Introduction to Social Network Analysis: Broad Overview and a Demonstration of a Novice's Guide Using R Statnet

June 28, 2018

The webinar will begin at 12:00 PM ET.

Please listen through the audio on your computer.

NACCHO
National Association of County & City Health Officials

Logistics

- The second portion of this webinar will involve an explanation and live demonstration of a recently developed toolkit from the Florida Department of Health in Orange County.
- This tool can be accessed by clicking the hyperlink, copying and pasting the web address below, or searching the NACCHO toolbox
  - A Novice's Guide: Social Network Analysis Using R Statnet
  - http://toolbox.naccho.org/pages/tool-view.html?id=5731

Logistics

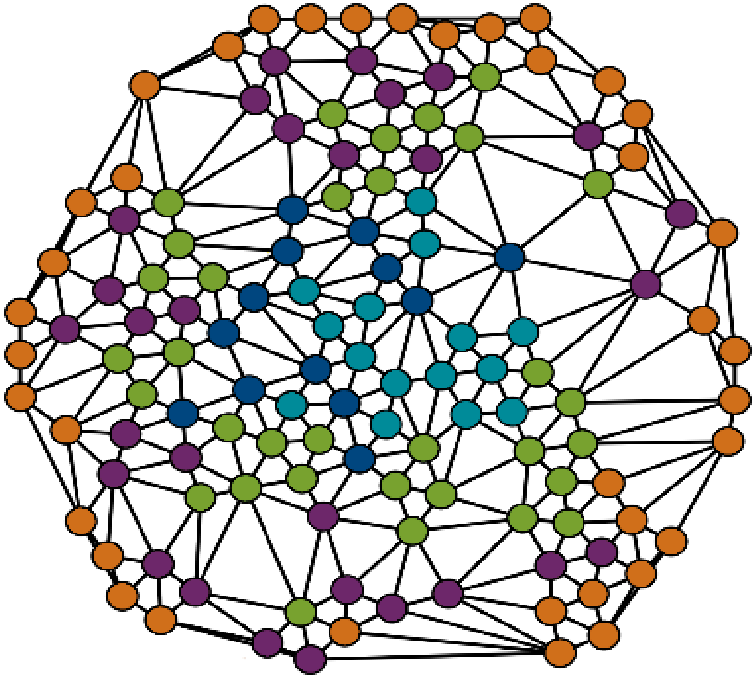
- Please listen through the audio on your computer
- This webinar is being recorded and the recording will be shared
- Submit questions through the Q&A Box at any time. We will discuss questions at the end of the presentation
- If you need technical assistance, please use the Q&A box or email infectiousdiseases@naccho.org

Speaker Introductions

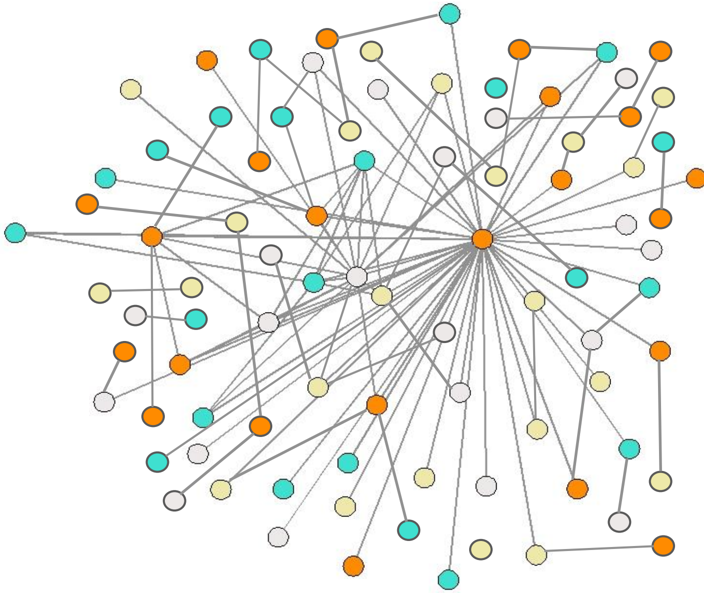
NACCHO
National Association of County & City Health Officials

