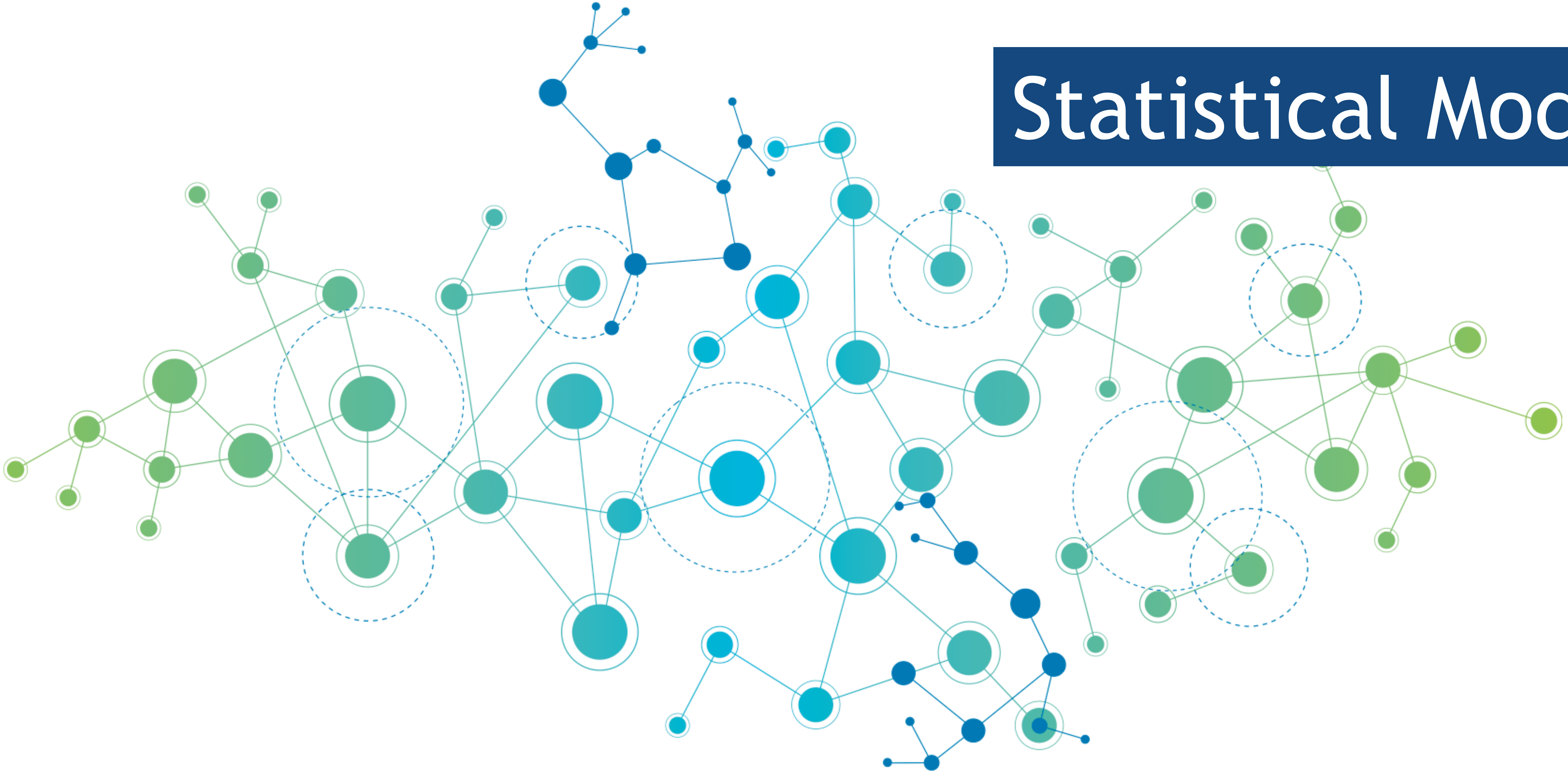


Statistical Models

Center for Public Health Systems Science

Brown School

Washington University in St. Louis

Goals

- Describe challenge of statistically modeling networks
- Explore reasons for modeling networks
- Explore exponential random graph models (*ergm*)
  - Conceptual foundations
  - Statistical and simulation aspects
  - Examples

Statistical Models

Modeling network ties and structures to explore and test hypotheses

What is a statistical model?

The word "model" means different things in different subfields

- A *statistical* model is a
  - formal representation of a
  - stochastic process
  - specified at one level (e.g., person, dyad) that
  - aggregates to a higher level (e.g., population, network)

Why take a statistical approach?

Descriptive vs. generative goals

- *Descriptive*: numerical summary measures
  - Nodal level: e.g., centrality, geodesic distribution
  - Configuration level: e.g., cycle census
  - Network level: e.g., centralization, clustering, small world, core/periphery
- *Generative*: micro foundations for macro patterns
  - Recover underlying dynamic process from x-sectional data
  - Test alternative hypotheses
  - Extrapolate and simulate from model

Statistical analysis, given a model

- Estimate parameters of the process
  - Joint estimation of multiple, possibly correlated, effects
- Inference from sampled data to population
  - Uncertainty in parameter estimates
- Goodness of fit
  - Traditional diagnostics
    - Model fit (BIC, AIC)
    - Estimation diagnostics (MCMC performance)
  - Network-specific GOF
    - Network statistics in the model as covariates
    - Network properties not in the model

ERGM – current solution to stochastic modeling of networks

- ERGM = exponential random graph model
  - Fit by Monte Carlo Markov Chain maximum likelihood (MCMC)
  - Simulation approach
- ERGMs are popular because
  - Handle complex dependencies in network data
  - Model is flexible and can handle many different types of covariates (i.e., predictors)
  - Produce generative models (overall network structure can be understand based on small number of local predictors)

What processes contribute to differences between observed & random networks?

- Some potential hypotheses:
  - Sociality: Network members are different in their tendency to form ties, some may form many ties while others form few
  - Homophily: Network member tend to form ties with similar others
  - Reciprocity: Network members tend to form mutual connections (directed networks only)
- Can use ERGM to test these hypotheses

Golder & Yardi, 2010; Goodreau et al., 2009

Network modeling today

- Statistical models of observed networks
  - Account for complex dependencies
  - Explain the observed structure of a network based on general network characteristics (e.g., density) and social processes (e.g., homophily)
  - Called exponential random graph models (ergm)

- Model form and interpretation are like logistic regression predicting the probability of a tie
  - Results can be used to predict the likelihood that there is a link between any two network members
    - P(tie) = density + sociality + homophily
  - Reporting can use odds ratios

ERGMs are a hybrid:

