INTRODUCTION TO SOCIAL NETWORK ANALYSIS AND NETWORK SCIENCE METHODS

Workshop, ICHPS 2023, Scottsdale, AZ

James O'Malley, Ph.D.

Department of Biomedical Data Science

The Dartmouth Institute for Health Policy and Clinical Practice

Geisel School of Medicine at Dartmouth

Email: James.OMalley@Dartmouth.edu

Link to Computing Materials Used in Workshop

- Github site with R scripts and data used in the illustrative examples presented in workshop:

https://github.com/kiwijomalley/ICHPS-Social-Network-Analysis-Workshop-2023

- GRANDPA algorithm
  - Use to generate random networks for analysis that emulate a base network.
  - Very useful if network data is confidential
  - Bobak CA, Zhao Y, Levy JJ, and O'Malley AJ. (2022). GRANDPA: GeneRAtive Network sampling using Degree and Property Augmentation applied to the analysis of partially confidential healthcare networks. doi.org/10.48550/arXiv.2211.15000
  - Presented in a conference poster by Carly Bobak on January 9, 2023

Example network and data: Small physician clinic in the Boston area

- Relational data (adjacency matrix form): ICHPS_PhysPractBin.txt
- Relational data (edgelist form): ICHPS_ClinEdgeList.txt
- Node attribute data: ICHPS_nodecov.dat

- References:
  - Keating, Ayanian, Cleary, and Marsden (2007)
  - O'Malley and Marsden (2008)

Segments of Code for Three Key Problems

I. Descriptive analyses involving networks
- Visualizing a network
- Descriptive features of networks

II. Statistical analyses of social influence or peer effects
- Formation of weight matrix
- Cross-sectional cases

III. Statistical analysis of relational data
- Exponential random graph models
- Latent-space models

I. Descriptive Analyses

- R script: WorkshopICHPS2023.R
  - Loads network
  - Forms directed and undirected binary-valued network objects
  - Makes plots of network
  - Computes summary measures of the network and of physicians' positions within the network

Loading Physician Practice Network

Adjacency matrix

- reldata <- scan(”ICHPS_PhysPractBin.txt")
- nr <- sqrt(length(reldir)) #Number of physicians
- reldir <- matrix(reldir,ncol=nr,nrow=nr,byrow=T)

#Directed

- reldir <- ifelse(reldir>0,1,0) #Directed
  - Can use more stringent threshold (greater # conversations) to obtain sparser network by using threshold of 1

#Undirected

- relmut <- ifelse(reldir+t(reldir)>0,1,0) #Undirected
  - Use “And Rule” (threshold of >1) to obtain sparser network
- pnet <- network(reldir,directed=TRUE,matrixtype="adjacency")

Aside: Forming an edgelist from an adjacency matrix with IDs from 1 to nr

- idto <- seq(1,nr)
- edgelist=c()
- for (i in 1:nr) {
-   edge <- idto[(reldir[i,]==1)]
-   nc=length(edge)
-   edgelist=rbind(edgelist,cbind(rep(i,nc),edge))
- }
- edgelist <- data.frame(edgelist)
- names(edgelist) <- c("source","target")
- pnet <- network(reldir,directed=TRUE,matrixtype=”edgelist")

More general loading of network data

- pnet <-
network(yourdatafile,directed=TRUE,matrixtype="edgelist
")

Use one of: adjacency, edgelist, incidence

Loading attribute data

Load attribute data

- covdata <- scan(“ICHPS_nodecov.dat")
- covdata <- matrix(covdata,nrow=nr,byrow=T)
- colnames(covdata) <- c("id","male","whexpert","pctwom","numsess","practice", "sumhrt")

Software for descriptive analysis of network

- R is convenient
  - sna package contains many built-in features
  - igraph package

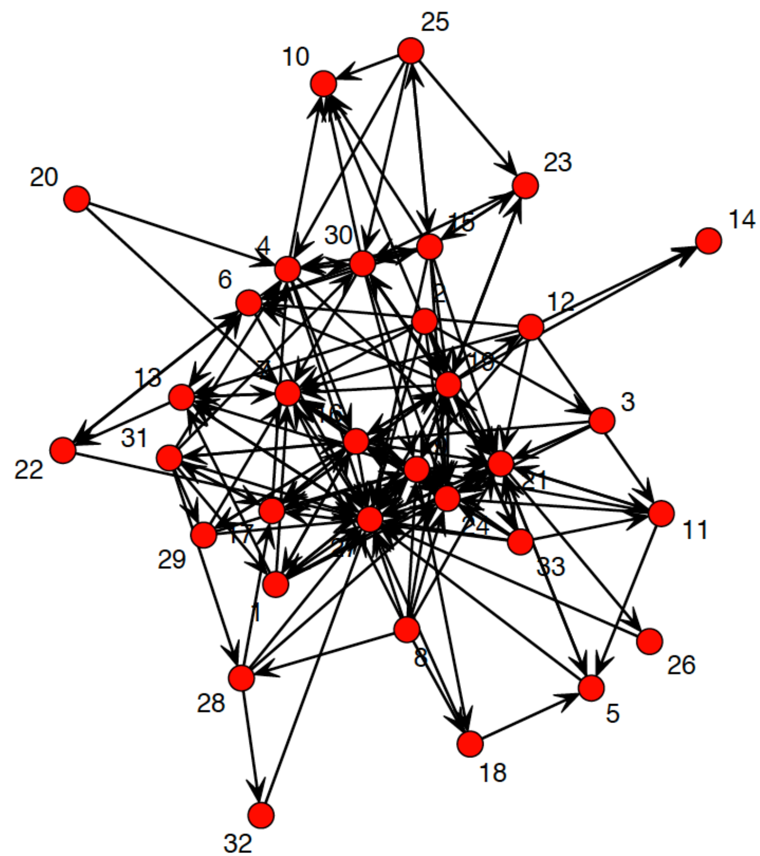
    R: install.packages(”sna”,”igraph”)

    R: library(sna)

    R: library(igraph
- Python
  - networkx package

Plotting network using SNA

- #Plot from adjacency matrix
- plot(pnet,mode="fruchtermanreingold",displaylabels=T)
  gplot(pnet,gmode="digraph",mode="fruchtermanreingold")



Using igraph

- Let edgelist be a N by 2 matrix in R with the network represented as an edgelist
- nodes <- unique(c(edgelist[,1],edgelist[,2]))