Danielle Rankin, MPH, CIC

Infection Prevention and Assessment Response Epidemiologist,
Florida Department of Health Bureau of Epidemiology



Introduction to Social Network Analysis Webinar: Broad Overview and a Walkthrough of a Novice's Guide Using R Statnet



Danielle A. Rankin, MPH, CIC

Infection Control Assessment & Response Epidemiologist

Bureau of Epidemiology

National Association of County & City Health Officials Webinar

June 28, 2018

Objectives

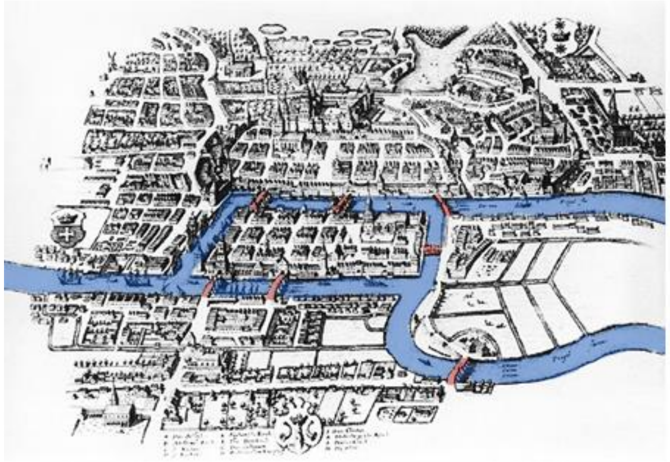
- Describe a social network analysis including general terminology and possible applications
- Locate and apply a toolkit that provides step-by-step instructions for a sample social network analysis
- Review the format of the social network analysis toolkit
- Visualize a tutorial of social network analysis on the R Statnet package (Version 2016.9)

Background

- Methodological and theoretical tools
- Transcribe the following:
  - Connections of people or partnerships
  - Disease transmission
  - Role of social support and social capital
  - Interorganizational structure of health care systems

History of Social Network Analysis

- Eighteenth Century
  - Königsberg Bridge Problem: Leohard Euler
- Foundation for social network analysis
  - Field of sociology: Auguste Comte and Georg Simmel
- Development of modern social network analysis
  - Current day sociograms: Jacob Moreno