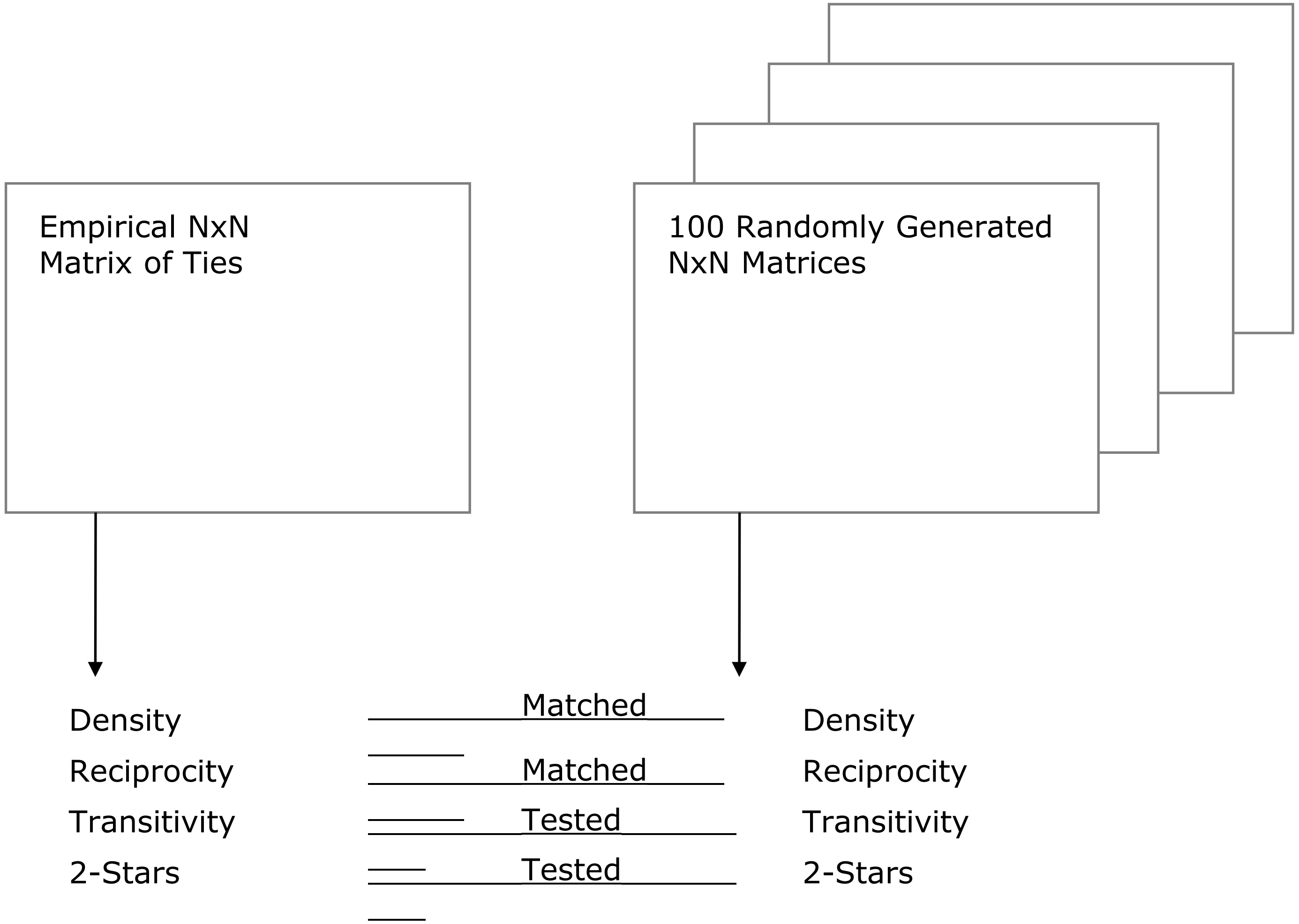
Traditional generalized linear models (statistical)

- *but*
  - Unit of analysis: relation (dyad)
  - Observations may be dependent (like a complex system)
  - Complex nonlinear and threshold effects
  - Estimation is different

Agent-based models (mathematical)

- *but*
  - Can estimate model parameters from data
  - Can test model goodness of fit

General *ergm* approach (Valente, 2010)



Basic *ergm* statistical model

$$\underbrace{P(Y=y)}_{\boxed{\text{Probability of the graph}}} = \exp\left\{\sum_{k=1}^{K} \underbrace{\theta_k g_k(y)}_{\boxed{\text{coefficient*covariate}}}\right\} / \kappa(\theta)$$

Re-expressed in terms of p(tie)

$$P(Y=y)=\exp\left\{\sum_{k=1}^{K}\theta_k g_k(y)\right\}/\kappa(\theta)$$

Probability of the graph

$$P(y_{ij}=1\,|\,Y^{(ij)})=P(Y^{+})/\{P(Y^{+})+P(Y^{-})\}$$

Probability of *ij* tie, conditional on the rest of the graph

$$\text{logit}[P(y_{ij}=1)\,|\,Y^{(ij)}]=\theta_1\partial_1(y^{(ij)})+\theta_2\partial_2(y^{(ij)})+\ldots+\theta_k\partial_k(y^{(ij)})$$

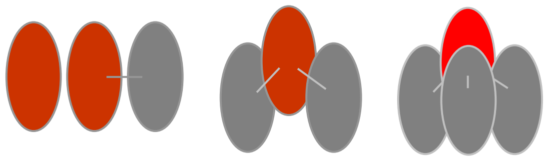
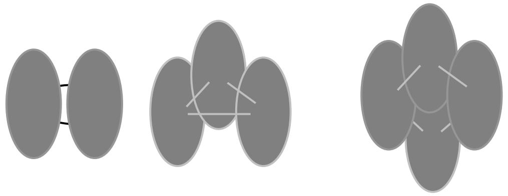
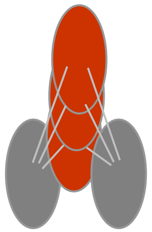
Conditional log odds of the tie.

δ is the "change statistic", the change in the value of the covariate g(y) when the *ij* tie changes from 1 to 0

So θ is the impact of the covariate on the log-odds of a tie

What kinds of covariates?

What creates heterogeneity in the probability of a tie being formed?

