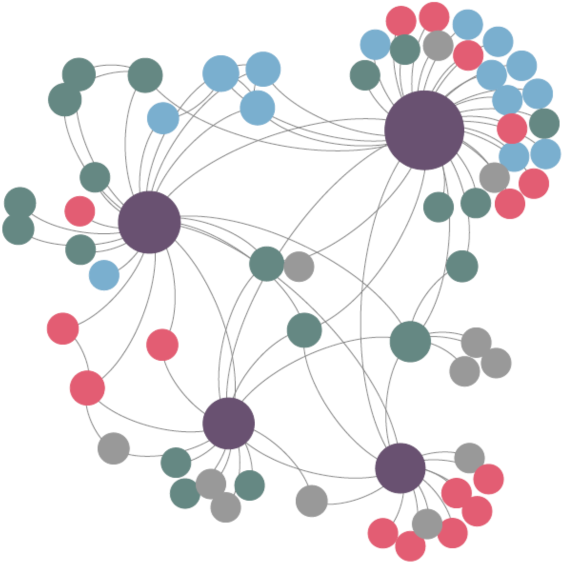
Network Analysis: An introduction

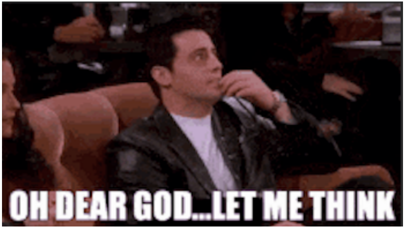
Welcome to the world of Social Network Analysis (SNA)



Objectives

- Learn basic concepts and data collection in SNA
- Use graphics to visualize and understand networks
- Gain familiarity with formatting/managing network data
- Learn to perform inferential network analysis using a variety of tools

What is SNA?



Origins and History of Network Analysis

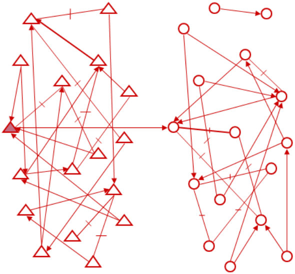
Early Puzzles

- **Sociologists** began using the term as early as 1887 and early 1990s
- Emile Durkheim, Jacob Moreno, and later Harrison with (among others) were interested in understanding social patterns and the relations between members of a system.
    - How do people feel towards one another? Why might this matter?

Early study of network analysis

Early Puzzles (1932): Individuals inside social groups

- Example: In 1932 there was a pandemic of runaways at the institution: within two weeks 14 girls ran away, which was 30 times more than the average number
- Moreno's finding: position in network predicted whether the girl would run away

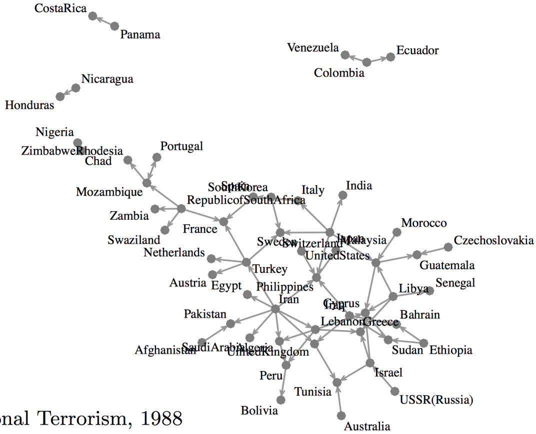


Quoting Moreno on the Analysis of Networks

"Such an invisible structure underlies society and has its influence in determining the conduct of society as a whole. Deep psychological evolutions have been evident throughout the world in the last few years in clashes between groups within nations, and between nations themselves. Until we have at least determined the nature of these fundamental structures which form the networks, we are working blindly in a hit-or-miss effort to solve problems which are caused by group attraction, repulsion and indifference."

What is a network (i.e., a graph)?