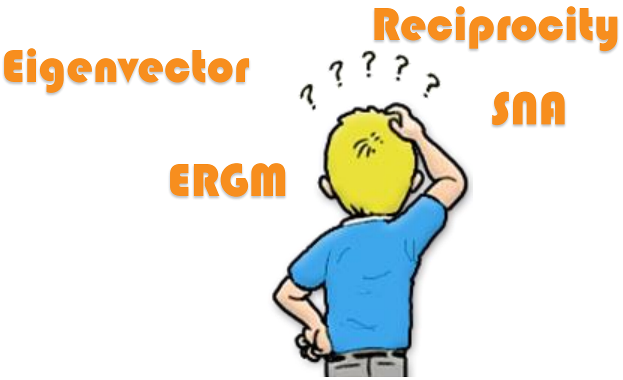


Source: www.mathscareers.org.uk/article/bridges-of-konigsberg-and-graph-theory/

Applications in Public Health

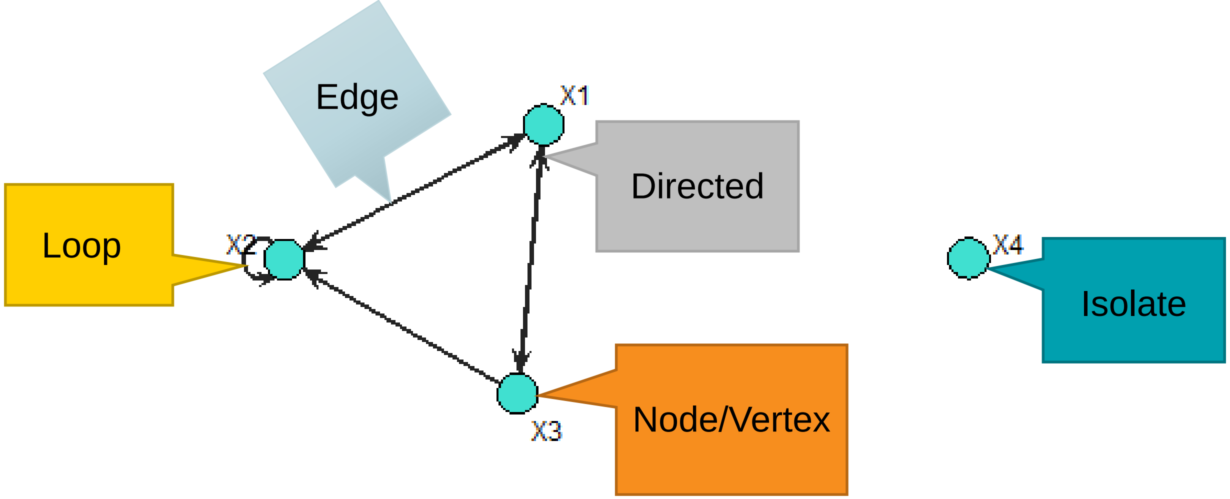
Increasing trend in exploration of social network analyses (SNAs)

- Application categories:
  1. Transmission networks
     - Disease transmission and information transmission
  2. Social networks
     - Social structure and relationships
  3. Organizational networks
     - Organizations and agencies



General Social Network Terminology

Graph/Sociogram Basics



Edge

Loop

Directed

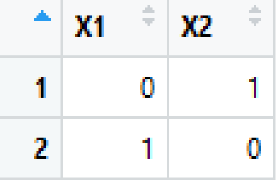
Isolate

Node/Vertex

Importation Formats

Adjacency Matrix

- Unweighted Adjacency Matrix



	X1	X2
1	0	1
2	1	0

- Weighted Adjacency Matrix

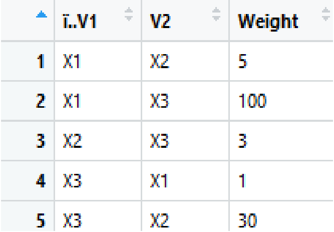
	X1	X2
1	0	5
2	17	0

Edge List

- Unweighted Edge List

	ï..V1	V2
1	X1	X2
2	X1	X3
3	X2	X3
4	X3	X1
5	X3	X2

- Weighted Edge List



	ï..V1	V2	Weight
1	X1	X2	5
2	X1	X3	100
3	X2	X3	3
4	X3	X1	1
5	X3	X2	30

Network Level Centrality Measure

Term and Definition	Interpretation
Density: the ratio of observable edges to the potential edges in a network 1. Potential Edges: $\frac{e * (e-1)}{2}$ 2. Network Density: $\frac{Observable\ Edges}{Potential\ Edges}$	Outputs range from 0 to 1. A value approaching 0 indicates a sparse network; a value approaching 1 indicates a tightly connected network. `[1] 0.4166667`

Node Level Centrality Measure

Term and Definition

Degree Centrality: the number of connections a vertex contains

- Outdegree: the number of connections a vertex exports/sends
- Indegree: the number of connections a vertex imports/receives