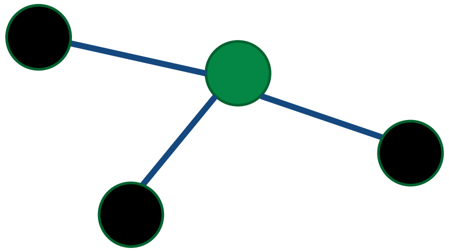
attributes of nodes	Heterogeneity by group – Average activity – Mixing by group Individual heterogeneity
attributes of links	Heterogeneity in – Duration – Types (sex, drug…)
configurations	Degree distributions (or stars)  Cycle distributions (2, 3, 4, etc.)  Shared partner distributions 

Dyad Independent Terms

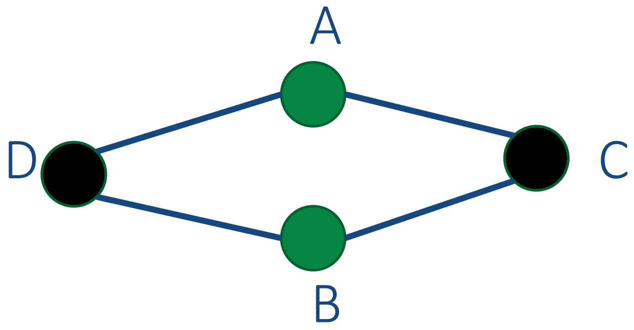
Dyad Dependent Terms

Structural terms in network models

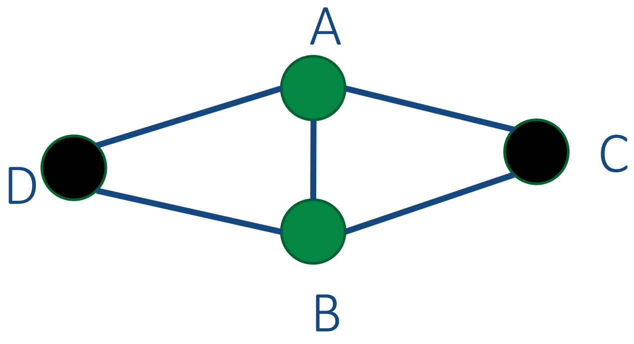
- GWDegree
  - Geometrically Weighted degree
  - Accounts for non-uniform degree distribution



- GWDSP
  - Geometrically Weighted Dyadwise (Non-edgewise) Shared Partnerships
  - Accounts for the distribution of shared partners for unlinked nodes (a measure of structural equivalence)

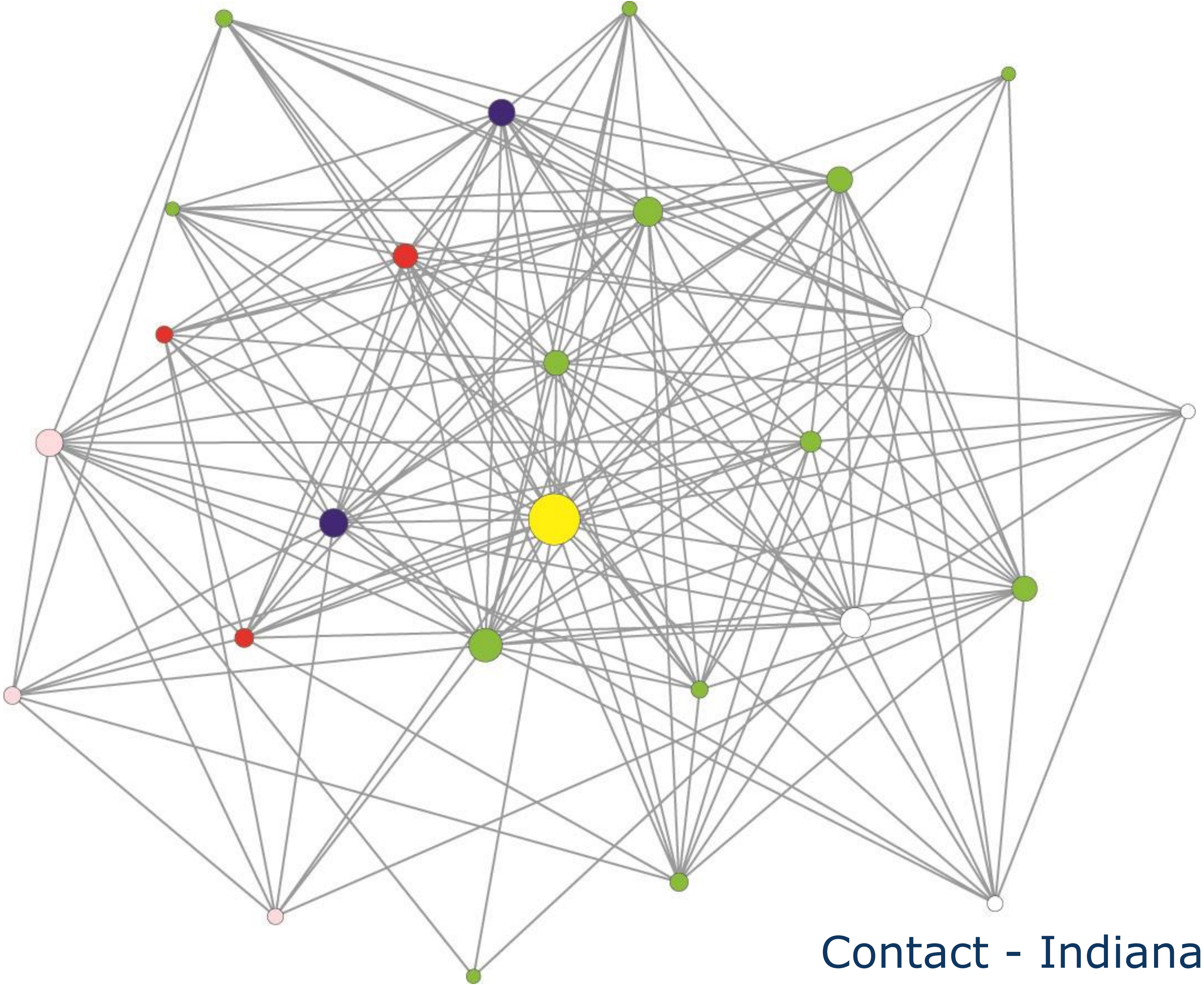


- GWESP
  - Geometrically Weighted Edgewise Shared Partnerships
  - Accounts for the distribution of shared partners for linked nodes (a measure of clustering)