Set of **nodes** and **relation(s)** defined on them



Transnational Terrorism, 1988

Defining Network Features & Measurements: What's a node?

- A **node** can be defined as an entity that can form relations with other entities. (A node is also called a vertex).

Synonyms:

- actor: from sociometry, common terminology in sociology and psychology
- vertex: from graph theory (i.e., math), common terminology in mathematics and physics
- Term node is common in statistics and applied sciences outside of soc and psych.

Defining Network Features & Measurements: What's a node?

- A **node** can be defined as an entity that can form relations with other entities. (A node is also called a vertex).

Synonyms:

- actor: from sociometry, common terminology in sociology and psychology
- vertex: from graph theory (i.e., math), common terminology in mathematics and physics
- Term node is common in statistics and applied sciences outside of soc and psych.

Let's brainstorm a list of nodes!

Go.

Defining Network Features & Measurements: What's a relation?

- A **relation** defines the existence of an attribute relating nodes.

Synonyms:

- link: common in computer science (e.g., huge lit on "Link Prediction") and social sciences
- edge: graph theoretic terminology common in physics and math, but also elsewhere

Ties can have characteristics:

- Weight
- Qualitative attributes
- Direction

Defining Network Features & Measurements: What's a relation?

- A **relation** defines the existence of an attribute relating nodes.

Synonyms:

- link: common in computer science (e.g., huge lit on "Link Prediction") and social sciences
- edge: graph theoretic terminology common in physics and math, but also elsewhere

Ties can have characteristics:

- Weight
- Qualitative attributes
- Direction

Let's brainstorm ties to go with our nodes!

Go.

Defining Network Features & Measurements

How can we capture these relationships? *The Sociomatrix*.
example: *PONIES*

Sociomatrix: Pony threats

Directed, targeted behavior

Table 3: Threat relationships between ponies on the hill. The table shows the frequency with which each pony threatened each other individual. Animals were ranked according to the method used by Schein and Fohrman (1955). Values which are underlined indicate the two ponies which each animal threatened relatively most except in the four lowest ranking ponies where only the identity of the most threatened individual is shown

Threatener	Threatened: WT	WH	WS	GA	BR	BA	TD	WG	PM	CA	GD	DA	2 B	2 D	2 G	2 S	TA	Total threats	Number of different ponies threatened
WT		2	8	6	8	10	8	15	5	12	6	14	9	15	4	3	9	134	16
WH			6	8	6	1	2		3	7	9	4	4		2	4	2	58	13
WS				1	9	8	7		1	9	11	10	7	1	6	1	4	75	13
GA					3	1	2		2	4	3	8	5	3	1	2		34	11
BR				1		2	3	12	4	9	11	6	6	13	5	3	7	82	13
BA		6			1		2		1	3	7	5	4	0	3		5	38	10
TD					1			7	1	4	4	1	3	6	3	5	3	38	11
WG		2	3				1			8	4	3	5	3	5	4	3	41	11
PM										6	7	6	9	9	7	9	6	59	8
CA									1		8	5	10	9	9	5	8	55	8
GD										2		18	8	4	6	8	5	51	7
DA									1	2	2		5	4	8	7	4	33	8
2 B											4	4		5	6	7	10	36	6
2 D													1		1	5	4	11	4
2 G																6	2	8	2
2 S												1		1	2		4	8	4
TA															1			1	1

Sociomatrix: Pony grooming

Undirected, mutual behavior

Table 5: Grooming relationships between ponies. The table shows the frequency with which each individual groomed each other pony. Animals were ranked according to their position in the threat hierarchy. Values which are underlined indicate the two ponies which each individual groomed with relatively most (see p. 210)