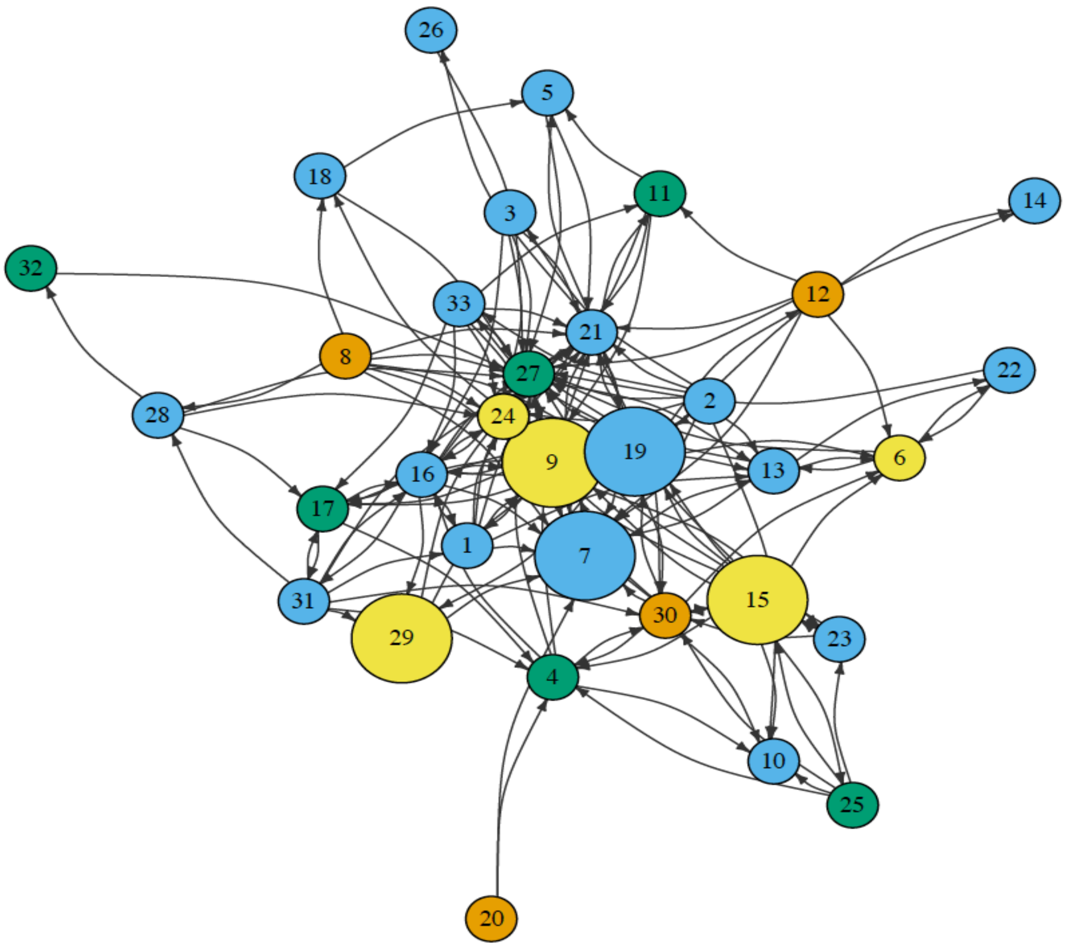
Make graph object from edgelist:

- gnet <- graph_from_data_frame(d=edgelist, vertices=nodes, directed=TRUE)
- print(gnet, e=TRUE, v=TRUE)

Plotting network in igraph

- V(gnet)$color <- covdata$practice
- V(gnet)$size <- 10*(covdata$whexpert+1)
- par(mar=c(0,0,0,0))
- plot(gnet,
-     vertex.color = V(gnet)$color, # Color of nodes
-     vertex.size = V(gnet)$size, # Size of nodes
-     vertex.label.color = "black", # change color of labels
-     vertex.label.cex = .75, # change size of labels to 75% of original size
-     edge.curved=.25, # add a 25% curve to the edges
-     edge.color="grey20", # change edge color to grey
-     edge.arrow.size=0.3)
- dev.copy2pdf(file=paste(outdir,"PhysNetNicePlot.pdf",sep=""), width=6, height=6) #to file

Visualization of network of physicians' professional relationships within a medical practice



Computing summary measures of networks in R

- Use SNA or igraph
  - You might detach igraph if want to use SNA as igraph masks functions (has other functions by the same name) or precede function name with package name (e.g., igraph::grecip)
  - detach("package:igraph", unload = TRUE)
- Dyad and Triad Census'
  - dyadc=dyad.census(reldir)
  - recip=grecip(reldir)
  - triadc_mut=triad.census(relmut,g="graph")
  - triadc_dir=triad.census(reldir,g="digraph")
  - trans=gtrans(reldir,mode="digraph")

Summary measures cont.

- Degree distributions
  - idegree=degree(reldir,cmode="indegree")
  - odegree=degree(reldir,cmode="outdegree")
  - central=centralization(reldir,degree)

- Centrality measures
  - closecent=closeness(reldir,gmode="digraph")
  - bcent=betweenness(reldir,gmode="digraph")
  - eigcent=evcent(reldir,gmode="digraph",use.eigen=FALSE)
  - powcent=bonpow(reldir,gmode="digraph")

Aside: Manual computation of eigenvector centrality

Column standardize as per Katz (1953): undirected network

- on=as.vector(rep(1,nrow(relmut)))
- rscale=as.vector(relmut %*% on)
- scnet=diag(rscale^(-1),nrow=length(cscale))
- srelmut=relmut %*% scnet #Row-stochastic adjacency mat
- seigcent=evcent(srelmut,gmode="digraph",use.eigen=FALSE)

Can't be 0!

#manual

- sesys=eigen(srelmut,symmetric=TRUE) #First eigenvalue is 1
- seigcent2=abs(sesys$vector[,1]) #Same as seigcent

Differers from:

- eigcent = evcent(relmut,gmode="digraph",use.eigen=FALSE)