Interpretation

- Degree, outdegree, and indegree: outputs range from 0 to ∞
- Degree

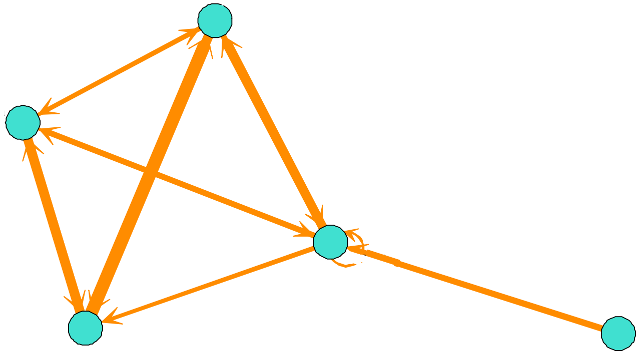
```
[1] 3 4 2 1
```

- Outdegree

```
[1] 1 2 1 1
```

- Indegree

```
[1] 2 2 1 0
```



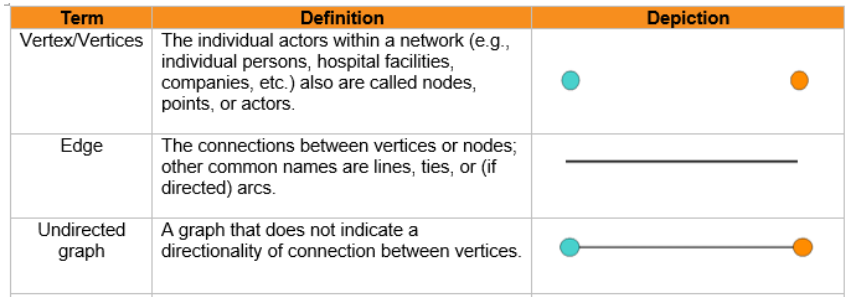
A Novice's Guide:
Social Network Analysis Using R Statnet

Structure and Format

- Overview
  - R Software
  - Downloading and installing R
  - General R coding rules
  - General R Statnet commands
- General SNA Terminology
- SNA Tutorial
  - Creating adjacency matrices and edge lists
  - Importing into R
- Social Network Centrality Measures
- Sociogram Analysis Tutorial
  - Network level calculations and interpretations
  - Node level calculations and interpretations
- Helpful Resources
  - YouTube
  - Text

General SNA Terminology

Example section in toolkit

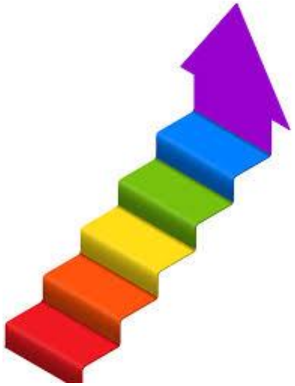


Term	Definition	Depiction
Vertex/Vertices	The individual actors within a network (e.g., individual persons, hospital facilities, companies, etc.) also are called nodes, points, or actors.	● ●
Edge	The connections between vertices or nodes; other common names are lines, ties, or (if directed) arcs.	———
Undirected graph	A graph that does not indicate a directionality of connection between vertices.	●———●

SNA Tutorial

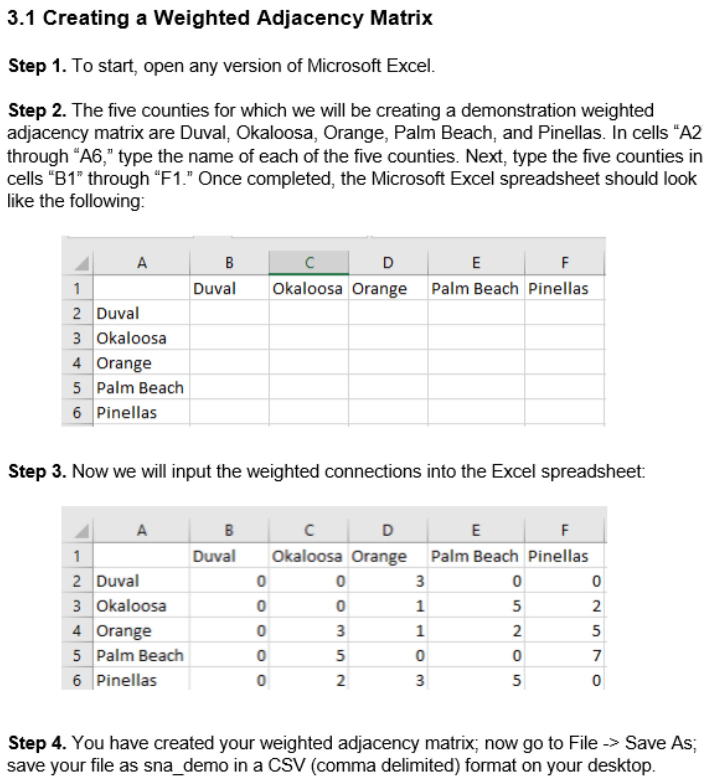
Step-by-step demonstration with a mock dataset

- Includes:
  - Creating a weighted adjacency matrix and edgelist
  - Importing and uploading with both dataset formats
  - Coding to create a sociogram with the mock dataset
  - Calculating and interpreting centrality measures of sociogram



Weighted Adjacency Matrix Tutorial

Example section in toolkit



3.1 Creating a Weighted Adjacency Matrix

Step 1. To start, open any version of Microsoft Excel.