Grooms with	WT	WH	WS	GA	BR	BA	TD	WG	PM	CA	GD	DA	2 B	2 D	2 G	2 S	TA	Total grooming session	Number of different ponies groomed with
WT			1			1		8			2							12	4
WH			5		33	25	2		2	4	5	2	1					79	9
WS	1	5			1	5		1			1			1				15	7
GA					11	1	3	1		2	2	3	2					25	7
BR		33	1	11		4	4	1	4	4	23	4	1	4	2	3	2	100	15
BA	1	25	5	1	4		14			12	3	4	2	1	1			73	12
TD		2		3	4	14								3				26	5
WG	8		1	1						1								11	4
PM		2			4					6	12	9	1	2	2	2		40	9
CA		4		2	4	12		1	6		8	2	1		1	1	2	44	12
GD	2	5	1	2	23	3			12	8		21	4	1	3	2		87	13
DA		2		3	4	4			9	2	21		1	3	6	6	1	62	12
2 B		1		2	1	2			1	1	4	1		2	15	3	3	36	12
2 D			1		4	1	3		2		1	3	2		4	1	4	26	11
2 G					2	1			2	1	3	6	15	4		12	7	53	10
2 S					3				2	1	2	6	3	1	12		2	32	9
TA					2					2		1	3	4	7	2		21	7

Where can we collect network data?

- Archival text data: surveillance documents (informant networks)

Where can we collect network data?

- Archival text data: surveillance documents (informant networks)
- Survey data: political behavior in Africa (close families and friends)

Where can we collect network data?

- Archival text data: surveillance documents (informant networks)
- Survey data: political behavior in Africa (close families and friends)
- Text: co-occurrence matrices (word co-occurrence $\rightarrow$ topic models)

Where can we collect network data?

- Archival text data: surveillance documents (informant networks)
- Survey data: political behavior in Africa (close families and friends)
- Text: co-occurrence matrices (word co-occurrence $\rightarrow$ topic models)
- Event data: conflict between actors, shared behavior between actors (rebel alliances or fighting networks)

Networks in Political Science

Dyads

- Introduced by the use of dyads, largely in International Relations literature

Networks in Political Science

Dyads

- Introduced by the use of dyads, largely in International Relations literature
- Early work in IR focused on the behavior and policies of individual states (for example, Morgenthau 1948).

Networks in Political Science

Dyads

- Introduced by the use of dyads, largely in International Relations literature
- Early work in IR focused on the behavior and policies of individual states (for example, Morgenthau 1948).
- Analysis of pairs of countries (trade, war, democracy, political ties).
  - Example:
    - US-Iraq 2003: War
    - US-Iran 2003: No War
    - Iran-Iraq 2003: No War

How the data look like (or why logits are not who you think they are)

Sender	Receiver	Event
i	j	y_{ij}
$\vdots$	k	y_{ik}
	l	y_{il}
j	i	y_{ji}
$\vdots$	k	y_{jk}
	l	y_{jl}
k	i	y_{ki}
$\vdots$	j	y_{kj}
	l	y_{kl}
l	i	y_{li}
$\vdots$	j	y_{lj}
	k	y_{lk}

$\longrightarrow$

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

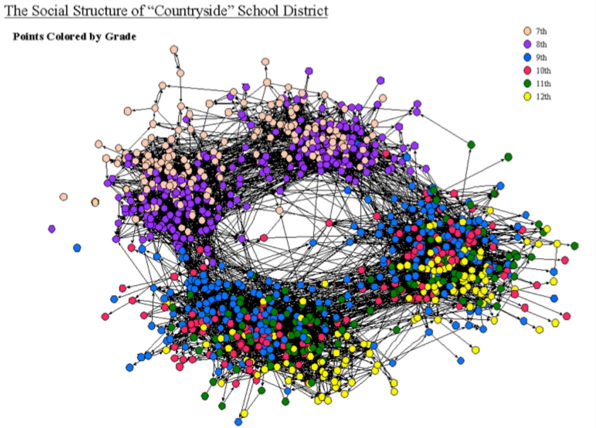
$\downarrow$

Networks in Political Science

Today: Systems (Dyads $\rightarrow$ Networks)