R Output

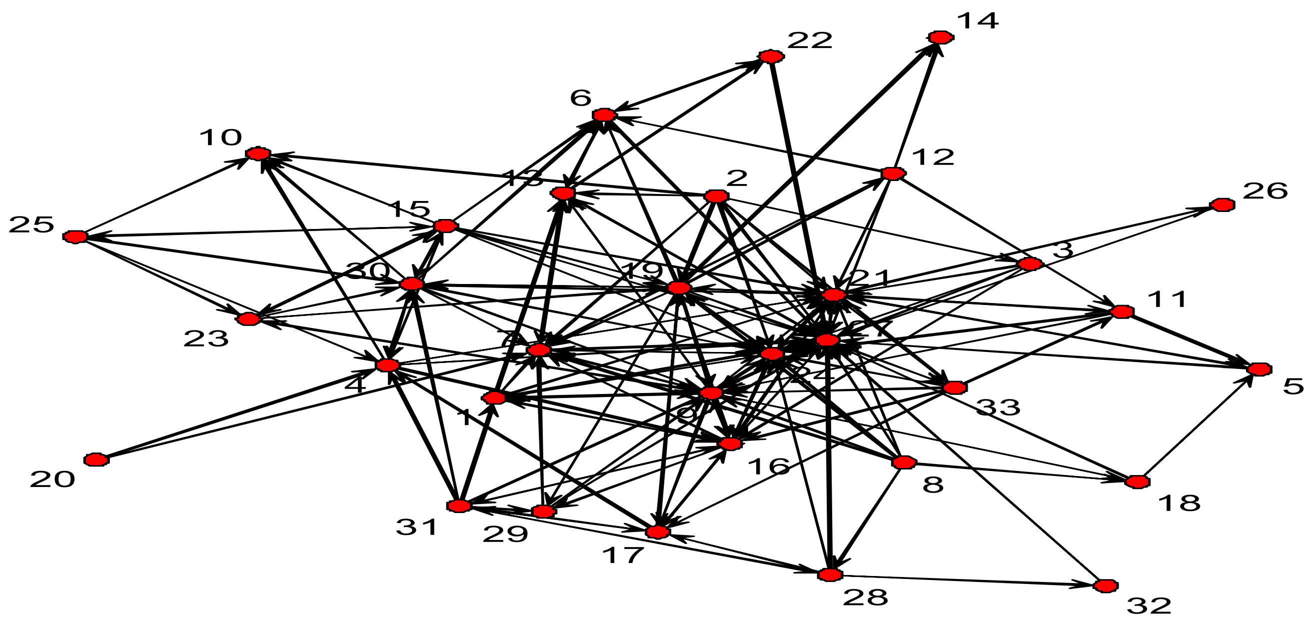
```
sesys$values
 [1]  1.00000000  0.58674579 -0.50090590  0.47303368 -0.43397276  0.42666878 -0.42491411 -0.39345582
 [9]  0.37479529 -0.37272600  0.34943660 -0.34833131  0.29465182 -0.29277386 -0.26530439 0.26150139
[17] -0.25725084 -0.25079539 -0.23908776  0.23292529  0.17182520 -0.15703102 -0.14330681 -0.12994348
[25]  0.12570645 -0.10902035 -0.10191533  0.09780081  0.06373494 -0.06033018  0.04989629 -0.04039014
[33]  0.01273312

> -sesys$vector[,1]
 [1] 0.14294834 0.16081688 0.08934271 0.17868542 0.07147417 0.12507980 0.23229105 0.14294834
 [9] 0.28589668 0.08934271 0.10721125 0.12507980 0.12507980 0.03573708 0.17868542 0.23229105
[17] 0.12507980 0.07147417 0.30376522 0.03573708 0.30376522 0.05360563 0.08934271 0.26802813
[25] 0.08934271 0.03573708 0.42884501 0.10721125 0.10721125 0.21442251 0.14294834 0.03573708
[33] 0.16081688
```

II. Statistical models involving comparative analysis of multiple networks

- Do social network characteristics correlate with other variables of interest?
- Often involves use of regression and hierarchical regression models
- Original data: Keating, Ayanian, Cleary, and Marsden (2007)
- R script: ICHPS2023models.R

Example plot of network and statistical analysis (O'Malley and Marsden, 2008)

- pnet <- network(physnetwork, directed=TRUE, matrixtype="adjacency",
  vertex.attr=nodecov,
  vertex.attrnames = c("male", "whexpert",
  "pctwom", "numsess", "practice", "bcma", "bima",
  "bpp", "wnhlth", "numcat", "pctcat"))
- plot(pnet,mode = "fruchtermanreingold“, displaylabels=T)



Simplest model

- model1a <- ergm(pnet~edges)
- Evaluating log-likelihood at the estimate.
- Formula: pnet ~ edges
- Iterations: 5 out of 20
- Monte Carlo MLE Results:
- Estimate Std. Error MCMC % p-value
- edges -1.70084 0.08517 0 <1e-04 ***
- ---
- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
- Null Deviance: 1463.9 on 1056 degrees of freedom
- Residual Deviance: 908.6 on 1055 degrees of freedom
-
- AIC: 910.6 BIC: 915.5 (Smaller is better.)

What is this the log-odds of?

Larger model cont.

- Formula: pnet ~ edges + mutual + nodeocov("whexpert") + nodeocov("pctwom") +
- nodeocov("numsess") + nodematch("male", diff = F) + nodematch("bcma",
- diff = F) + nodematch("bima", diff = F) + nodematch("bpp",
- diff = F) + nodematch("wnhlth", diff = F)

Ego (originator) covariate

Homophily covariate

- Monte Carlo MLE Results:

	Estimate	Std. Error	MCMC %	p-value	
edges	-4.560879	0.510615	0	< 1e-04	***
mutual	0.851292	0.293985	0	0.003862	**
nodeocov.whexpert	-0.391256	0.279232	0	0.161455	
nodeocov.pctwom	-0.001583	0.004530	0	0.726830	
nodeocov.numsess	0.159686	0.039360	0	< 1e-04	***
nodematch.male	0.646502	0.181881	0	0.000396	***
nodematch.bcma	0.739658	0.278889	0	0.008119	**
nodematch.bima	0.277871	0.199195	0	0.163321	
nodematch.bpp	1.396742	0.255284	0	< 1e-04	***
nodematch.wnhlth	-0.046937	0.221694	0	0.832366	

- ---
- Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

- AIC: 838.3 BIC: 887.9 (Smaller is better.)

Separate node-match (homophily) coefficients under alternative specification

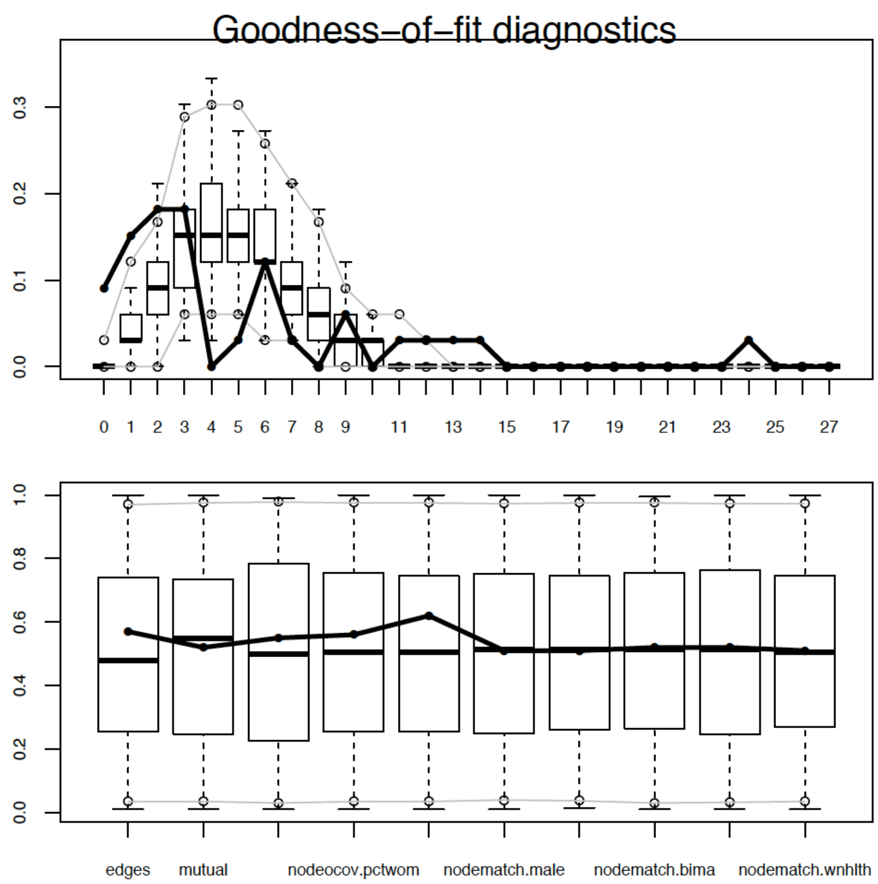
- model1f <- ergm(pnet~edges + mutual + nodeocov("whexpert") +nodeocov("pctwom") + nodeocov("numsess") + nodematch("male",diff=T) + nodematch("practice",diff=T))

