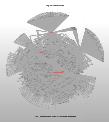
Affiliation networks

Visualization of small and real world data

Manuel S. González Canché & Cecilia Rios Aguilar
msgc@email.arizona.edu
Cecilia.Rios-Aguilar@cgu.edu

University of Arizona
Claremont Graduate University

INCHER, University of Kassel
June 26, 2012

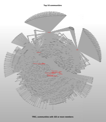


Purpose and Outline

At the end of the session(s) the participants are expected to understand and replicate the procedures followed to analyze and visualize affiliation networks using UCINET and R.

The specific topics to be covered are:

1. Displaying and analyzing affiliation data structures in UCINET
   - Displaying affiliation data
   - Using centrality measures as attributes
   - Transforming affiliation data to one-mode network
   - Replicating these procedures in UCINET
2. Displaying and analyzing real-world data in R



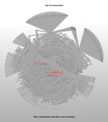
Purpose and Outline

At the end of the session(s) the participants are expected to understand and replicate the procedures followed to analyze and visualize affiliation networks using UCINET and R.

The specific topics to be covered are:

1. Displaying and analyzing affiliation data structures in UCINET
   - Displaying affiliation data
   - Using centrality measures as attributes
   - Transforming affiliation data to one-mode network
   - Replicating these procedures in UCINET

2. Displaying and analyzing real-world data in R



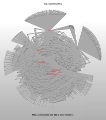
Affiliation data structures

Two-mode networks?

- Also known as affiliation, two-mode, or bipartite networks.
- Sets of relations connecting actors and events.
- **Actors**: Students, Faculty members, international organizations...
- **Events**: Associations, social movements, firms, alliances, communities...

Table 1: Two-mode data array

ID	Ev. 1	Ev. 2	Ev. 3
Bob	0	0	1
Carol	0	1	1
Ted	1	1	0
Alice	1	0	0



Representing affiliation networks

Table 1 can be represented as:

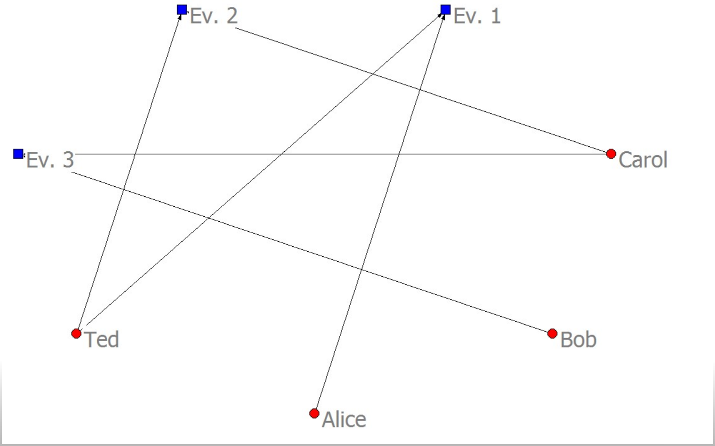


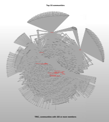
Figure 1: Representation of Table 1

Affiliation networks also have attributes

- The social network perspective focuses on the relations between actors, more than on the attributes of actors.
- Yet, SNA often takes measures of these relationships to use them as attributes.
- Both, rows and columns will have measures of centrality based on their positions in the network.

Table 2: Attribute and centrality data

ID	Sex	Age	EiVa		ID	EiVa
Bob	M	42	0.271		Ev. 1	0.500
Carol	F	44	0.653		Ev. 2	0.707
Ted	M	39	0.653		Ev. 3	0.500
Alice	F	27	0.271		—	—



Should we draw the map corresponding to Table 2?