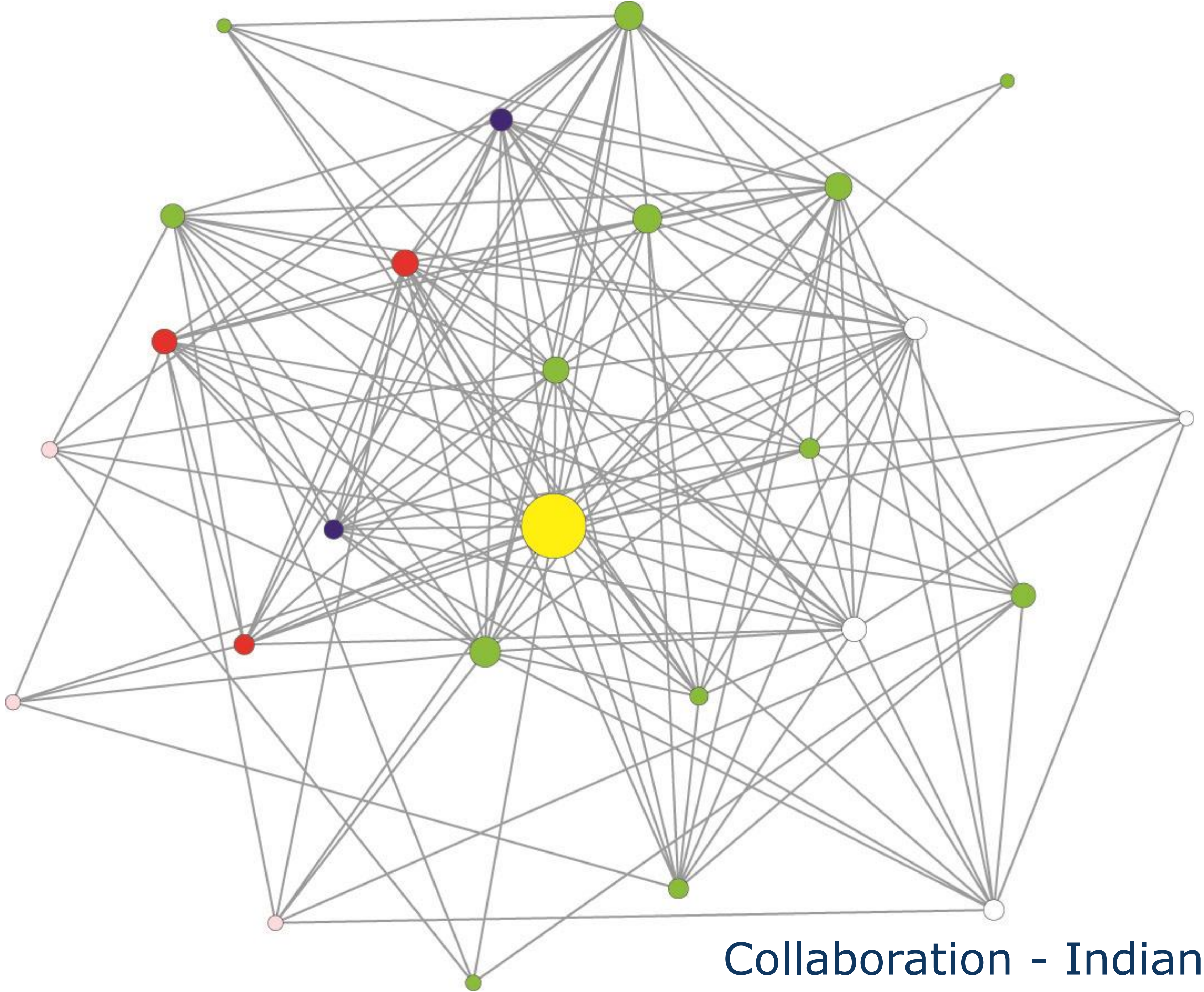


Example – Predictors of dissemination across a national public health network

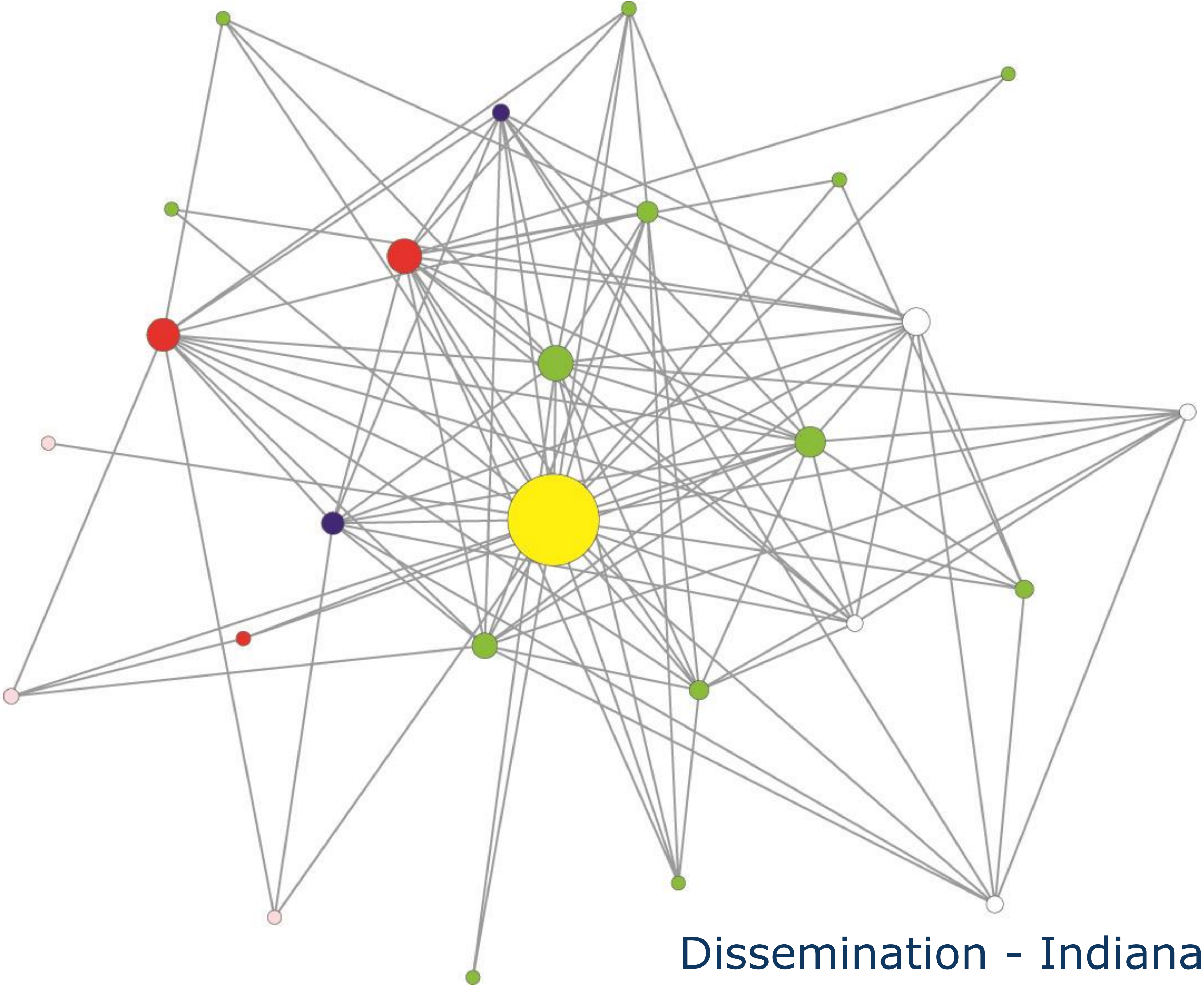
- Identifying predictors of dissemination of evidence-based guidelines in statewide tobacco control programs
- Part of a CDC-funded evaluation project of state tobacco control programs
  - Examining how states are disseminating and implementing the evidence-based *Best Practices* guidelines
  - Luke, D. A., Harris, J. K., Shelton, S., Allen, P., Carothers, B. J., & Mueller, N. B. (2010). Systems analysis of collaboration in 5 national tobacco control networks. *American Journal of Public Health, 100,* 1290-1297