Step 2. The five counties for which we will be creating a demonstration weighted adjacency matrix are Duval, Okaloosa, Orange, Palm Beach, and Pinellas. In cells "A2 through "A6," type the name of each of the five counties. Next, type the five counties in cells "B1" through "F1." Once completed, the Microsoft Excel spreadsheet should look like the following:

	A	B	C	D	E	F
1		Duval	Okaloosa	Orange	Palm Beach	Pinellas
2	Duval					
3	Okaloosa					
4	Orange					
5	Palm Beach					
6	Pinellas					

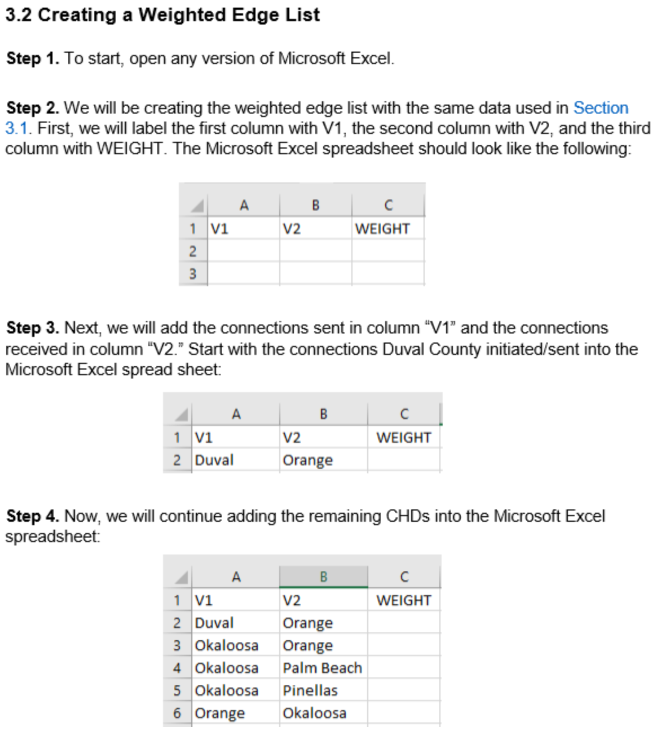
Step 3. Now we will input the weighted connections into the Excel spreadsheet:

	A	B	C	D	E	F
1		Duval	Okaloosa	Orange	Palm Beach	Pinellas
2	Duval	0	0	3	0	0
3	Okaloosa	0	0	1	5	2
4	Orange	0	3	1	2	5
5	Palm Beach	0	5	0	0	7
6	Pinellas	0	2	3	5	0

Step 4. You have created your weighted adjacency matrix; now go to File -> Save As; save your file as sna_demo in a CSV (comma delimited) format on your desktop.

Weighted Edge List Tutorial

Example section in toolkit



3.2 Creating a Weighted Edge List

Step 1. To start, open any version of Microsoft Excel.

Step 2. We will be creating the weighted edge list with the same data used in Section 3.1. First, we will label the first column with V1, the second column with V2, and the third column with WEIGHT. The Microsoft Excel spreadsheet should look like the following:

	A	B	C
1	V1	V2	WEIGHT
2			
3			

Step 3. Next, we will add the connections sent in column "V1" and the connections received in column "V2." Start with the connections Duval County initiated/sent into the Microsoft Excel spread sheet:

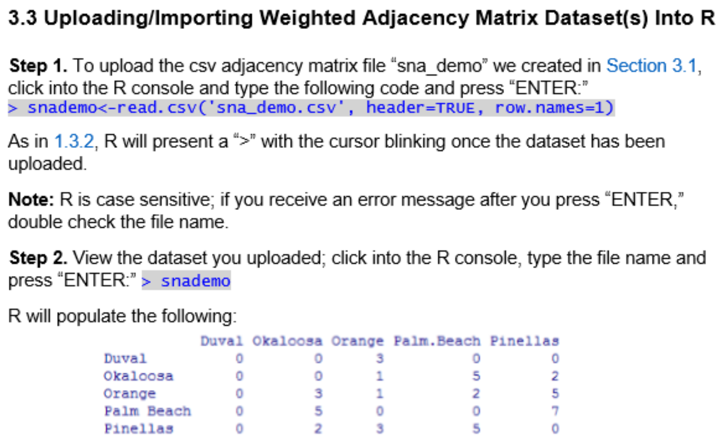
	A	B	C
1	V1	V2	WEIGHT
2	Duval	Orange	

Step 4. Now, we will continue adding the remaining CHDs into the Microsoft Excel spreadsheet:

	A	B	C
1	V1	V2	WEIGHT
2	Duval	Orange	
3	Okaloosa	Orange	
4	Okaloosa	Palm Beach	
5	Okaloosa	Pinellas	
6	Orange	Okaloosa	

Uploading/Importing Weighted Adjacency Matrix Dataset(s)

Example sections in toolkit



3.3 Uploading/Importing Weighted Adjacency Matrix Dataset(s) Into R

Step 1. To upload the csv adjacency matrix file "sna_demo" we created in Section 3.1, click into the R console and type the following code and press "ENTER:"

```
> snademo<-read.csv('sna_demo.csv', header=TRUE, row.names=1)
```

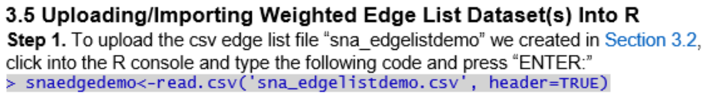
As in 1.3.2, R will present a ">" with the cursor blinking once the dataset has been uploaded.

Note: R is case sensitive; if you receive an error message after you press "ENTER," double check the file name.

Step 2. View the dataset you uploaded; click into the R console, type the file name and press "ENTER:" `> snademo`

R will populate the following:

```
            Duval Okaloosa Orange Palm.Beach Pinellas
Duval           0        0      3          0        0
Okaloosa        0        0      1          5        2
Orange          0        3      1          2        5
Palm Beach      0        5      0          0        7
Pinellas        0        2      3          5        0
```



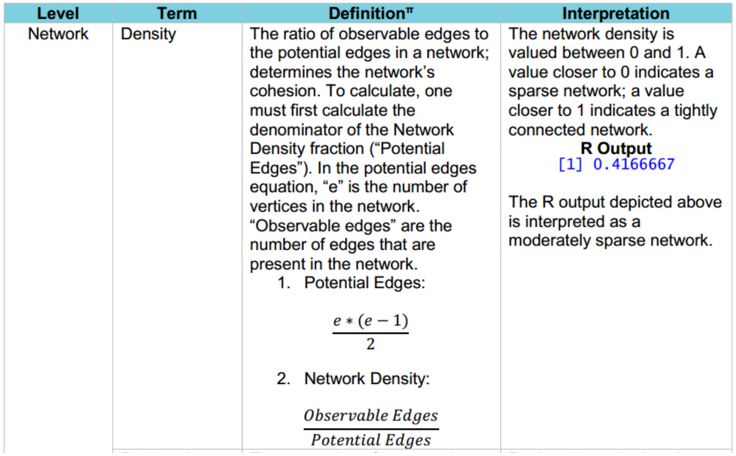
3.5 Uploading/Importing Weighted Edge List Dataset(s) Into R

Step 1. To upload the csv edge list file "sna_edgelistdemo" we created in Section 3.2, click into the R console and type the following code and press "ENTER:"

```
> snaedgedemo<-read.csv('sna_edgelistdemo.csv', header=TRUE)
```