Mapping Table 2 weighted by eigenvector centrality

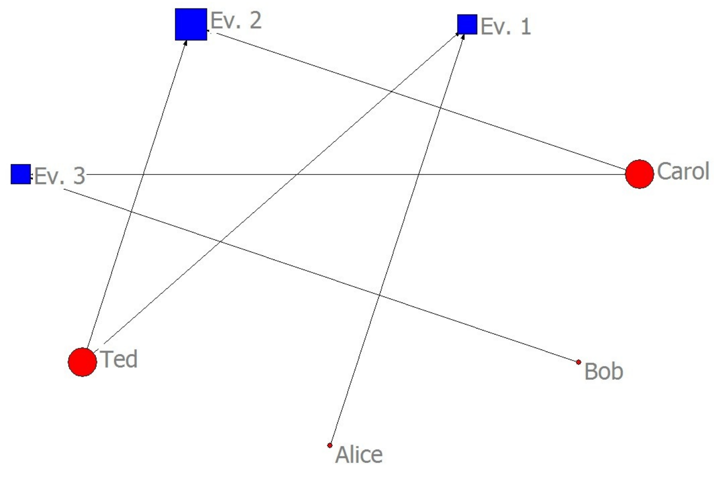


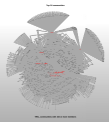
Figure 2: Why is Ev. 2 bigger than Evs. 1 and 3

Let's look at Table 1 again

Table 1: Two-mode data array

ID	**Ev. 1**	**Ev. 2**	**Ev. 3**
Bob	0	0	1
Carol	0	1	1
Ted	1	1	0
Alice	1	0	0

- Can you see what participants are attending the same events?
- It seems that Bob and Carol attend Ev. 3 ...
- Who else attended the same events?
- Can you do this with hundreds of cases?
- Let's automatize the process ...

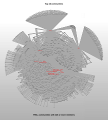


Let's look at Table 1 again

Table 1: Two-mode data array

ID	**Ev. 1**	**Ev. 2**	**Ev. 3**
Bob	0	0	1
Carol	0	1	1
Ted	1	1	0
Alice	1	0	0

- Can you see what participants are attending the same events?
- It seems that Bob and Carol attend Ev. 3 . . .
- Who else attended the same events?
- Can you do this with hundreds of cases?
- Let's automatize the process . . .

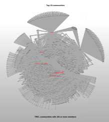


Let's look at Table 1 again

Table 1: Two-mode data array

ID	Ev. 1	Ev. 2	Ev. 3
Bob	0	0	1
Carol	0	1	1
Ted	1	1	0
Alice	1	0	0

- Can you see what participants are attending the same events?
- It seems that Bob and Carol attend Ev. 3 . . .
- Who else attended the same events?
- Can you do this with hundreds of cases?
- Let's automatize the process . . .

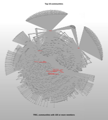


Let's look at Table 1 again

Table 1: Two-mode data array

ID	**Ev. 1**	**Ev. 2**	**Ev. 3**
Bob	0	0	1
Carol	0	1	1
Ted	1	1	0
Alice	1	0	0

- Can you see what participants are attending the same events?
- It seems that Bob and Carol attend Ev. 3 ...
- Who else attended the same events?
- Can you do this with hundreds of cases?
- Let's automatize the process ...

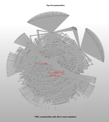


Let's look at Table 1 again

Table 1: Two-mode data array

ID	**Ev. 1**	**Ev. 2**	**Ev. 3**
Bob	0	0	1
Carol	0	1	1
Ted	1	1	0
Alice	1	0	0

- Can you see what participants are attending the same events?
- It seems that Bob and Carol attend Ev. 3 ...
- Who else attended the same events?
- Can you do this with hundreds of cases?
- Let's automatize the process ...



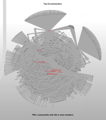
Notation and Matrix algebra

Matrix multiplication is useful to:

- Obtain actor's co-memberships or co-attendance at the same events.
- Event-event connections via overlap or interlocks with shared actors.
- These two dual networks can be measured by either pre- or post-multiplying an affiliation network and its transpose.