Contact - Indiana

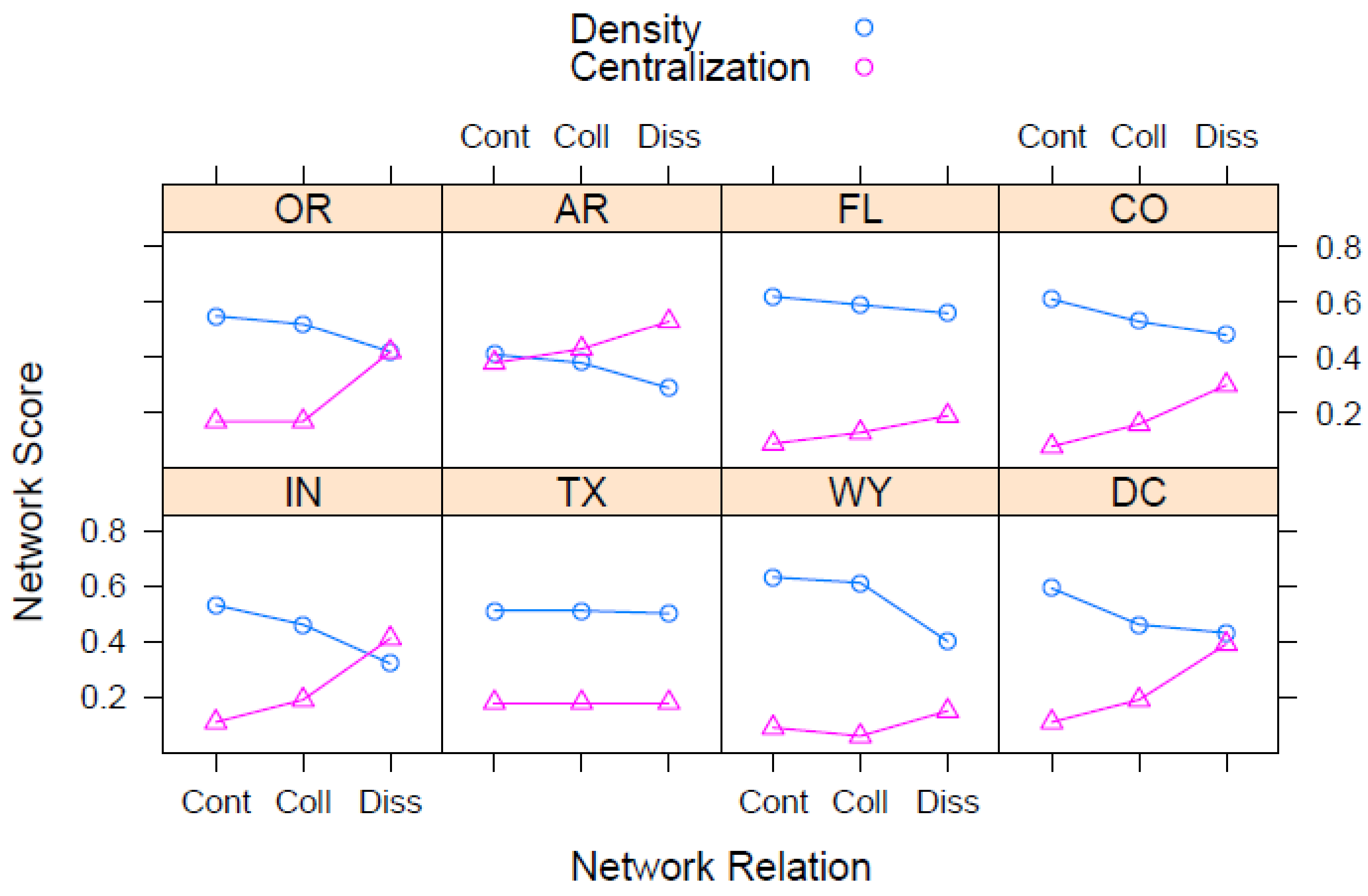


Collaboration - Indiana



Dissemination - Indiana

Network characteristics across 3 types of relationships: Contact, Collaboration, Dissemination



Prediction of EBG dissemination

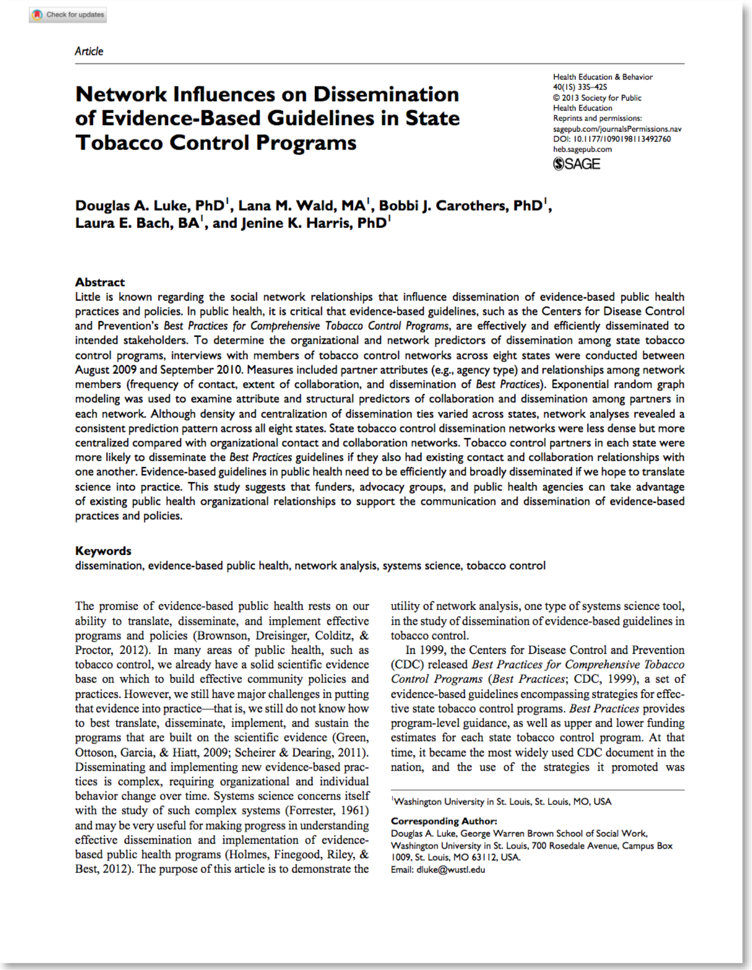
	Indiana (g=26)			Texas (g=20)			Wyoming (g=20)			DC (g=19)		
	M1	M2	M3	M1	M2	M3	M1	M2	M3	M1	M2	M3
Parameters												
Edges	-2.63	-4.21	-4.05	-0.16	-0.97	-0.93	-1.05	-1.87	-2.23	-2.32	-4.37	-5.43
TC Experience	0.11 (.02)*	0.10 (.02)*	0.08 (.02)*	0.03 (.02)	0.02 (.03)	0.01 (.03)	0.09 (.03)*	0.07 (.03)*	0.07 (.03)*	0.18 (.04)*	0.16 (.04)*	0.20 (.05)*
Agency Level (Homophily)	0.59 (.08)*	0.61 (.10)*	0.32 (.10)*	0.83 (.10)*	0.85 (.10)*	0.78 (.11)*	-0.22 (.11)	-0.46 (.10)*	-0.45 (.10)*	0.17 (.10)	.02 (.12)	0.58 (.13)*
Agency Distance	.000 (.000)	.000 (.000)	.000 (.000)	-.001 (.000)*	-.001 (.000)*	-.001 (.000)*	-.000 (.000)	-.001 (.000)*	-.000 (.000)	-.001 (.00)*	-.001 (.00)*	-.000 (.00)
Degree (GWDegree)	2.82 (1.39)*	2.87 (1.63)	4.53 (1.79)*	200.6 (7.3)*	296.4 (7.6)*	302.3 (7.6)*	-4.58 (.44)*	-3.33 (.55)*	-3.41 (.54)*	3.54 (1.44)*	3.32 (1.75)	1.97 (1.70)
Network Contact(OR) (95% CI)		2.69 (2.54-2.85)	2.01 (1.90-2.14)		2.05 (1.90-2.22)	1.73 (1.63-1.84)		2.03 (1.92-2.16)	1.68 (1.59-1.78)		3.63 (3.36-3.93)	1.86 (1.69-2.05)
Network Collaboration(OR) (95% CI)			1.65 (1.55-1.75)			1.32 (1.25-1.40)			1.38 (1.30-1.46)			2.83 (2.57-3.12)
Model Fit												
AIC	372.2	285.4	272.2	272.1	237.5	236.1	212.1	187.55	186.8	203.0	137.3	120.0
p	.000	.000	.000	.000	.000	.066	.000	.000	.096	.000	.000	.000
GOF (%)	.74	.84	.92	.85	.93	.88	.74	.75	.78	.82	.90	.93