SNA Centrality Measure Terminology

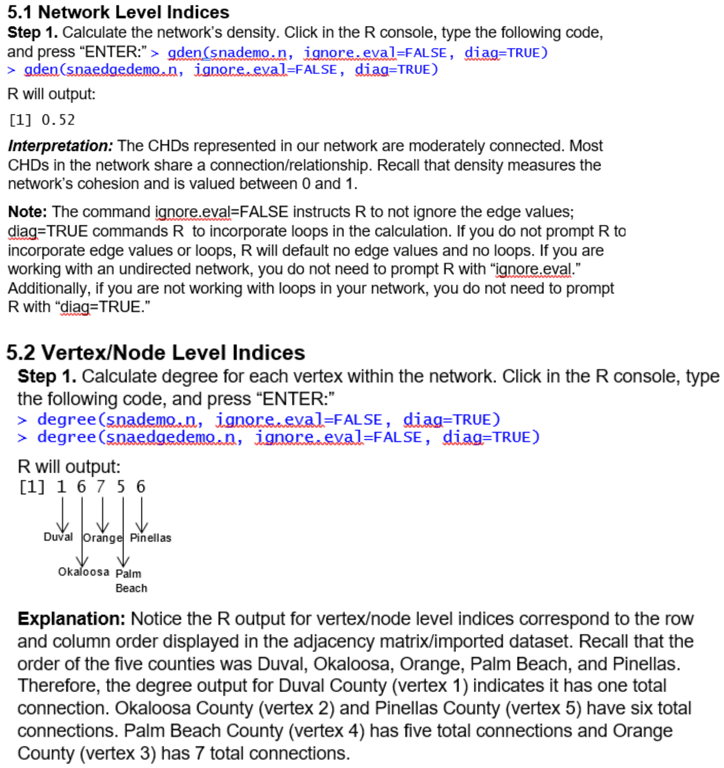
Example sections in toolkit



Level	Term	Definition$^{\pi}$	Interpretation
Network	Density	The ratio of observable edges to the potential edges in a network; determines the network's cohesion. To calculate, one must first calculate the denominator of the Network Density fraction ("Potential Edges"). In the potential edges equation, "e" is the number of vertices in the network. "Observable edges" are the number of edges that are present in the network. 1. Potential Edges: $\frac{e*(e-1)}{2}$ 2. Network Density: $\frac{Observable\ Edges}{Potential\ Edges}$	The network density is valued between 0 and 1. A value closer to 0 indicates a sparse network; a value closer to 1 indicates a tightly connected network. **R Output** `[1] 0.4166667` The R output depicted above is interpreted as a moderately sparse network.

SNA Centrality Measure Tutorial

Example sections in toolkit



5.1 Network Level Indices

Step 1. Calculate the network's density. Click in the R console, type the following code, and press "ENTER:"

```
> gden(snademo.n, ignore.eval=FALSE, diag=TRUE)
> gden(snaedgedemo.n, ignore.eval=FALSE, diag=TRUE)
```

R will output:

```
[1] 0.52
```

Interpretation: The CHDs represented in our network are moderately connected. Most CHDs in the network share a connection/relationship. Recall that density measures the network's cohesion and is valued between 0 and 1.

Note: The command ignore.eval=FALSE instructs R to not ignore the edge values; diag=TRUE commands R to incorporate loops in the calculation. If you do not prompt R to incorporate edge values or loops, R will default no edge values and no loops. If you are working with an undirected network, you do not need to prompt R with "ignore.eval." Additionally, if you are not working with loops in your network, you do not need to prompt R with "diag=TRUE."

5.2 Vertex/Node Level Indices

Step 1. Calculate degree for each vertex within the network. Click in the R console, type the following code, and press "ENTER:"

```
> degree(snademo.n, ignore.eval=FALSE, diag=TRUE)
> degree(snaedgedemo.n, ignore.eval=FALSE, diag=TRUE)
```

R will output:

```
[1] 1 6 7 5 6
```

Explanation: Notice the R output for vertex/node level indices correspond to the row and column order displayed in the adjacency matrix/imported dataset. Recall that the order of the five counties was Duval, Okaloosa, Orange, Palm Beach, and Pinellas. Therefore, the degree output for Duval County (vertex 1) indicates it has one total connection. Okaloosa County (vertex 2) and Pinellas County (vertex 5) have six total connections. Palm Beach County (vertex 4) has five total connections and Orange County (vertex 3) has 7 total connections.