$A * A^T =$ Actor's co-memberships[1]

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



[1] $A^T * A =$ Event's overlaps

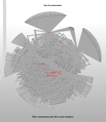
Notation and Matrix algebra

Matrix multiplication is useful to:

- Obtain actor's co-memberships or co-attendance at the same events.
- Event-event connections via overlap or interlocks with shared actors.
- These two dual networks can be measured by either pre- or post-multiplying an affiliation network and its transpose.

$$A * A^T = \text{Actor's co-memberships}$$ [1]

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



[1] $A^T * A$ = Event's overlaps

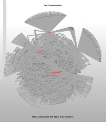
Notation and Matrix algebra

Matrix multiplication is useful to:

- Obtain actor's co-memberships or co-attendance at the same events.
- Event-event connections via overlap or interlocks with shared actors.
- These two dual networks can be measured by either pre- or post-multiplying an affiliation network and its transpose.

$A * A^T =$ Actor's co-memberships[1]

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



[1] $A^T * A =$ Event's overlaps

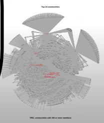
Notation and Matrix algebra

Matrix multiplication is useful to:

- Obtain actor's co-memberships or co-attendance at the same events.
- Event-event connections via overlap or interlocks with shared actors.
- These two dual networks can be measured by either pre- or post-multiplying an affiliation network and its transpose.

$$A * A^T = \text{Actor's co-memberships}^{1}$$

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$



[1] $A^T * A$ = Event's overlaps

Mapping the Actor's co-memberships

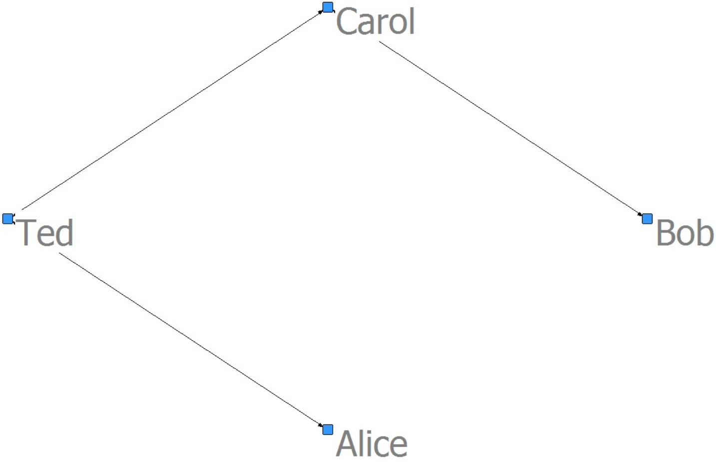
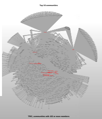


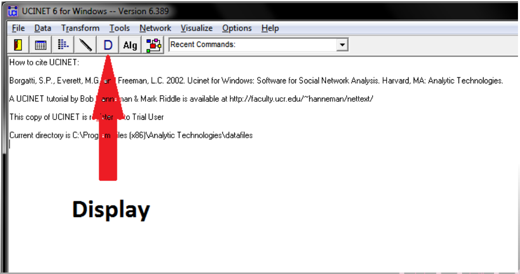
Figure 3: Let's replicate it in UCINET



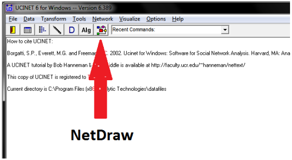
Data to be analyzed

Davis-Affiliation dataset

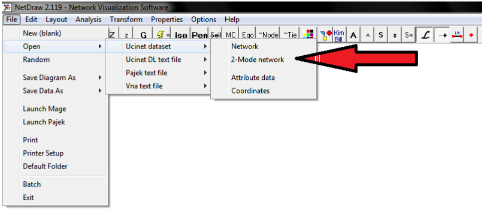
- The davis-Aff dataset comes from a study of women and the events they attended conducted by Davis in the 1930s.
- Attendance at 14 social events by 18 Southern women.
- Person-by-event matrix: cell (i,j)
- Let's view the data in UCINET, the name is **davis.##d**



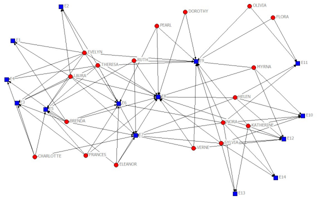
Displaying data, go back to UCINET



(a) Open NetDraw

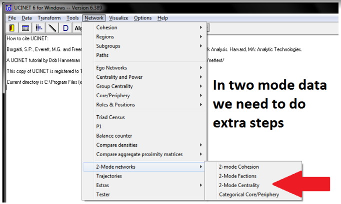


(b) In NetDraw open the dataset

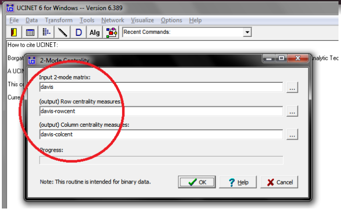


(c) Can you see what events are more important?

