R news and tutorials contributed by (563) R bloggers

Welcome!

Here you will find daily **news and tutorials about R**, contributed by over 555 bloggers. There are many ways to **follow us -**

By e-mail:

On Twitter:

On Facebook:

If you are an R blogger yourself you are invited to add your own R content feed to this site (**Non-English** R bloggers should add themselves- here)

Jobs for R-users

- Improve UI for our general online data analysis and visualization tools (R)
- Looking for a partner to code an algorithm which will trade base on twitter and other social data feed
- R Integration with D3 charts
- Data Analyst at uSwitch

- Director, Pew Research Center Labs

Search & Hit Enter

Popular Searches

- heatmap
- web scraping
- hadoop
- twitter
- PCA
- shiny
- finance
- boxplot
- maps
- ggplot
- ggplot2
- animation
- LaTeX
- trading
- time series
- excel
- quantmod
- how to import image file to R
- knitr
- googlevis
- tutorial
- rstudio
- market research
- eclipse
- rattle
- gis
- rcmdr
- ecdf
- coplot
- sweave

Recent Posts

- Microsoft to acquire Revolution Analytics: Now this gets interesting…
- RSS feeds for statistics and related journals
- R Open Source and Commercial Support
- Assertive R programming in dplyr/magrittr pipelines
- Revolution Analytics joins Microsoft
- Shiny 0.11, themes, and dashboard
- R Package to Download Fitbit Data
- A first look at Spark
- Data Wrangling Webinar and IFAR Chapter

Update
- Sequential Monte Carlo 2015 workshop
- Four Famous Laws & How You Can Visualize Them
- Easy error propagation in R
- How to reliably access network resources in R
- Practical introduction to Shiny – workshop write-up
- A beautiful story about NYC weather

Other sites

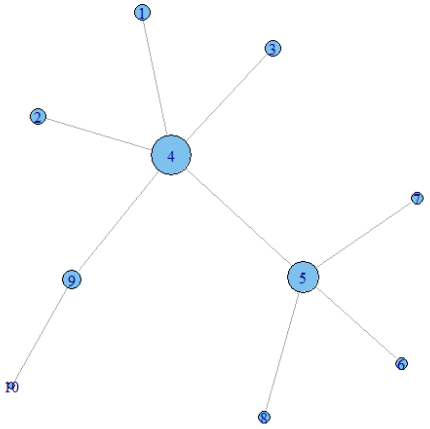
- Statistics of Israel
- SAS blogs
- Jobs for R-users

Using SNA in Predictive Modeling

April 10, 2012
By Matt Bogard

(This article was first published on **Econometric Sense**, and kindly contributed to R-bloggers)

In a previous post, I described the basics of social network analysis. I plan to extend that example here with an application in predictive analytics. Let's suppose we have the following network (visualized in R)



Suppose we have used the igraph package in R to derive measures of centrality, and we combined that information with other information gathered from our data base, for example income and loan default data. We can then incorporate the centrality measures into a model that predicts default.

Notwithstanding the issues related to linear regression with binary

dependent variables, the extremely small sample size, and the fact that this data is totally made up, for illustrative purposes, lets say we have the following data set:

id	income	eig	default
1	35000	0.403323	1
2	37000	0.403323	1
3	43000	0.403323	1
4	63000	1	0
5	72000	0.787754	0
6	27000	0.31772	1
7	30000	0.31772	1
8	34000	0.31772	1
9	45000	0.481678	1
10	55000	0.194272	1

A basic linear regression gives the following:

```
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.775e+00  1.424e-01  12.464 4.93e-06 ***
creditrisk$eig   -1.172e+00  2.642e-01  -4.438  0.00302 **
creditrisk$income -9.807e-06  4.362e-06  -2.249  0.05932 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1362 on 7 degrees of freedom
Multiple R-squared: 0.9188,     Adjusted R-squared: 0.8956
F-statistic:  39.6 on 2 and 7 DF,  p-value: 0.0001526
```

Although the data is made up, we can see that eigenvector centrality is a significant predictor in the regression. While logistic regression or decision trees may be more appropriate, I could not obtain results for illustrative purposes.