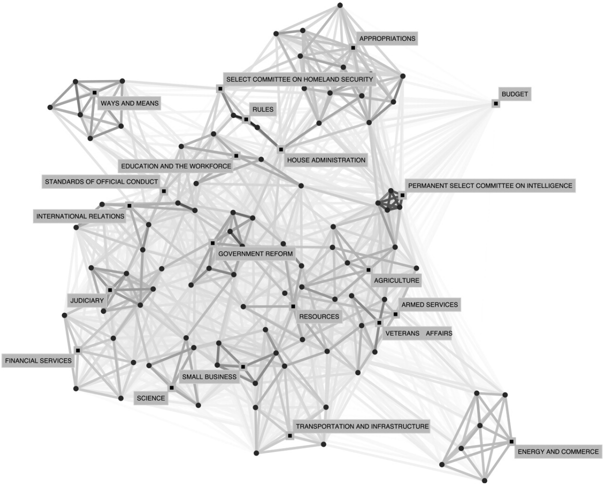
- Researchers recognize that dyads cannot be studied independently
- Network analysis is seen in a wide variety of applications both within and beyond Political Science:
  - geography
  - spatial analysis
  - conflict studies
  - peer-networks
  - congressional voting

Network graphs can reveal important structure



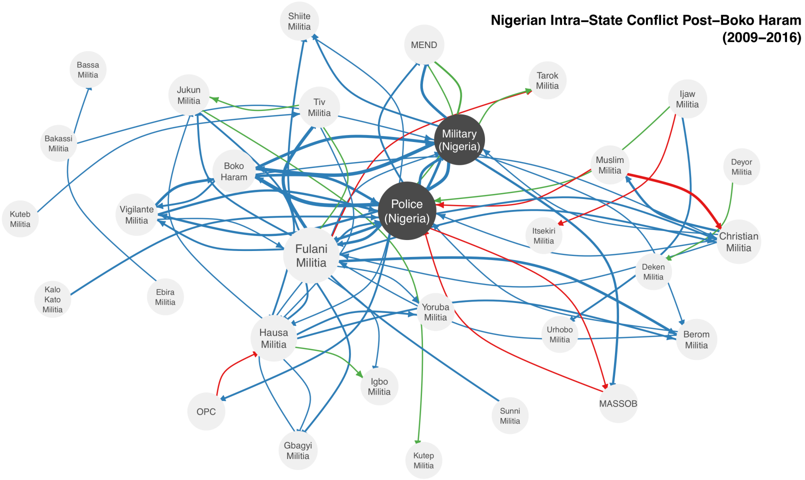
Adolescent Social Structure by **Jim Moody**

Committees and Subcommittees



A network analysis of committees in the US House of Representatives by **Porter et. al 2005**

Networks of Violence in Nigeria



Networks of Violence: Predicting Conflict in Nigeria by **Dorff, Gallop, & Minhas**

Dealing with Data

You might begin with either a matrix or information stored separately about edges and nodes. This depends on your data collection strategy.

Useful terminology for working in R:

- Matrices: the adjacency matrix
- Edges: linkages between actors or nodes
- Vertices: nodes (or actors) in your system

How can we go beyond dyads?

How do we restructure a dyadic data frame, such as alliances from COW, into a matrix format?

```
load('defAlly.rda')
head(defAlly)
```

```
  ccode1 ccode2    ij defAlly year
1      2     20  2_20       1 2012
2      2     31  2_31       0 2012
3      2     41  2_41       0 2012
4      2     42  2_42       0 2012
5      2     51  2_51       0 2012
6      2     52  2_52       0 2012
```

```
#str(defAlly)
```

Transforming dyadic data

(Hint the `spread` function from the `tidyr` package might be useful … also easy to do with a for loop or the `cast` function from the `reshape2` package)