Again in Ucinet



(d) Open NetDraw



(e) Centrality actors and events

2-Mode Centrality Measures for ROWS of davis

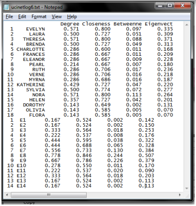
		1 Degree	2 Closeness	3 Betweenne	4 Eigenvect
1	EVELYN	0.571	0.800	0.097	0.335
2	LAURA	0.500	0.727	0.051	0.309
3	THERESA	0.571	0.800	0.088	0.371
•					
•					
•					
18	FLORA	0.143	0.585	0.005	0.070

2-Mode Centrality Measures for COLUMNS of davis

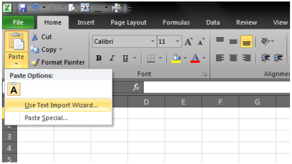
		1 Degree	2 Closeness	3 Betweenne	4 Eigenvect
1	E1	0.167	0.524	0.002	0.142
2	E2	0.167	0.524	0.002	0.150
3	E3	0.333	0.564	0.018	0.253
•					
•					
•					
14	E14	0.167	0.524	0.002	0.113

(f) Summary two mode centrality

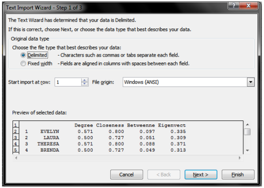
We need Excel or a similar data editor



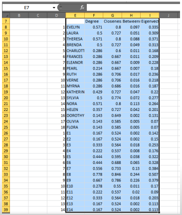
(g) Remove extra info. and copy



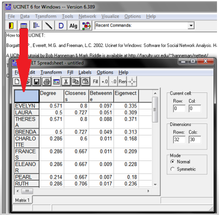
(h) Use Text Import Wizard



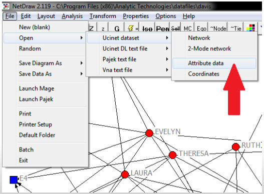
(i) If fine, just hit finish



(j) Fix header and copy

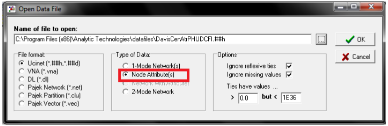


(k) Use Matrix Spread-Sheet Editor and save

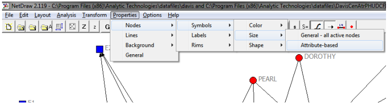


(l) In NetDraw load the attribute data just created

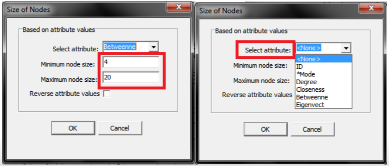
Loading and plotting attribute data



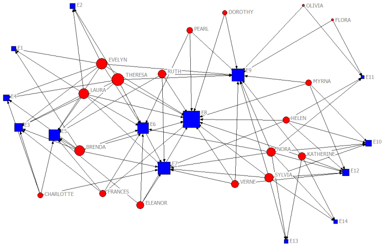
(m) Be sure you load it as attribute



(n) Modifying size of nodes

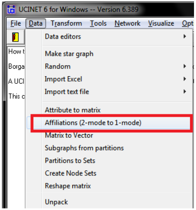


(o) Select the attribute and size

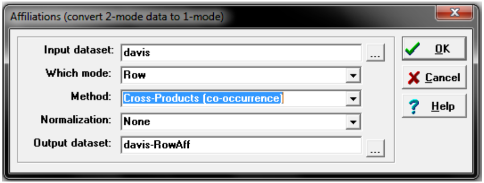


(p) Easier to see?