Live Demonstration

Conclusions

- The toolkit was created to provide applied public health officials guidance in construction of social networks
- SNAs can be informative for applied public health officials
- SNAs can assist in formulating guided discussions, enhancing prioritization of targeted interventions, and targeting disease control

Acknowledgements

CDC Division of Healthcare Quality Promotion

Rachel Slayton, PhD, MPH

Florida Department of Health

Taylor Campion, MPH

Alvina Chu, MHS

Karen Elliott, MPH

Scott Pritchard, MPH

Michael Wydotis

National Association of County and City Health Officials (NACCHO)

References

1. Luke D, Harris J. Network Analysis in Public Health: History, Methods, and Applications. ANNUAL REVIEW OF PUBLIC HEALTH [serial online]. 2007:69. Available from: British Library Document Supply Centre Inside Serials & Conference Proceedings, Ipswich, MA. Accessed July 21, 2017.
2. R Core Team [computer program]. Version 3.3.2. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org/. Published 2016. Accessed February 2, 2018.
3. RStudio Team [computer program]. Version 1.1.383. Boston, MA: RStudio: Integrated Development for R. www.rstudio.com/. Published 2016. Accessed February 2, 2018.
4. Handcock M, Hunter D, Butts C, Goodreau S, Krivitsky P, Bender-deMoll S, Morris M. Statnet: software tools for the statistical analysis of network data. The Statnet Project. www.statnet.org. Published 2016. Accessed February 2, 2018.
5. Acton R., Butts C., Goodreau S.M. An introduction to network analysis with r and statnet. Workshop presented at: Sunbelt XXXII Workshop Series; March 13, 2012. Available from: https://statnet.org/trac/raw-attachment/wiki/Resources/introToSNAinR_sunbelt_2012_tutorial.pdf. Accessed February 2, 2018.
6. Butts C, Social network analysis with sna. Journal of Statistical Software. 2008; 24(6). Accessed February 2, 2018.
7. Petrescu-Prahova M, Belza B, Leith K, et al. Using social network analysis to assess mentorship and collaboration in a public health network. Preventing Chronic Disease. 2015:12.
8. McCulloch J. Betweenness centrality [video]. YouTube. www.youtube.com/watch?v=0CCrq62TF7U&t=638s. Published September 10, 2017. Accessed February 2, 2018.
9. McCulloch J. Centralization [video]. YouTube. www.youtube.com/watch?v=-ANEqyrJOac. Published October 25, 2017. Accessed February 2, 2018.
10. McCulloch J. Degree centrality [video]. YouTube. www.youtube.com/watch?v=iiVeQkIELyc&t=419s. Published September 10, 2017. Accessed February 2, 2018.
11. McCulloch J. Graph measures [video]. YouTube. www.youtube.com/watch?v=N3wbYrsJrbk&t=230s. Published October 25, 2017. Accessed February 2, 2018.
12. McCulloch J. Mod02D closeness [video]. YouTube. www.youtube.com/watch?v=2ELP6yd21tw. Published September 10, 2017. Accessed February 2, 2018.
13. McCuloch J. Mod02A terminology [video]. YouTube. www.youtube.com/watch?v=5Hw1OmWOLA8. Published September 10, 2017. Accessed February 2, 2018.
14. Clarkson M. A guide to the gplot function of the sna library for r. www.melissaclarkson.com/resources/R_guides/. Accessed February 2, 2018.

Questions

Danielle A. Rankin, MPH, CIC
Infection Prevention Assessment & Response Epidemiologist
Bureau of Epidemiology
Florida Department of Health
Danielle.Rankin@flhealth.gov

Discussion

Please enter your

questions or comments in to the

Q&A box

Thank you for joining today's webinar!

Contact us with questions

Email: infectiousdiseases@naccho.org

Download A Novice's Guide: Social Network Analysis Using R Statnet by visiting: http://toolbox.naccho.org/pages/tool-view.html?id=5731