Allows nodematch.male coefficient to vary by sex

	Estimate	Std. Error	MCMC %	p-value	
edges	-1.134295	0.417598	0	0.00671	**
mutual	0.615252	0.316615	0	0.05226	.
nodeocov.whexpert	-0.509118	0.293552	0	0.08315	.
nodeocov.pctwom	-0.022326	0.005604	0	< 1e-04	***
nodeocov.numsess	-0.007532	0.045438	0	0.86837	
nodematch.male.0	1.534152	0.247805	0	< 1e-04	***
nodematch.male.1	-1.112795	0.420071	0	0.00819	**
nodematch.practice.1	1.320874	0.849377	0	0.12022	
nodematch.practice.2	0.466084	0.215630	0	0.03088	*
nodematch.practice.3	2.131637	0.423777	0	< 1e-04	***
nodematch.practice.4	1.981935	0.489350	0	< 1e-04	***

Assessing Goodness of Fit with Respect to In-Degree Distribution

- model1e.gof <- gof(model1e~idegree,control=control.gof.ergm(nsim=100),verbose=T)
- plot(model1e.gof)



Estimating Latent-space models in R

- ergmm(formula, response = NULL, family = "Bernoulli", fam.par = NULL, control = control.ergmm(), user.start = list(), prior = ergmm.prior(), tofit = c("mcmc", "mkl", "mkl.mbc", "procrustes", "klswitch"), Z.ref = NULL, Z.K.ref = NULL, seed = NULL, verbose = FALSE)
  - "family" command allows different distributions in the exponential family (as for generalized linear models)
  - "prior" command gives some control over prior distributions for Bayesian analysis
  - "tofit" controls which elements of estimation are performed
- ?ergmm to get help and then terms.ergmm to get list of terms that are supported
  - Note that there are no mutual, triadic or higher-order network statistics are allowed in this model!

Equating Latentnet's ERGMM and Statnet's ERGM

- **ERGMM Formula: pnet ~ nodematch("male", diff = F)**
- **Attribute: edges (included by default in ERGMM)**
- **Model: Bernoulli**
- **MCMC sample of size 4000, draws are 10 iterations apart, after burnin of 10000 iterations.**
- **Covariate coefficients posterior means:**
- **Estimate 2.5% 97.5% 2*min(Pr(>0),Pr(<0))**
- **(Intercept) -2.15650 -2.44573 -1.8870 < 2.2e-16 ***
- **nodematch.male 0.78345 0.43672 1.1398 < 2.2e-16 ***

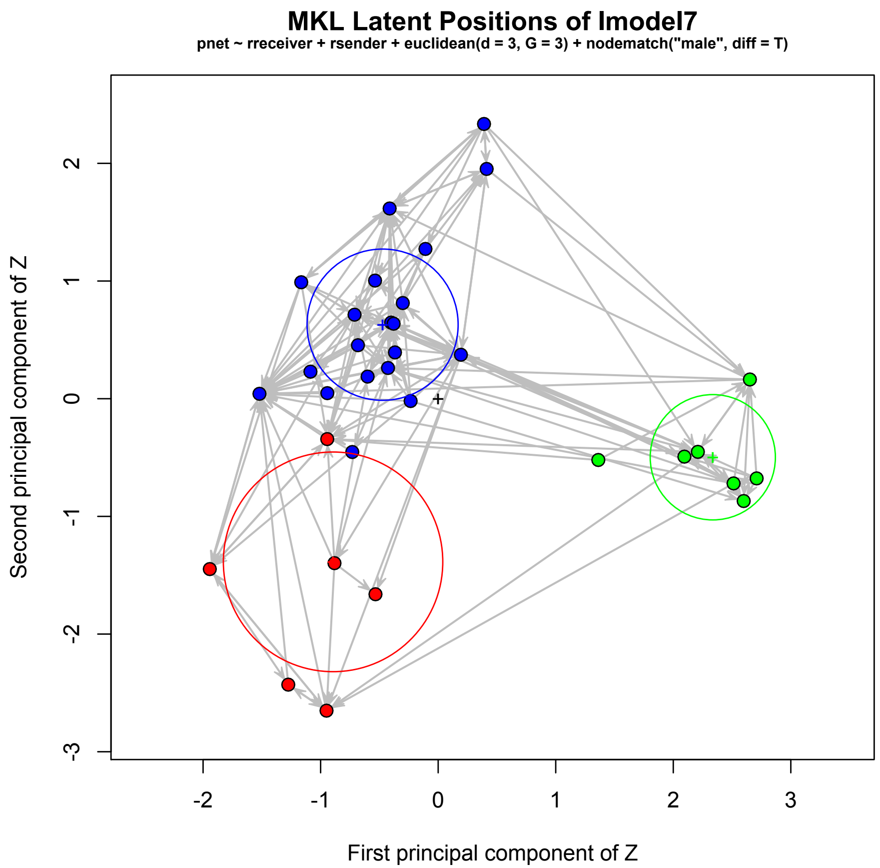
- **ERGM Formula: pnet ~ edges + nodematch("male", diff = F)**
-
- **Iterations: 5 out of 20**
-
- **Monte Carlo MLE Results:**
- **Estimate Std. Error MCMC % p-value**
- **edges -2.1552 0.1437 0 <1e-04 ***
- **nodematch.male 0.7898 0.1794 0 <1e-04 ***

WARNING: Currently, the ERGMM procedure in R can only estimate models in which edges are conditionally independent. So cannot include network statistics involving mutual or triadic terms!