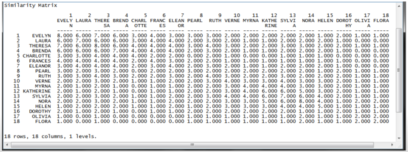
Getting actor’s co-memberships



(q) Transforming



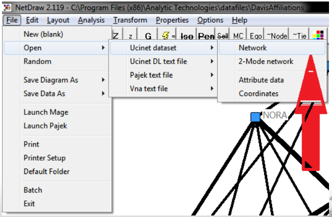
(r) Select rows or columns


Similarity Matrix

		1 EVELYN	2 LAURA	3 THERESA	4 BRENDA	5 CHARLOTTE	6 FRANCES	7 ELEANOR	8 PEARL	9 RUTH	10 VERNE	11 MYRNA	12 KATHERINE	13 SYLVIA	14 NORA	15 HELEN	16 DOROTHY	17 OLIVIA	18 FLORA
1	EVELYN	8.000	6.000	7.000	6.000	3.000	4.000	3.000	3.000	3.000	2.000	2.000	2.000	2.000	2.000	1.000	2.000	1.000	1.000
2	LAURA	6.000	7.000	6.000	6.000	3.000	4.000	4.000	2.000	3.000	2.000	1.000	1.000	2.000	2.000	2.000	1.000	0.000	0.000
3	THERESA	7.000	6.000	8.000	6.000	4.000	4.000	4.000	3.000	4.000	3.000	2.000	2.000	3.000	3.000	2.000	2.000	1.000	1.000
4	BRENDA	6.000	6.000	6.000	7.000	4.000	4.000	4.000	2.000	3.000	2.000	1.000	1.000	2.000	2.000	2.000	1.000	0.000	0.000
5	CHARLOTTE	3.000	3.000	4.000	4.000	4.000	2.000	2.000	0.000	2.000	1.000	0.000	0.000	1.000	1.000	1.000	0.000	0.000	0.000
6	FRANCES	4.000	4.000	4.000	4.000	2.000	4.000	3.000	2.000	2.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.000	0.000
7	ELEANOR	3.000	4.000	4.000	4.000	2.000	3.000	4.000	2.000	3.000	2.000	1.000	1.000	2.000	2.000	2.000	1.000	0.000	0.000
8	PEARL	3.000	2.000	3.000	2.000	0.000	2.000	2.000	3.000	2.000	2.000	2.000	2.000	2.000	2.000	1.000	2.000	1.000	1.000
9	RUTH	3.000	3.000	4.000	3.000	2.000	2.000	3.000	2.000	4.000	3.000	2.000	2.000	3.000	2.000	2.000	2.000	1.000	1.000
10	VERNE	2.000	2.000	3.000	2.000	1.000	1.000	2.000	2.000	3.000	4.000	3.000	3.000	4.000	3.000	3.000	2.000	1.000	1.000
11	MYRNA	2.000	1.000	2.000	1.000	0.000	1.000	1.000	2.000	2.000	3.000	4.000	4.000	4.000	3.000	3.000	2.000	1.000	1.000
12	KATHERINE	2.000	1.000	2.000	1.000	0.000	1.000	1.000	2.000	2.000	3.000	4.000	6.000	6.000	5.000	3.000	2.000	1.000	1.000
13	SYLVIA	2.000	2.000	3.000	2.000	1.000	1.000	2.000	2.000	3.000	4.000	4.000	6.000	7.000	6.000	4.000	2.000	1.000	1.000
14	NORA	2.000	2.000	3.000	2.000	1.000	1.000	2.000	2.000	2.000	3.000	3.000	5.000	6.000	8.000	4.000	1.000	2.000	2.000
15	HELEN	1.000	2.000	2.000	2.000	1.000	1.000	2.000	1.000	2.000	3.000	3.000	3.000	4.000	4.000	5.000	1.000	1.000	1.000
16	DOROTHY	2.000	1.000	2.000	1.000	0.000	1.000	1.000	2.000	2.000	2.000	2.000	2.000	2.000	1.000	1.000	2.000	1.000	1.000
17	OLIVIA	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	2.000	1.000	1.000	2.000	2.000
18	FLORA	1.000	0.000	1.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	2.000	1.000	1.000	2.000	2.000

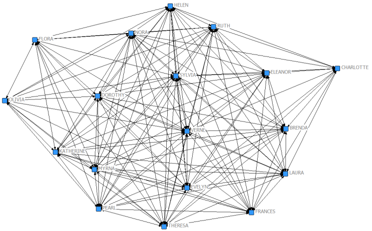
18 rows, 18 columns, 1 levels.

