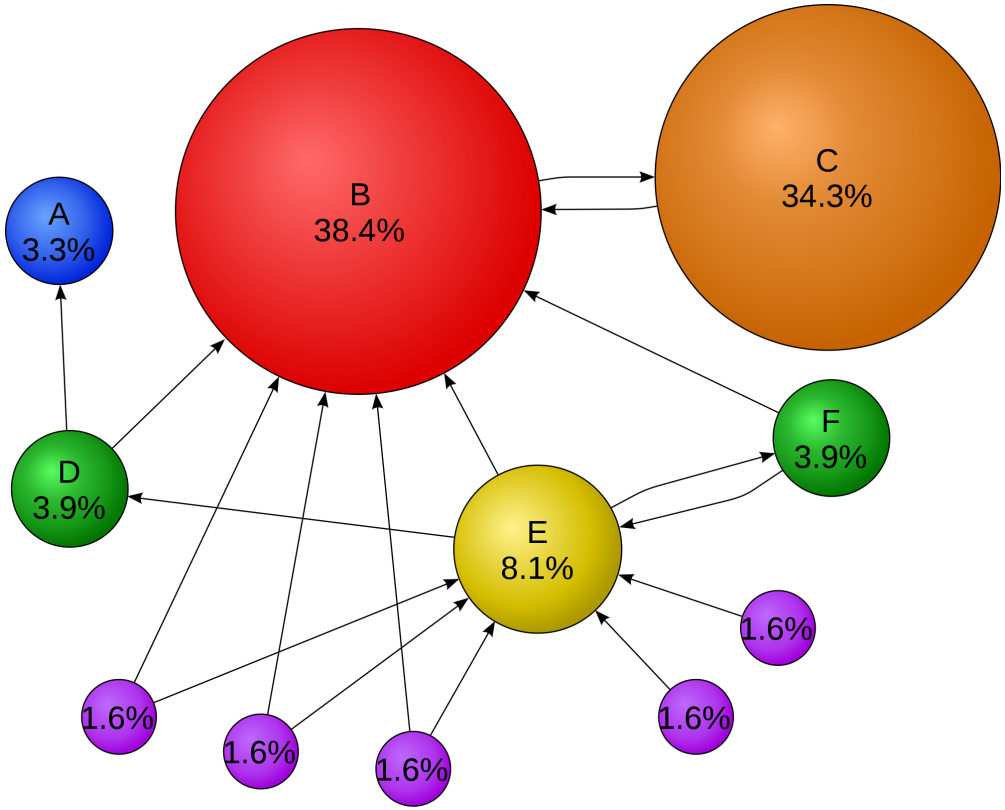
Networks are everywhere: Political Network: Obama In The Media

- http://www.youtube.com/watch?v=5etSid8G6EU&feature=player_detailpage

Networks are everywhere: the Spread of Obesity

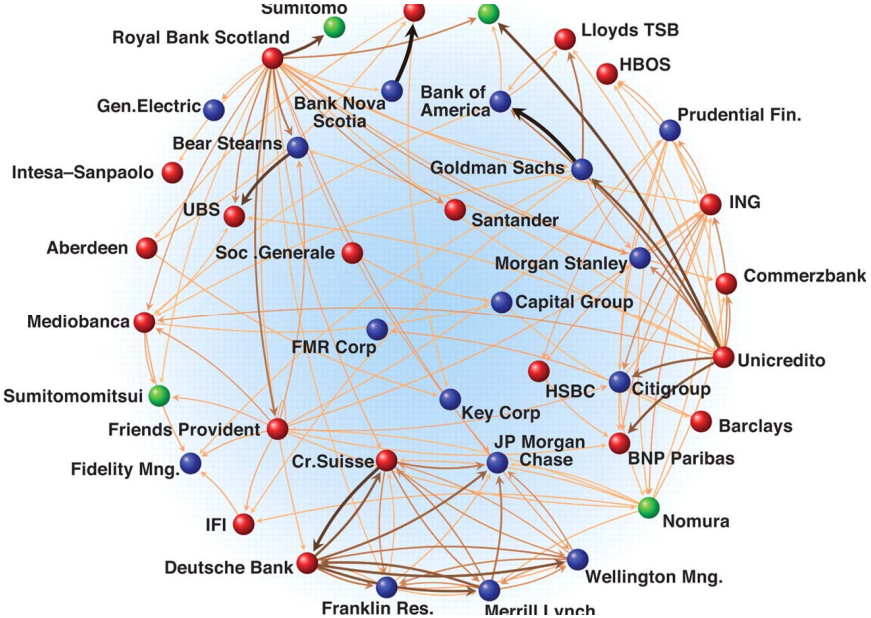
- http://www.youtube.com/watch?v=8aEtyRD1j5U&feature=player_embedded

Networks are everywhere: The Web: Google PageRank



- http://stackoverflow.com/questions/12268697/how-to-sort-and-visualize-a-directed-graph

Networks are everywhere: International Financial Network



European Union members (red), North America (blue), other countries (green). This indicates that the financial sector is strongly interdependent, which may affect market competition and systemic risk and make the network vulnerable to instability.