Euclidean distance with $\mathrm{d} = 3$ dimensions and 3 clusters (or groups)

lmodel7 <- ergmm(pnet ~ rreceiver+rsender+euclidean(d=3,G=3)+nodematch("male",diff=T))

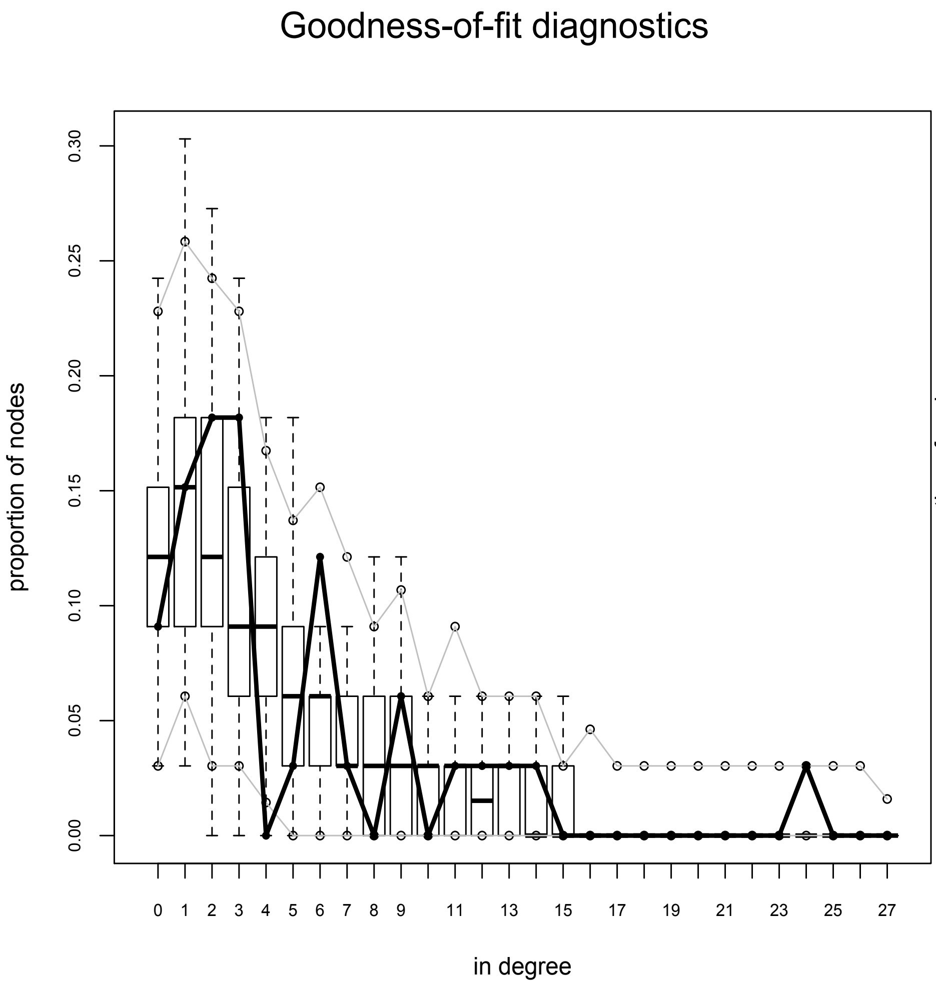
Obvious application is to clusters actors into groups!