Transforming dyadic data

This is how we go beyond dyads

```
defAllyMat[1:20,1:20]

##     2 20 31 41 42 51 52 53 54 55 56 57 58 60 70 80 90 91 92 93
## 2  NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 20  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 31  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 41  0  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 42  0  0  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 51  0  0  0  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0  0
## 52  0  0  0  0  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0  0
## 53  0  0  0  0  0  0  0 NA  0  0  0  0  0  0  0  0  0  0  0  0
## 54  0  0  0  0  0  0  0  0 NA  0  0  0  0  0  0  0  0  0  0  0
## 55  0  0  0  0  0  0  0  0  0 NA  0  0  0  0  0  0  0  0  0  0
## 56  0  0  0  0  0  0  0  0  0  0 NA  0  0  0  0  0  0  0  0  0
## 57  0  0  0  0  0  0  0  0  0  0  0 NA  0  0  0  0  0  0  0  0
## 58  0  0  0  0  0  0  0  0  0  0  0  0 NA  0  0  0  0  0  0  0
## 60  0  0  0  0  0  0  0  0  0  0  0  0  0 NA  0  0  0  0  0  0
## 70  0  0  0  0  0  0  0  0  0  0  0  0  0  0 NA  0  0  0  0  0
## 80  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 NA  0  0  0  0
## 90  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 NA  0  0  0
## 91  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 NA  0  0
## 92  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 NA  0
## 93  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 NA
```

```r
cntries = unique(defAlly$ccode1)
defAllyMat = matrix(0,
  nrow=length(cntries), ncol=length(cntries),
  dimnames=list(cntries,cntries)
  )
diag(defAllyMat) = NA

defMat = defAlly %>%
  dplyr::filter(year==2012) %>%
  dplyr::select(ccode1, ccode2, defAlly) %>%
  spread(ccode2, defAlly) %>%
  data.matrix()

rownames(defMat) = defMat[,1]
defMat = defMat[,-1]
defMat[is.na(defMat)] = 0
diag(defMat) = NA
```

What does moving towards a sociomatrix buy us?

Sender heterogeneity

An actor can induce dependence across its "recievers." Thus values across a row, say $\{y_{ij}, y_{ik}, y_{il}\}$, may be more similar to each other than other values in the adjacency matrix because each of these values has a common sender i.

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Sender heterogeneity

An actor can induce dependence across its "recievers." Thus values across a row, say $\{y_{ij}, y_{ik}, y_{il}\}$, may be more similar to each other than other values in the adjacency matrix because each of these values has a common sender i.

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Receiver heterogeneity

Additionally, values across a column, say $\{y_{ji}, y_{ki}, y_{li}\}$, may be more similar to each other than other values in the adjacency matrix because each of these values has a common receiver i.

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Sender-Receiver Covariance

Actors who are more likely to send ties in a network may also be more likely to receive them.

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Reciprocity

Values of y_{ij} and y_{ji} may be statistically dependent. Dyads might exhibit high reciprocity because there is a tendency for actors to treat each other similarily, i.e., "respond in kind" to these behaviors.

	i	j	k	l
i	NA	y_{ij}	y_{ik}	y_{il}
j	y_{ji}	NA	y_{jk}	y_{jl}
k	y_{ki}	y_{kj}	NA	y_{kl}
l	y_{li}	y_{lj}	y_{lk}	NA

Let's explore some data

```r
library(amen) # Load additive and multiplicative effects pkg
data(IR90s) # Load trade data
```

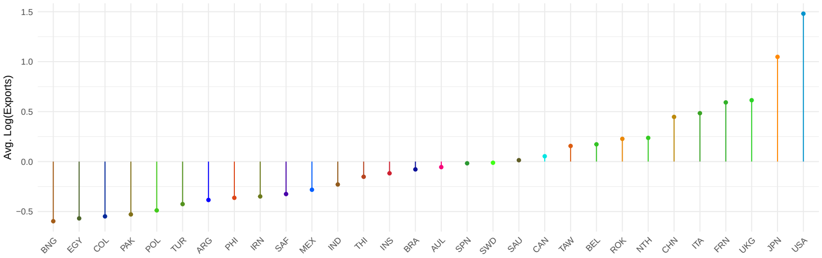
```r
Y[1:5,1:5] # Data organized in an adjacency matrix
```