- http://www.sciencemag.org/content/325/5939/422/F2.expansion.html

Networks exhibit interesting phenomenon

- Small world phenomenon or six degree of separation
- The friendship paradox
- Scale-free degree distribution of real networks
- Strength of weak ties
- Giant component
- ...

Vast applications in different displines

- SNA has its origin from sociology and has gained a significant following in

- anthropology,
- biology
- communication studies
- economics
- geography
- history
- information science
- organizational studies
- political science
- social psychology
- development studies
- sociolinguistics
- ...

Networks analysis are useful in practice

- PageRank by Google
- Google trend in prediction: flu: Ginsberg et al. (2009), stock: Preis et al. (2013)
- Graph Search by Facebook
- EdgeRank by Twitter
- Sentiment analysis of Twitter
- Early detection of flu
- The Internet
- Team and collaboration: `//www.nature.com/news/2008/081008/full/455720a.html`
- Biological network Bastolla et al. (2009), Sugihara and Ye (2009)
- Brain network
- Disease network
- Gene network
- Economy network Schweitzer et al. (2009)
- Organization network Piepenbrink and Gaur (2013)
- Recipe network
- Financial network: `http://www.nature.com/news/2008/080201/full/news.2008.541.html`
- Terrorist network
- Movie box office prediction
- Stock market prediction
- ...

Tools needed to analyze social network

- Graph theory
- Game theory
- Statistics
- Computer science
- Statics physics
- ...

Topis to be covered

- **Basic graph and network knowledge**: degree, path, connectivity, distance, diameter, Breadth-first search, betweeness, clustering coefficient, etc.
- **Basic game theory knowledge**: Nash Equilibrium, dominated strategy, and dynamic games etc.
- **Network Structure**: Strong and weak ties, Centrality and prestige, Positivity and negative relationship, Clustering, Diameter, Giant component etc.
- **Network dynamics**: population modes: Power law distribution, Rich-get-richer modes; and structural modes: Random network models, Erdos-Reyni, preferential attachment, Kleinberg, Cascading behavior in networks, Small-world phenomenon
- **World wide web and internet**: The structure of the web, PageRank, web search and link analysis

Online network data

- R package: library(igraphdata)
- Mark Newman's network data repository: http://www-personal.umich.edu/~mejn/netdata/
- Laszlo Barabasi's network data collection: http://www3.nd.edu/~networks/resources.htm
- Stanford Large Network Dataset Collection: http://snap.stanford.edu/data/
- Indiana University data set: http://iv.slis.indiana.edu/db/index.html
- UCINet data sets: http://vlado.fmf.uni-lj.si/pub/networks/data/UciNet/UciData.htm
- http://code.google.com/p/open-advertising-dataset/
- The UCI Network Data Repository : http://networkdata.ics.uci.edu/
- http://nexus.igraph.org/api/dataset_info
- Dataset in textbook "Data Mining and Business Analytics with R" by Johannes Ledolter: http://www.biz.uiowa.edu/faculty/jledolter/DataMining/datatext.html

Journals

- Nature: http://www.nature.com/
- Science: http://www.sciencemag.org/
- PNAS: http://www.pnas.org/
- Scientific Reports: http://www.nature.com/srep/index.html
- PLOS ONE: http://www.plosone.org/
- Social Networks: http://www.journals.elsevier.com/social-networks/

Software to be covered

- R
- Netlogo
- Gephi

Creating graphs in igraph

- The igraph homepage: http://igraph.sourceforge.net/
- igraph manual: http://cran.r-project.org/web/packages/igraph/index.html
- Tutorial site: http://igraph.sourceforge.net/igraphbook/igraphbook-creating.html

```
rm(list=ls())# clear memory
library(igraph)# load package igraph
...
```

Import network data from different resources: igraphdata

```
rm(list=ls())# clear memory
library(igraph)# load package igraph
library(igraphdata)# load package igraphdata

data(package="igraphdata") #get a list of data sets included in this package

>Data sets in package igraphdata:

>Koenigsberg Bridges of Koenigsberg from Euler's times
>UKfaculty        Friendship network of a UK university faculty
>USairports       US airport network, 2010 December
>foodwebs         A collection of food webs
>immuno           Immunoglobulin interaction network
>karate           Zachary's karate club network
>macaque          Visuotactile brain areas and connections
>yeast            Yeast protein interaction network

data(foodwebs) # read in a named list of directed igraph graph objects
foodwebs[[1]]

data(karate)  #Social network between members of a university karate club
plot(karate)
```

Import network data from different resources: Edge list

```
rm(list=ls())# clear memory
library(igraph)# load package igraph

##I. Edge lists: graph() and get.edgelist(): graph() id starts from 1.
g_el2 <- graph( c(1,2, 1,3, 2,3, 3,4 ))
summary(g1)
plot(g_el1)
# "directed" parameter can be changed to FALSE to create
#undirected graphs from the default directed  graphs
g_el2 <- graph( c(1,2, 1,3, 2,3, 3,4 ), directed=FALSE)
summary(g_el2)
plot(g_el2)

#If you happen to have the edge list of a graph in a two-column matrix
edgelist<-get.edgelist(g_el1) # get the deglist
g_el3<-graph( t(edgelist))
plot(g_el3)
```