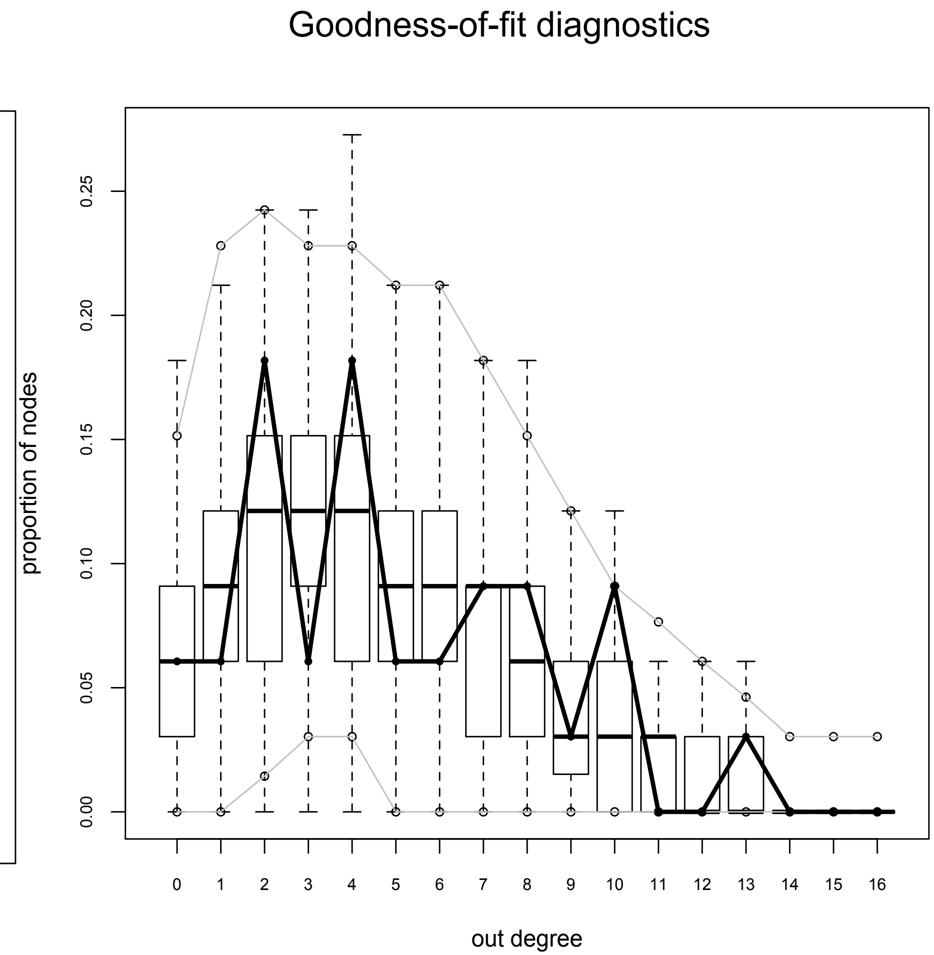


Assessment of Goodness of fit

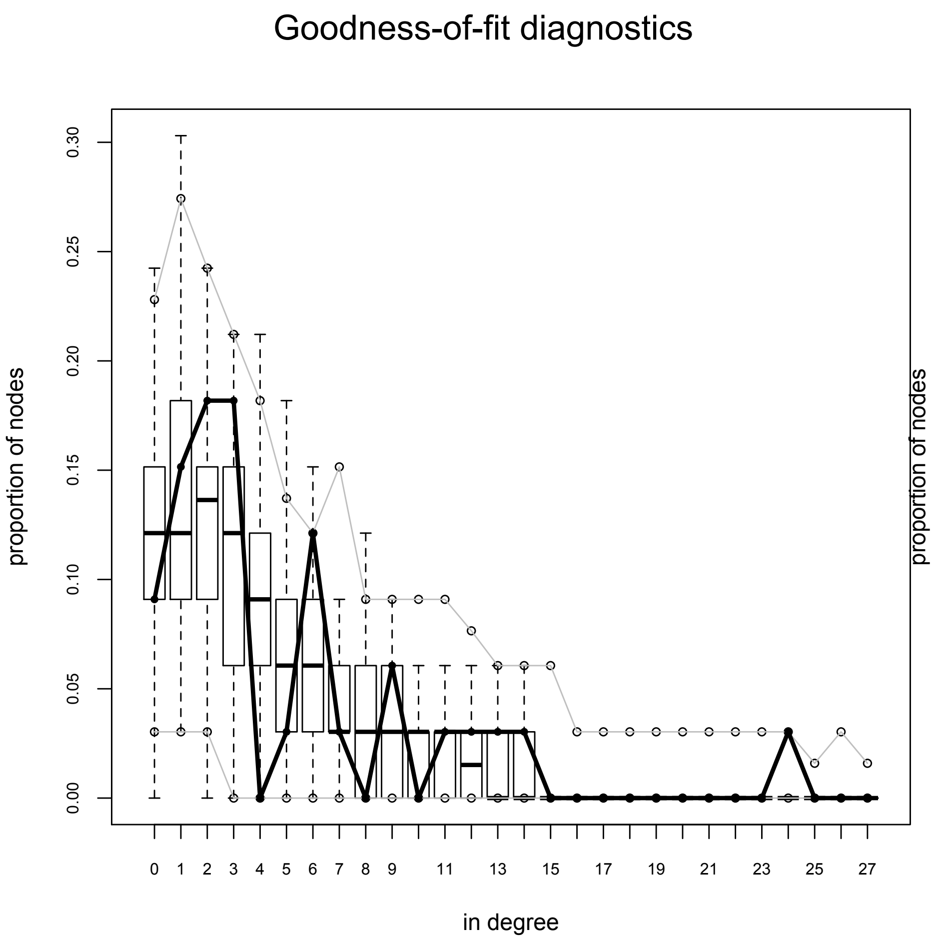
- lmodel7i.gof <- gof(lmodel7,GOF=~idegree,control=ergmm.control(nsim=100),verbose=T)
- plot(lmodel7i.gof)
- lmodel7o.gof <- gof(lmodel7,GOF=~odegree,control=ergmm.control(nsim=100),verbose=T)
- plot(lmodel7o.gof)

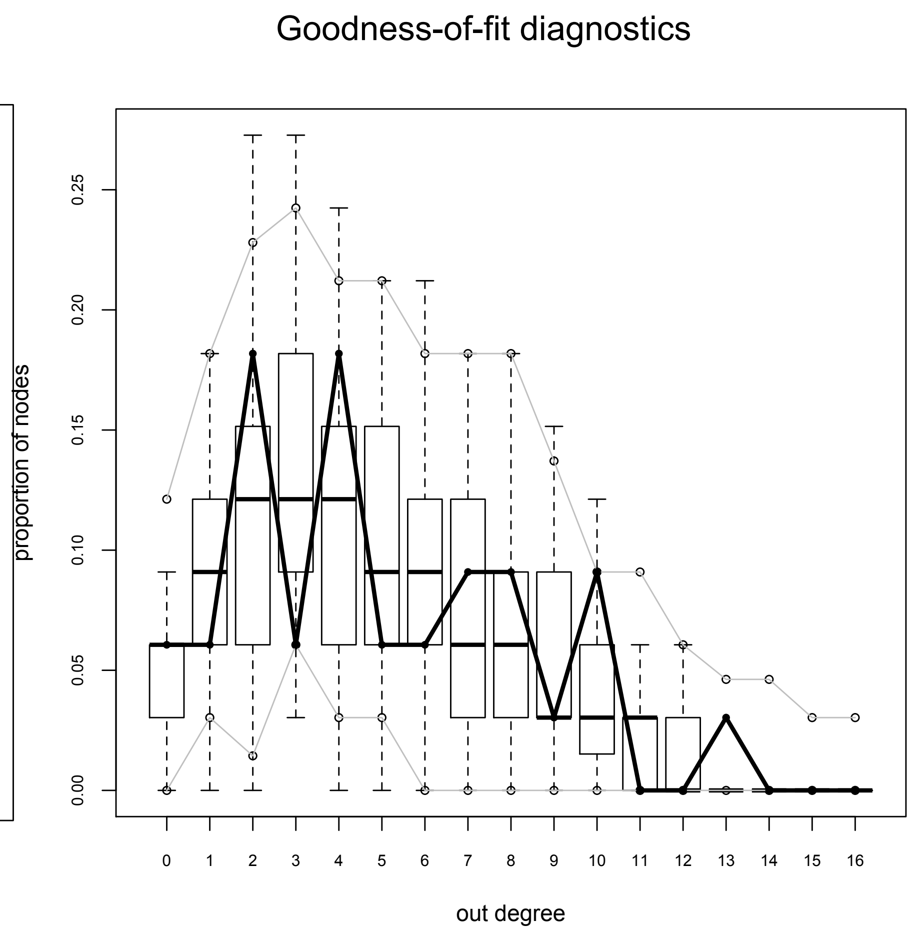
Goodness of fit of degree distributions with Euclidean latent-space





Goodness of fit of degree distributions with Bilinear latent-space





III. Statistical analyses of social influence or peer effects

- Use lnam function
  - Uses numerical approximation to second derivatives in Newton-family optimization routines
- Develop your own estimation routines
  - Use optim function for maximum likelihood inference
- Amenable to Bayesian inference

Data wrangling required to form peer actor outcome or attribute weighted average outcome

Data manipulation

- on <- as.vector(rep(1,nr))
- x <- as.matrix(cbind(on,regdata[,c("male", "pctwom","numalters")]))

Peer outcome predictor

- hrtalt <- wtreldir %*% as.vector(covdata$sumhrt)
- regdata <- data.frame(covdata,hrtalt=hrtalt,noalters=noalters,numalters=numalters)

Estimation of linear regression and autoregressive outcome models for cross-sectional network influence data in R using lnam function

Autoregressive outcome model estimation in R

- reg.adj <- lm(sumhrt~x+hrtalt-1, data=regdata)
- lnam1.adj <- lnam(regdata$sumhrt,x,wtreldir)

Network autocorrelation model estimation in R

- lnam2.adj <- lnam(regdata$sumhrt,x,NULL,wtreldir)

Results and Peer-effect Estimate Interpretation (isoequal=0; assume isolates not influenced)