Contact and collaboration predict dissemination

	Indiana (g=26)		Texas (g=20)		Wyoming (g=20)		DC (g=19)	
	M2	M3	M2	M3	M2	M3	M2	M3
Network *Contact* (OR)	2.69	2.01	2.05	1.73	2.03	1.68	3.63	1.86
(95% CI)	(2.54-2.85)	(1.90-2.14)	(1.90-2.22)	(1.63-1.84)	(1.92-2.16)	(1.59-1.78)	(3.36-3.93)	(1.69-2.05)
Network *Collaboration* (OR)		1.65		1.32		1.38		2.83
(95% CI)		(1.55-1.75)		(1.25-1.40)		(1.30-1.46)		(2.57-3.12)

	Oregon (g=17)		Arkansas (g=17)		Florida (g=16)		Colorado (g=15)	
	M2	M3	M2	M3	M2	M3	M2	M3
Network *Contact* (OR)	2.92	1.45	5.00	2.64	4.53	3.56	3.53	2.92
(95% CI)	(2.7-3.15)	(1.31-1.60)	(4.45-5.63)	(2.26-3.09)	(4.02-5.09)	(3.17-4.01)	(3.13-3.97)	(2.59-3.28)
Network *Collaboration* (OR)		3.46		2.89		1.35		1.35
(95% CI)		(3.13-3.81)		(2.52-3.31)		(1.20-1.52)		(1.20-1.52)

ERGMs being used more widely



Check for updates

Article

Health Education & Behavior
40(1S) 33S–42S

DOI: 10.1177/1090198113492760
heb.sagepub.com
SAGE

Network Influences on Dissemination of Evidence-Based Guidelines in State Tobacco Control Programs

Douglas A. Luke, PhD[1], Lana M. Wald, MA[1], Bobbi J. Carothers, PhD[1], Laura E. Bach, BA[1], and Jenine K. Harris, PhD[1]

Abstract

Little is known regarding the social network relationships that influence dissemination of evidence-based public health practices and policies. In public health, it is critical that evidence-based guidelines, such as the Centers for Disease Control and Prevention's *Best Practices for Comprehensive Tobacco Control Programs*, are effectively and efficiently disseminated to intended stakeholders. To determine the organizational and network predictors of dissemination among state tobacco control programs, interviews with members of tobacco control networks across eight states were conducted between August 2009 and September 2010. Measures included partner attributes (e.g., agency type) and relationships among network members (frequency of contact, extent of collaboration, and dissemination of *Best Practices*). Exponential random graph modeling was used to examine attribute and structural predictors of collaboration and dissemination among partners in each network. Although density and centralization of dissemination ties varied across states, network analyses revealed a consistent prediction pattern across all eight states. State tobacco control dissemination networks were less dense but more centralized compared with organizational contact and collaboration networks. Tobacco control partners in each state were more likely to disseminate the *Best Practices* guidelines if they also had existing contact and collaboration relationships with one another. Evidence-based guidelines in public health need to be efficiently and broadly disseminated if we hope to translate science into practice. This study suggests that funders, advocacy groups, and public health agencies can take advantage of existing public health organizational relationships to support the communication and dissemination of evidence-based practices and policies.

Keywords

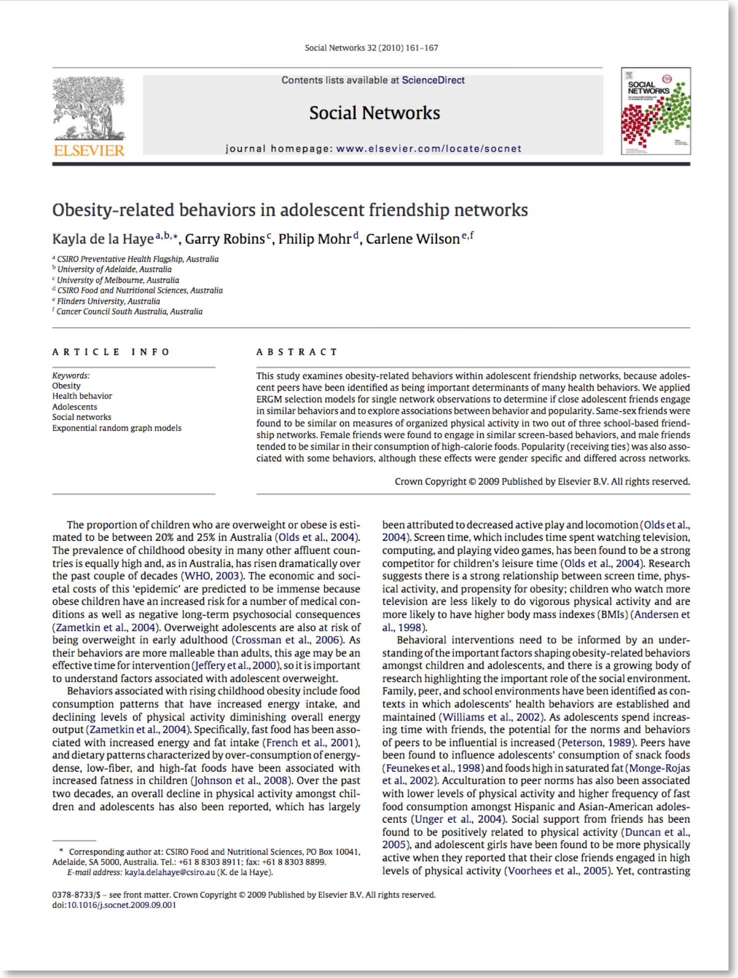
dissemination, evidence-based public health, network analysis, systems science, tobacco control