Import network data from different resources: Adjacency matrices

```
rm(list=ls())# clear memory
library(igraph)# load package igraph
##II. Adjacency matrices:  graph.adjacency() and get.adjacency()
adjm_u<-matrix(
c(0, 1, 0, 0, 1, 0,
1, 0, 1, 0, 1, 0,
0, 1, 0, 1, 0, 0,
0, 0, 1, 0, 1, 1,
1, 1, 0, 1, 0, 0,
0, 0, 0, 1, 0, 0), # the data elements
nrow=6,                # number of rows
ncol=6,                # number of columns
byrow = TRUE)          # fill matrix by rows
g_adj_u <- graph.adjacency(adjm_u, mode="undirected")
plot(g_adj_u)
get.adjacency(g_adj_u) # get the adjacency matrix
```

Import network data from different resources: data frame on the fly

```
rm(list=ls())# clear memory
library(igraph)# load package igraph

## III. create graph from data frame after creating data frame:
# First createa data frame
node1 = c("Her", "You", "Him")
node2 = c("Him", "Her", "You")
weight = c(10, -2, 3)
df = data.frame(node1, node2, weight)
# Use graph.data.frame() to create a gaph
g <- graph.data.frame(df, directed=FALSE)
V(g)$name # node names
E(g)$weight # edge weights
plot(g)
```

Import network data from different resources: data frame from file

```
rm(list=ls())# clear memory
library(igraph)# load package igraph

## IV. create graph from data frame in file:  graph.data.frame()
file_path<-file.path(getwd(), "data/")
local_file<-paste(file_path, "g1.csv", sep="")
write.csv(df, file=local_file, row.names = FALSE) # write datafeme into a local f
df_g<-read.csv(local_file) # read graph data froma
# Use graph.data.frame() to create a gaph
g <- graph.data.frame(df_g, directed=FALSE)
plot(g)

# another example where the file alreay exists
local_file<-paste(file_path, "www.dat", sep="")
df_g<-read.table(local_file)
g <- graph.data.frame(df_g, directed=TRUE)
# or
g<-read.graph(local_file, directed=TRUE)
```

Case study: http://igraph.sourceforge.net/igraphbook/import.R

```
rm(list=ls())# clear memory
library(igraph)# load package igraph

# Read the files first
file_path<-file.path(getwd(), "data/")
local_file1<-paste(file_path, "traits.csv", sep="")
local_file2<-paste(file_path, "relations.csv", sep="")
traits <- read.csv(local_file1, head=FALSE)
rel <- read.csv(local_file2, head=FALSE)

# Create the graph, add the vertices
g <- graph.empty()
g <- add.vertices(g, nrow(traits),
                     name=as.character(traits[,1]), age=traits[,2],
                     gender=as.character(traits[,3]))

# Extract first names from the full names
names <- sapply(strsplit(V(g)$name, " "), "[",1)
ids <- 1:length(names)
names(ids) <- names

# Create the edges
from <- as.character(rel[,1])
to <- as.character(rel[,2])
edges <- matrix(c(ids[from], ids[to]), nc=2)

# Add the edges
g <- add.edges(g, t(edges),
               room=as.character(rel[,3]),
       friend=rel[,4], advice=rel[,5])
```

References I

Barabási, A.-L. (2012). Network science: Luck or reason. *Nature*, 489(7417):507–508.

Bastolla, U., Fortuna, M. A., Pascual-García, A., Ferrera, A., Luque, B., and Bascompte, J. (2009). The architecture of mutualistic networks minimizes competition and increases biodiversity. *Nature*, 458(7241):1018–1020.

Borgatti, S. P., Mehra, A., Brass, D. J., and Labianca, G. (2009). Network analysis in the social sciences. *science*, 323(5916):892–895.

Butts, C. T. (2009). Revisiting the foundations of network analysis. *science*, 325(5939):414–416.

References II

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014.

Piepenbrink, A. and Gaur, A. S. (2013). Methodological advances in the analysis of bipartite networks an illustration using board interlocks in indian firms. *Organizational Research Methods*, 16(3):474–496.

Preis, T., Moat, H. S., and Stanley, H. E. (2013). Quantifying trading behavior in financial markets using google trends. *Scientific reports*, 3.

Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., and White, D. R. (2009). Economic networks: The new challenges. *science*, 325(5939):422.

References III

Scott, J. and Carrington, P. J. (2011). *The SAGE handbook of social network analysis*. SAGE publications.

Sugihara, G. and Ye, H. (2009). Complex systems: Cooperative network dynamics. *Nature*, 458(7241):979–980.

Wasserman, S. (1994). *Social network analysis: Methods and applications*, volume 8. Cambridge university press.

Watts, D. J. (2007). A twenty-first century science. *Nature*, 445(7127):489–489.