	Linear regression			Autoregressive outcome model		
Term	Estimate	SE	P-value	Estimate	SE	P-value
on	16.049	4.236	0.001	16.078	3.902	0.000
male	-0.327	2.492	0.897	-0.206	2.277	0.928
pctwom	-0.054	0.052	0.305	-0.052	0.047	0.275
numalters	-0.363	0.218	0.107	-0.353	0.199	0.077
hrtalt or rho1.1	0.299	0.195	0.136	0.277	0.171	0.106

Peer effect parameter under the autoregressive outcome model estimated by Inam

Estimation of network autocorrelation model in R and Interpretation of Peer-Effect Estimates

lnam2.adj <- lnam(regdata$sumhrt,x,NULL,wtreldir)

	Autoregressive outcome model			Network autocorrelation model		
Term	Estimate	SE	P-value	Estimate	SE	P-value
on	16.078	3.902	0.000	16.471	4.089	0.000
male	-0.206	2.277	0.928	1.324	2.159	0.540
pctwom	-0.052	0.047	0.275	-0.023	0.050	0.645
numalters	-0.353	0.199	0.077	-0.209	0.227	0.356
rho1.1 or rho2.1	0.277	0.171	0.106	0.026	0.445	0.953

lnam also accommodates models with both autoregressive outcome and network autocorrelation terms

Much larger SE

References: Descriptive network measures

- Wasserman S. and Faust K. (1994), Social Network Analysis. Cambridge: Cambridge University Press
- Bonacich, P. (1987), Power and Centrality: A Family of Measures," American Journal of Sociology, 92, 1170-1182
- Freeman, L. (1979), Centrality in Social Networks, Conceptual Clarification, Social Networks, 1, 215-239
- Faust, K. (1997), Centrality in Affiliation Networks, Social Networks, 19, 157-191
- O'Malley, A. J. and Marsden, P. V. (2008), The Analysis of Social Networks, Health Services & Outcomes Research Methodology, 8, 222-269

References: Bipartite Networks and Comparison of Multiple Networks

- Borgatti, S. and Everett, M. (1997), Network Analysis of 2-Mode Data," Social Networks, 19, 243-269
- Barnett, M. L., Christakis, N. A., O'Malley, A. J., Onnela, J.-P., Keating, N. L., and Landon, B. E. (2012), Physician Patient-Sharing Networks and the Cost and Intensity of Care in US Hospitals," Medical Care, 50, 152-160
- Landon, B. E., Keating, N. L., Barnett, M. L., Onnela, J. P., Paul, S., O'Malley, A. J., Keegan, T., and Christakis, N. A. (2012), Variation in Patient-Sharing Networks of Physicians Across the United States, Journal of the American Medical Association, 308, 265-273.

References: Bipartite Networks and Comparison of Multiple Networks cont.

- An C, O’Malley AJ, Rockmore DN, Stock CD. Analysis of the U.S. Patient Referral Network. Statistics in Medicine, 2018, 37, (5), 847-866. doi: 10.1002/sim.7565. PMID: 29205445
- Moen EL, Bynum JPW, Austin AM, Chakraborti G, Skinner JS, O’Malley AJ. Assessing variation in implantable cardioverter defibrillator therapy guideline adherence with physician and hospital patient-sharing networks. Medical Care, 2018, 56, (4), 350-357
- Moen EL, Austin AM, Bynum JP, Skinner JS, O’Malley AJ. An analysis of patient-sharing physician networks and implantable cardioverter defibrillator therapy. Health Services and Outcomes Research Methodology, 2016, 16, 132-153
- Landon BE, Keating NL, Onnela J-P, Zaslavsky AM, Christakis NAC, O’Malley AJ. Patient-Sharing Networks of Physicians and Healthcare Utilization and Spending Among Medicare Beneficiaries. JAMA Internal Medicine, 2018, 178 (1), 66-73
- Keating NL, O'Malley AJ, Onnela J-P, Landon BE. Assessing the impact of colonoscopy complications on use of colonoscopy among primary care physicians and other connected physicians: an observational study of older Americans. BMJ Open, 7, (6), e014239. doi:10.1136/bmjopen-2016-014239
- Moen EL, Bynum JPW, Skinner JS, O’Malley AJ. Physician network position and patient outcomes following implantable cardioverter defibrillator therapy. Health Services Research, 2019, 54 (4), 880-889. PMID: 30937894

References: General Network Analyses and Applications

- Barabasi, A.-L. and Albert, R. (1999), Emergence of Scaling in Random Networks," Science, 286, 509-512
- Keating, N. L., Ayanian, J. Z., Cleary, P. D., and Marsden, P. V. (2007), Factors affecting influential discussions among physicians: a social network analysis of a primary care practice," Journal of general internal medicine, 22, 794-798
- Coleman, J., Katz, E., and Menzel, H. (1957), The diffusion of innovations among physicians," Sociometry, 20, 253-270
- Coleman, J., Katz, E., and et al. (1966), Medical Innovation: A Diffusion Study, Bobbs-Merrill
- Pollack, C. E., Soulos, P. R. & Gross, C. P. Physician's peer exposure and the adoption of a new cancer treatment modality. *Cancer* **121,** 2799–2807 (2015).
- Hidalgo, C. A., Blumm, N., Barabasi, A.-L., and Christakis, N. A. (2009), A Dynamic Network Approach for the Study of Human Phenotypes, PLoS Computational Biology, 5, e1000353. doi:10.1371/journal.pcbi.1000353

References: cross-sectional models of relationships (social selection)

- Erdos, P. and Renyi, A. (1959), Random Graphs, Publicationes Mathematicae, 6, 290:297.
- Wang, W. and Wong, G. (1987), Stochastic Blockmodels for Directed Graphs, Journal of the American Statistical Association, 82, 8-19.
- Choi D, Wolfe P. and Airoldi E. (2010) Stochastic blockmodels with growing number of classes. arXiv:1011.4644.
- Fienberg, S. and S. Wasserman. (1981). Categorical data analysis of single sociometric relations. In Sociological Methodology, edited by S. Leinhardt, pp. 156-92. San Francisco: Jossey-Bass.
- Frank, O. and D. Strauss. (1986). Markov graphs. Journal of the American Statistical Association 81: 832-42.
- Goldenberg, A., Zheng, A. X., Fineberg, S. E. and Airoldi, E. M. (2009). A survey of statistical network models. In press: Foundations and Trends in Machine Learning.
- Handcock, M. S., Robins, G. L., Snijders, T. A. B., Moody, J. and Besag, J. (2003). Assessing degeneracy in statistical models of social networks. Journal of the American Statistical Association, 76: 33-50.
- Holland, P. and S. Leinhardt. (1981). An exponential family of probability-distributions for directed-graphs. Journal of the American Statistical Association 76(373): 33-50.

References: cross-sectional models of relationships cont.