The promise of evidence-based public health rests on our ability to translate, disseminate, and implement effective programs and policies (Brownson, Dreisinger, Colditz, & Proctor, 2012). In many areas of public health, such as tobacco control, we already have a solid scientific evidence base on which to build effective community policies and practices. However, we still have major challenges in putting that evidence into practice—that is, we still do not know how to best translate, disseminate, implement, and sustain the programs that are built on the scientific evidence (Green, Ottoson, Garcia, & Hiatt, 2009; Scheirer & Dearing, 2011). Disseminating and implementing new evidence-based practices is complex, requiring organizational and individual behavior change over time. Systems science concerns itself with the study of such complex systems (Forrester, 1961) and may be very useful for making progress in understanding effective dissemination and implementation of evidence-based public health programs (Holmes, Finegood, Riley, & Best, 2012). The purpose of this article is to demonstrate the utility of network analysis, one type of systems science tool, in the study of dissemination of evidence-based guidelines in tobacco control.

In 1999, the Centers for Disease Control and Prevention (CDC) released *Best Practices for Comprehensive Tobacco Control Programs* (*Best Practices*; CDC, 1999), a set of evidence-based guidelines encompassing strategies for effective state tobacco control programs. *Best Practices* provides program-level guidance, as well as upper and lower funding estimates for each state tobacco control program. At that time, it became the most widely used CDC document in the nation, and the use of the strategies it promoted was

[1]Washington University in St. Louis, St. Louis, MO, USA

Corresponding Author:
Douglas A. Luke, George Warren Brown School of Social Work, Washington University in St. Louis, 700 Rosedale Avenue, Campus Box 1009, St. Louis, MO 63112, USA.
Email: dluke@wustl.edu



Social Networks 32 (2010) 161–167

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

ELSEVIER

Obesity-related behaviors in adolescent friendship networks

Kayla de la Haye[a,b,*], Garry Robins[c], Philip Mohr[d], Carlene Wilson[e,f]

[a] *CSIRO Preventative Health Flagship, Australia*
[b] *University of Adelaide, Australia*
[c] *University of Melbourne, Australia*
[d] *CSIRO Food and Nutritional Sciences, Australia*
[e] *Flinders University, Australia*
[f] *Cancer Council South Australia, Australia*

A R T I C L E   I N F O

Keywords:
Obesity
Health behavior
Adolescents
Social networks
Exponential random graph models

A B S T R A C T

This study examines obesity-related behaviors within adolescent friendship networks, because adolescent peers have been identified as being important determinants of many health behaviors. We applied ERGM selection models for single network observations to determine if close adolescent friends engage in similar behaviors and to explore associations between behavior and popularity. Same-sex friends were found to be similar on measures of organized physical activity in two out of three school-based friendship networks. Female friends were found to engage in similar screen-based behaviors, and male friends tended to be similar in their consumption of high-calorie foods. Popularity (receiving ties) was also associated with some behaviors, although these effects were gender specific and differed across networks.

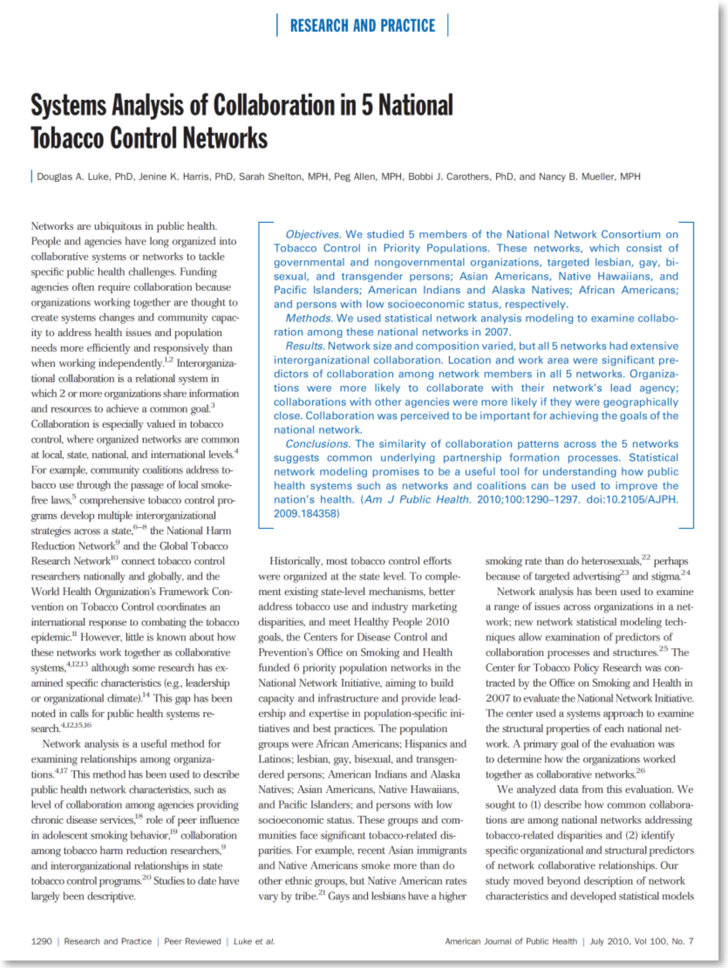
The proportion of children who are overweight or obese is estimated to be between 20% and 25% in Australia (Olds et al., 2004). The prevalence of childhood obesity in many other affluent countries is equally high and, as in Australia, has risen dramatically over the past couple of decades (WHO, 2003). The economic and societal costs of this 'epidemic' are predicted to be immense because obese children have an increased risk for a number of medical conditions as well as negative long-term psychosocial consequences (Zametkin et al., 2004). Overweight adolescents are also at risk of being overweight in early adulthood (Crossman et al., 2006). As their behaviors are more malleable than adults, this age may be an effective time for intervention (Jeffery et al., 2000), so it is important to understand factors associated with adolescent overweight.