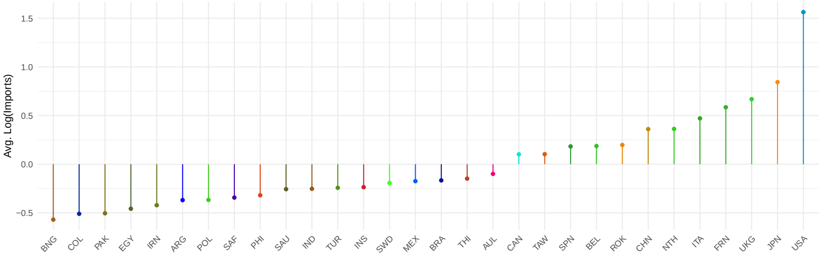
```
          ARG        AUL       BEL        BNG        BRA
ARG        NA 0.05826891 0.2468601 0.03922071 1.76473080
AUL 0.0861777         NA 0.3784364 0.10436002 0.21511138
BEL 0.2700271 0.35065687        NA 0.01980263 0.39877612
BNG 0.0000000 0.01980263 0.1222176         NA 0.01980263
BRA 1.6937791 0.23901690 0.6205765 0.03922071         NA
```

Parsing the hairball - Nodal Effects

```
effPlot = function(dimID, ylabel, colors=ccols, net=Y){
  globalMean = mean(Y, na.rm = TRUE)
  avgActivity = apply(net, dimID, mean, na.rm=TRUE)
  avgActivity_demean = avgActivity - globalMean
  muDf = data.frame(mu=avgActivity_demean); muDf$id = rownames(muDf)
  muDf$id = factor(muDf$id, levels=muDf$id[order(muDf$mu)])
  muDf$ymax = with(muDf, ifelse(mu>=0,mu,0))
  muDf$ymin = with(muDf, ifelse(mu<0,mu,0))
  gg = ggplot(muDf, aes(x=id, y=mu, color=id)) +
    geom_point() + geom_linerange(aes(ymax=ymax,ymin=ymin)) +
    xlab('') + ylab(ylabel) +
    scale_color_manual(values=colors) +
    theme(
      legend.position='none', axis.ticks=element_blank(),
      axis.text.x=element_text(angle=45, hjust=1),
      panel.border=element_blank() )
  return(gg) }
grid.arrange(
  effPlot(1, 'Avg. Log(Exports)'),
  effPlot(2, 'Avg. Log(Imports)'), nrow=2)
```