- Handcock, M. S., Hunter, D. R., Butts, C. T., Goodreau, S. M., Krivitsky, P. N., and Morris, M. (2010), ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks, http://CRAN.R-project.org/package=ergm. Version 2.2-6. Project home page at http://statnetproject.org
- Handcock, M. S., Robins, G. L., Snijders, T. A. B., Moody, J., and Besag, J. (2003), Assessing degeneracy in statistical models of social networks, Journal of American Statistical Association, 76, 33-50
- Robins, G. L., Snijders, T. A. B.,Wang, P., Handcock, M. S., and Pattison, P. E. (2007), Recent developments in exponential random graph (p∗) models for social networks, Social Networks, 29, 192-215
- Hoff, P. (2005). Bilinear mixed-effects models for dyadic data. Journal of the American Statistical Association 100: 286-95.
- Hoff, P., A. Raftery, and M. Handcock. (2002). Latent space approaches to social network analysis. Journal of the American Statistical Association 97: 1090-98.
- Hoff P. (2008), Modeling Homophily and Stochastic Equivalence in Symmetric Relational Data, in: Advances in Neural Information Processing Systems, volume 20, MIT Press, 657-664
- Paul S, Keating NL, Landon BE, O’Malley AJ. Results from using a new dyadic-dependence model to analyze sociocentric physician networks. Social Science & Medicine, 125, 2015, 51-59

References: dynamic network models of relationships

- Krivitsky P.N. and Handcock M.S. (2010). A Separable Model for Dynamic Networks. arXiv:1011.1937v1[stat.ME].
- Nowicki K and Snijders T.A.B. (2001). Estimation and prediction for stochastic blockstructures. Journal of the American Statistical Association, 96, 1077-1087.
- O'Malley AJ, Christakis NA. (2011). Longitudinal Analysis of Large Social Networks: estimating the Effect of Health Traits on changes in Friendship Ties. Statistics in Medicine, 30, 9, 950-964
- Paul S. and O'Malley A.J. (2013). Hierarchical longitudinal models of relationships in social networks. Journal of the Royal Statistical Society, Series C.
- Pattison, P. and S. Wasserman. (1999). Logit models and logistic regressions for social networks: II. Multivariate relations. British Journal of Mathematical and Statistical Psychology 52 (Pt 2): 169-93.
- Van Duijn, M., T. Snijders, and B. Zijlstra. (2004). A Random Effects Model with Covariates for Directed Graphs. Statistica Neerlandica 58(2): 234-54.

References: dynamic network models of relationships cont.

- Snijders, T. A. B. (2006). "Statistical methods for network dynamics." In S. R. Luchini et al., editors, Proceedings of the XLIII Scientic Meeting, Italian Statistical Society, pages 281-296, Padova: CLEUP.
- Snijders, T.A.B. (2005), Models for longitudinal social network data," in Models and Methods in Social Network Analysis, Cambridge University Press, 215-247.
- Steglich, C.E.G., Snijders, T.A.B. and Pearson, M. (2010). Dynamic Networks and Behavior: Separating Selection from Influence. Sociological Methodology, 40, 329-393.
- Westveld, A. H. and Hoff, P. D. (2011), A Mixed Effect Model for Longitudinal Relational and Network Data, With Applications To International Trade and Conflict," The Annals of Applied Statistics, 5, 843-872.

References: social influence analyses

- Christakis NA, Fowler JH. (2007). The Spread of Obesity in a Large Social Network over 32 Years. New England Journal of Medicine, 357, 370--379
- Christakis NA, Fowler, JH. (2008). Dynamics of Smoking Behavior in a Large Social Network. New England Journal of Medicine, 358, 2249--2258
- Fowler JH, Christakis NA. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. British Medical Journal, 337, doi:10.1136/bmj.a2338.
- Cohen-Cole E, Fletcher JM. (2008). Detecting Implausible Social Network Effects in Acne, Height, and Headaches: Longitudinal Analyses. British Medical Journal, 337, a2533.
- Lyons, R. (2011), The spread of evidence-poor medicine via flawed social network analyses, Statistics, Politics and Policy, 2, 1-26, doi:10.2202/2151-7509.1024
- Shalizi CR, Thomas AC. (2011). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies, Sociological Methods and Research, 40, 211--239.
- O’Malley AJ, Elwert F, Rosenquist JN, Zaslavsky AM, Christakis NA. Estimating peer effects in longitudinal dyadic data using instrumental variables. Biometrics, 2014, 70, 3, 506–515. (Published online: 29 APR 2014, DOI:10.1111/biom.12172).

References: social influence analyses cont.

- Christakis, N. A. and Fowler, J. H. (2013), Social Contagion Theory: Examining Dynamic Social Networks and Human Behavior," Statistics in Medicine, 32, 556-577
- Manski, C. A. (1993), Identfication of endogenous social effects: The Reflection Problem, Review of Economic Studies, 60, 531-542
- Marsden, P. V. and Friedkin, N. E. (1993), Network Studies of Social Influence, Sociological Methods and Research, 22, 127-151
- McPherson, M. L., Smith-Lovin, Cook, and et al (2001), Birds of a Feather: Homophily in Social Networks, Annual Review of Sociology, 27, 415-444
- VanderWeele, T. J. (2011), Sensitivity Analysis for Contagion Effects in Social Networks, Sociological Methods & Research, 40, 240-255
- VanderWeele, T. J., Ogburn, E. L., and Tchetgen Tchetgen, E. J. (2012), Why and When "Flawed" Social Network Analyses still yield Valid Tests of no Contagion, Statistics, Politics, and Policy, Manuscript 1050
- Keating NL, O'Malley AJ, Onnela J-P, Gray SQ, Landon BE. Influence of Peer Physicians on Intensity of End-of-Life Care for Cancer Decedents. Medical Care, 57, 6, 468-474. PMID: 31033059

References: network-influence related models

- Land KC, Deane G. (1992). On the Large-Sample Estimation of Regression Models with Spatial or Network Effect Terms: A Two-Stage Least-Squares Approach. Sociological Methodology, (ed: Peter V. Marsden), Oxford, UK: Basil Blackwell, Ltd, 221-248
- Anselin L. (1988). Spatial Econometrics: Methods and Models, Kluwer Academic Publishers: Dordrecht, The Netherlands
- Sargan JD. (1958). The Estimation of Econometric Relationships Using Instrumental Variables, Econometrica, 26, 393-415
- Haining, R.P. (1978). The moving average model for spatial interaction. Transactions of the Institute of British Geographers, 3, 202-225
- Kelejian, H.H., Robinson, D.P., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297-312
- O’Malley AJ, Moen EL, Bynum JPW, Austin AM, Skinner JS. Modeling Peer Effect Modification by Network Position: The Diffusion of Implantable Cardioverter Defibrillators in the US Hospital Network. In Press: Statistics in Medicine