(s) Co-Members

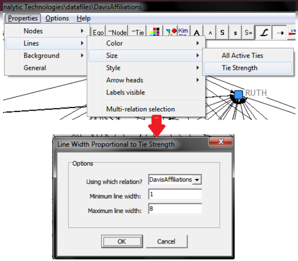


(t) NetDraw one-mode

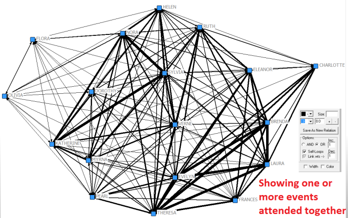
Why do I need to do this?



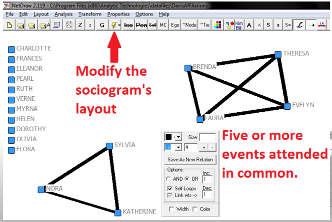
(u) Sociogram co-members



(v) Adding weight

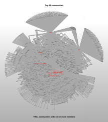


(w) Co-Members



(x) Strong co-members

Outline



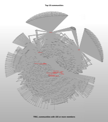
What is real-world data?

Real-world or High-dimensional data

- Term adopted to describe massive amounts of data
- Requires more computer and packages power
- Requires to adopt **data mining** techniques

  The science and art of finding hidden structure in large amounts of data, dropping waste while ensuring that valuable information is kept.

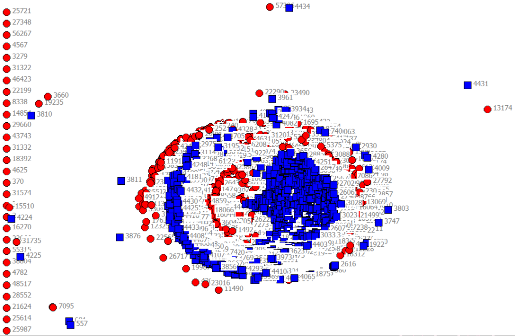
- We will analyze student's affiliations to virtual communities.
- Actors=4,064 and Communities=4,445
- Almost impossible to be handled by UCINET with regular computers.

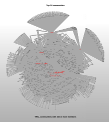


Why R?

Even though R could be difficult to learn

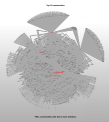
- Has THE MOST advanced techniques in many disciplines, including SNA.
- If UCINET does not crash when trying to read the data, the outcome is practically unusable.





Shortest story ever

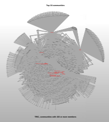
- R is an open source language and environment for statistical computing and graphics.
- A group of grad students in (multidisciplinary) statistical methodology did not want to pay for the package S and decided to build their own platform.
- R, like S, is designed around a true computer language, allowing users to continue creating functions and packages.
- R has more than 3,898 packages and counting.
- Please, bear with us, we will **be learning R by doing**.



R Basics

We will be using the following commands:

```
Everything after # is a comment and will be ignored
<- is used to assign values to objects
Objects can be data frames, matrices, arrays, vectors, scalars
ls() #Tells what objects are in the current workspace
setwd(‘‘path to your folder’’) #Tells R where to read/save data
dta<-read.csv("Data.csv", header=TRUE, sep=",")#Reads csv files
dim(dta) #Tells us dimensions (No. of rows and columns)
colnames(dta)[1]<-"id" #Changes the name of column 1 to id
rownames(dta)<-dta$id #Adds the value of id to actors
install.packages("package name")#Connects to an R repository
library("package name")#Makes its functions available
dta<-t(dta) #Transposes object dta
```



R Time?

Let's initialize R