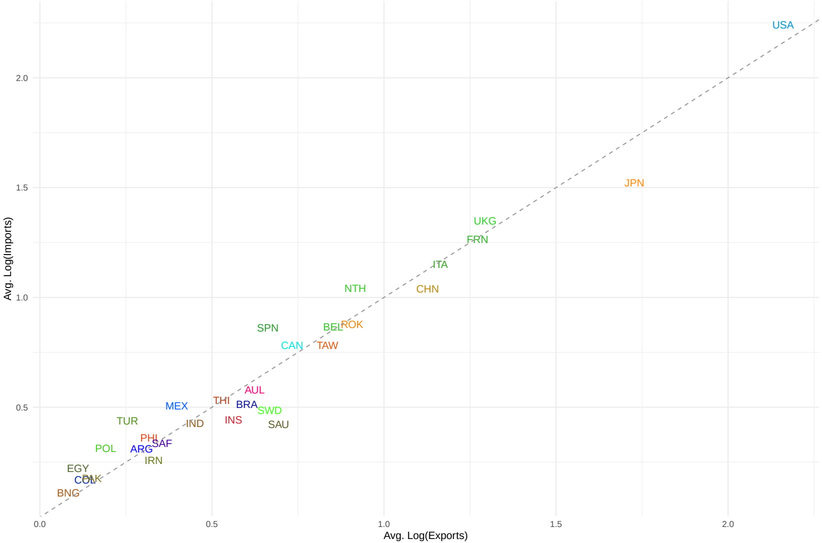
Parsing the hairball - Nodal Effects





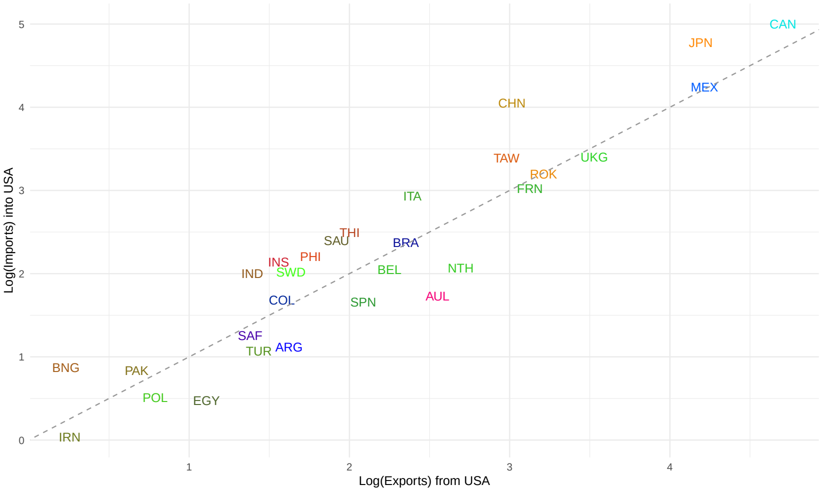
Parsing the hairball - Covariance

```
senColEff = data.frame(
  exp=apply(Y, 1, mean, na.rm=TRUE),
  imp=apply(Y, 2, mean, na.rm=TRUE))
senColEff$id = rownames(senColEff)

ggplot(senColEff, aes(x=exp, y=imp,label=id, color=id)) +
  geom_text() +
  geom_abline(
    slope=1, intercept=0, linetype='dashed', color='grey60') +
  ylab('Avg. Log(Imports)') + xlab('Avg. Log(Exports)') +
  scale_color_manual(values=ccols) +
  theme(
    legend.position = 'none',
    axis.ticks=element_blank(),
    panel.border=element_blank())
```


USA
JPN
UKG
FRN
ITA
NTH
CHN
SPN
BEL
ROK
CAN
TAW
AUL
THI
BRA
MEX
SWD
INS
SAU
TUR
IND
PHI
SAF
POL
ARG
IRN
EGY
COL
PAK
BNG
Avg. Log(Imports)
Avg. Log(Exports)
0.0
0.5
1.0
1.5
2.0

Anything else going on … reciprocity?

```
usa = na.omit(data.frame(send=Y['USA',], rec=Y[,'USA']))
usa$id = rownames(usa)
dat = data.frame(
  x=c(min(usa$send),max(usa$send)), ymin=min(usa$rec),
  y=c(max(usa$rec),min(usa$rec)), ymax=max(usa$rec))

ggplot(data=usa, aes(x=send, y=rec, label=id, color=id)) +
  geom_text() +
  geom_abline(
    slope=1, intercept=0, linetype='dashed', color='grey60') +
  ylab('Log(Imports) into USA') + xlab('Log(Exports) from USA') +
  scale_color_manual(values=ccols) +
  theme(
    legend.position = 'none',
    axis.ticks=element_blank(),
    panel.border=element_blank())
```


CAN
JPN
MEX
CHN
TAW
UKG
ROK
FRN
ITA
THI
SAU
BRA
PHI
INS
IND
SWD
BEL
NTH
COL
AUL
SPN
SAF
ARG
TUR
BNG
PAK
POL
EGY
IRN
0
1
2
3
4
5
1
2
3
4
Log(Imports) into USA
Log(Exports) from USA