For a couple of actual applications with real data see:

Analysis of Variance Application:

Investigating Student Communities with Network Analysis of Interactions in a Physics Learning Center. Brewe, Eric; Kramer, Laird; O'Brien, George
2009 PHYSICS EDUCATION RESEARCH CONFERENCE. AIP Conference Proceedings, Volume 1179, pp. 105-108 (2009).

Path Analysis:

Ties That Bind: A Social Network Approach to Understanding Student Integration and Persistence
Scott L. Thomas. The Journal of Higher Education , Vol. 71, No. 5 (Sep. - Oct., 2000), pp. 591-615

The R used in this illustration follows:

```
# *------------------------------------------------------------
# | PROGRAM NAME: R_BASIC_SNA
# | DATE: 4/9/12
# | CREATED BY: MATT BOGARD
# | PROJECT FILE: P:\R  Code References\SNA
# *------------------------------------------------------------
# | PURPOSE: DEMONSTRATION OF BASIC CONCEPTS OF NETWORK ANALYSIS
# | REFERENCES: Conway, Drew. Social Network Analysis in R.
# | New York City R User Group Meetup Presentation August 6, 200
# | http://www.drewconway.com/zia/wp-content/uploads/2009/08/sna
# *------------------------------------------------------------

setwd("P:\\TOOLS AND REFERENCES (Copy)\\Social Network Analysis\

#---------------------------------------
#
# analytics
#
#---------------------------------------

library(igraph)

#---------------------------------------
# get data
#---------------------------------------
```

```
# specify the adjacency matrix
M <- matrix(c(0, 0, 0, 1, 0, 0, 0, 0, 0, 0,0, 0, 0, 1, 0, 0, 0,
G<-graph.adjacency(M, mode=c("undirected"))               # conv
cent<-data.frame(bet=betweenness(G),eig=evcent(G)$vector) # calc
res<-as.vector(lm(eig~bet,data=cent)$residuals)           # calc
cent<-transform(cent,res=res)                             # add
write.csv(cent,"r_keyactorcentrality.csv")                  # sa

#-------------------------------------
# visualize the network
#-------------------------------------

# plot that reflects correct vertex names and scaled by centrali

plot(G, layout = layout.fruchterman.reingold, vertex.size = 20*e

#-----------------------------------------
# create analysis data set
#-----------------------------------------

id <- c(1,2,3,4,5,6,7,8,9,10)  # create individual id's for refe
income <- c(35000, 37000, 43000, 63000, 72000, 27000, 30000, 340
default <- c(1,1,1,0,0,1,1,1,1,1) # default indicator

#-----------------------------------------
# basic regression
#-----------------------------------------

creditrisk <- cbind(id,income, cent, default) # combine with eig

# model default risk as a function of income and network relatio

# OLS
model1 <- lm(creditrisk$default~  creditrisk$eig + creditrisk$in
summary(model1)

#Logistic regression
model2 <- glm(creditrisk$default~  creditrisk$eig + creditrisk$i
summary(model2)

# decision tree
model3 <- rpart(creditrisk$default~  creditrisk$eig + creditrisk
summary(model3)
```

Created by Pretty R at inside-R.org

To **leave a comment** for the author, please follow the link and comment on his blog: **Econometric Sense**.

Tags: R, R code, social network analysis

Comments are closed.

1. Learn Statistics and R online from Harvard
2. Installing R packages
3. In-depth introduction to machine learning in 15 hours of expert videos
4. Scatterplots
5. Using apply, sapply, lapply in R
6. Basics of Histograms
7. Visualization series: Insight from Cleveland and Tufte on plotting numeric data by groups
8. R in Business Intelligence
9. K-means clustering is not a free lunch

Sponsors