Behaviors associated with rising childhood obesity include food consumption patterns that have increased energy intake, and declining levels of physical activity diminishing overall energy output (Zametkin et al., 2004). Specifically, fast food has been associated with increased energy and fat intake (French et al., 2001), and dietary patterns characterized by over-consumption of energy-dense, low-fiber, and high-fat foods have been associated with increased fatness in children (Johnson et al., 2008). Over the past two decades, an overall decline in physical activity amongst children and adolescents has also been reported, which has largely been attributed to decreased active play and locomotion (Olds et al., 2004). Screen time, which includes time spent watching television, computing, and playing video games, has been found to be a strong competitor for children's leisure time (Olds et al., 2004). Research suggests there is a strong relationship between screen time, physical activity, and propensity for obesity; children who watch more television are less likely to do vigorous physical activity and are more likely to have higher body mass indexes (BMIs) (Andersen et al., 1998).

Behavioral interventions need to be informed by an understanding of the important factors shaping obesity-related behaviors amongst children and adolescents, and there is a growing body of research highlighting the important role of the social environment. Family, peer, and school environments have been identified as contexts in which adolescents' health behaviors are established and maintained (Williams et al., 2002). As adolescents spend increasing time with friends, the potential for the norms and behaviors of peers to be influential is increased (Peterson, 1989). Peers have been found to influence adolescents' consumption of snack foods (Feunekes et al., 1998) and foods high in saturated fat (Monge-Rojas et al., 2002). Acculturation to peer norms has also been associated with lower levels of physical activity and higher frequency of fast food consumption amongst Hispanic and Asian-American adolescents (Unger et al., 2004). Social support from friends has been found to be positively related to physical activity (Duncan et al., 2005), and adolescent girls have been found to be more physically active when they reported that their close friends engaged in high levels of physical activity (Voorhees et al., 2005). Yet, contrasting

* Corresponding author at: CSIRO Food and Nutritional Sciences, PO Box 10041, Adelaide, SA 5000, Australia. Tel.: +61 8 8303 8911; fax: +61 8 8303 8899.
E-mail address: kayla.delahaye@csiro.au (K. de la Haye).

0378-8733/$ – see front matter. Crown Copyright © 2009 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2009.09.001



RESEARCH AND PRACTICE

Systems Analysis of Collaboration in 5 National Tobacco Control Networks

Douglas A. Luke, PhD, Jenine K. Harris, PhD, Sarah Shelton, MPH, Peg Allen, MPH, Bobbi J. Carothers, PhD, and Nancy B. Mueller, MPH

Objectives. We studied 5 members of the National Network Consortium on Tobacco Control in Priority Populations. These networks, which consist of governmental and nongovernmental organizations, targeted lesbian, gay, bisexual, and transgender persons; Asian Americans, Native Hawaiians, and Pacific Islanders; American Indians and Alaska Natives; African Americans; and persons with low socioeconomic status, respectively.

Methods. We used statistical network analysis modeling to examine collaboration among these national networks in 2007.

Results. Network size and composition varied, but all 5 networks had extensive interorganizational collaboration. Location and work area were significant predictors of collaboration among network members in all 5 networks. Organizations were more likely to collaborate with their network's lead agency; collaborations with other agencies were more likely if they were geographically close. Collaboration was perceived to be important for achieving the goals of the national network.

Conclusions. The similarity of collaboration patterns across the 5 networks suggests common underlying partnership formation processes. Statistical network modeling promises to be a useful tool for understanding how public health systems such as networks and coalitions can be used to improve the nation's health. (*Am J Public Health.* 2010;100:1290–1297. doi:10.2105/AJPH.2009.184358)

Networks are ubiquitous in public health. People and agencies have long organized into collaborative systems or networks to tackle specific public health challenges. Funding agencies often require collaboration because organizations working together are thought to create systems changes and community capacity to address health issues and population needs more efficiently and responsively than when working independently.[1,2] Interorganizational collaboration is a relational system in which 2 or more organizations share information and resources to achieve a common goal.[3] Collaboration is especially valued in tobacco control, where organized networks are common at local, state, national, and international levels.[4] For example, community coalitions address tobacco use through the passage of local smoke-free laws,[5] comprehensive tobacco control programs develop multiple interorganizational strategies across a state,[6–8] the National Harm Reduction Network[9] and the Global Tobacco Research Network[10] connect tobacco control researchers nationally and globally, and the World Health Organization's Framework Convention on Tobacco Control coordinates an international response to combating the tobacco epidemic.[11] However, little is known about how these networks work together as collaborative systems,[4,12,13] although some research has examined specific characteristics (e.g., leadership or organizational climate).[14] This gap has been noted in calls for public health systems research.[4,12,15,16]

Network analysis is a useful method for examining relationships among organizations.[4,17] This method has been used to describe public health network characteristics, such as level of collaboration among agencies providing chronic disease services,[18] role of peer influence in adolescent smoking behavior,[19] collaboration among tobacco harm reduction researchers,[9] and interorganizational relationships in state tobacco control programs.[20] Studies to date have largely been descriptive.

Historically, most tobacco control efforts were organized at the state level. To complement existing state-level mechanisms, better address tobacco use and industry marketing disparities, and meet Healthy People 2010 goals, the Centers for Disease Control and Prevention's Office on Smoking and Health funded 6 priority population networks in the National Network Initiative, aiming to build capacity and infrastructure and provide leadership and expertise in population-specific initiatives and best practices. The population groups were African Americans; Hispanics and Latinos; lesbian, gay, bisexual, and transgendered persons; American Indians and Alaska Natives; Asian Americans, Native Hawaiians, and Pacific Islanders; and persons with low socioeconomic status. These groups and communities face significant tobacco-related disparities. For example, recent Asian immigrants and Native Americans smoke more than do other ethnic groups, but Native American rates vary by tribe.[21] Gays and lesbians have a higher smoking rate than do heterosexuals,[22] perhaps because of targeted advertising[23] and stigma.[24]

Network analysis has been used to examine a range of issues across organizations in a network; new network statistical modeling techniques allow examination of predictors of collaboration processes and structures.[25] The Center for Tobacco Policy Research was contracted by the Office on Smoking and Health in 2007 to evaluate the National Network Initiative. The center used a systems approach to examine the structural properties of each national network. A primary goal of the evaluation was to determine how the organizations worked together as collaborative networks.[26]

We analyzed data from this evaluation. We sought to (1) describe how common collaborations are among national networks addressing tobacco-related disparities and (2) identify specific organizational and structural predictors of network collaborative relationships. Our study moved beyond description of network characteristics and developed statistical models

R-statnet and ergm resources

- Several articles/tutorials about statnet, ergm, and other network analysis packages in R are included in the materials we distributed
- The statnet website has tutorials, resources, and other materials that are extremely useful for learning and using ergm: http://statnet.csde.washington.edu/
- Two online groups that are active and provide excellent technical assistance:
  - Statnet user group: https://mailman.u.washington.edu/mailman/listinfo/statnet_help
  - Socnet listserv: http://www.insna.org/pubs/socnet.html
- Of course, Jenine's Sage book on *ergm* models