A Précis of Key Types of Social Network Analyses and Recent Applications Involving Physician Networks

James O'Malley, Ph.D.
Geisel Medical School at Dartmouth

April 20, 2018

Funding

P01 AG 0309301
U01 AG 046830

Acknowledgements

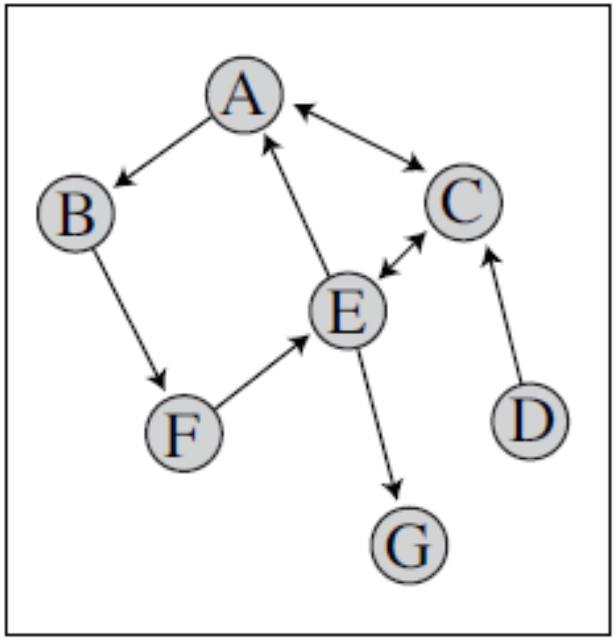
Erika Moen, Julie Bynum, Andrea Austin, Gouri Chakraborti, Jon Skinner

Outline

1. Introduction to social networks and social network analyses
    - Three key general types of problems in networks
2. Example 1: Accounting for variation in whether use of implantable cardiac defibrillators (ICDs) is within guidelines
3. Example 2: Modeling the inter-hospital diffusion in adoption of capability to implant ICDs

Definition of a social network

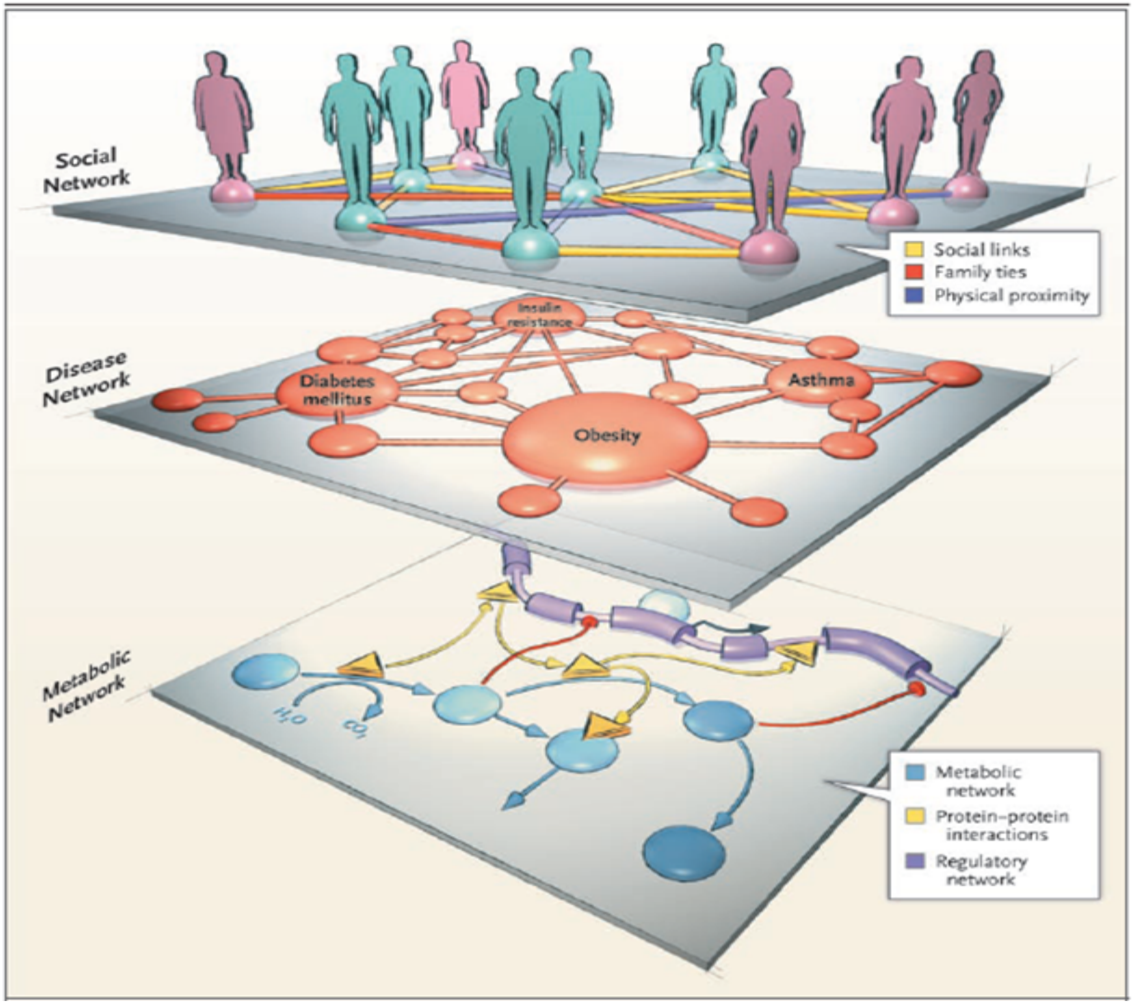
- A social network consists of one or more sets of actors—also known as “units,” “nodes,” or “vertices”—together with the possibly directed relationships or social ties among them



Components of a social network

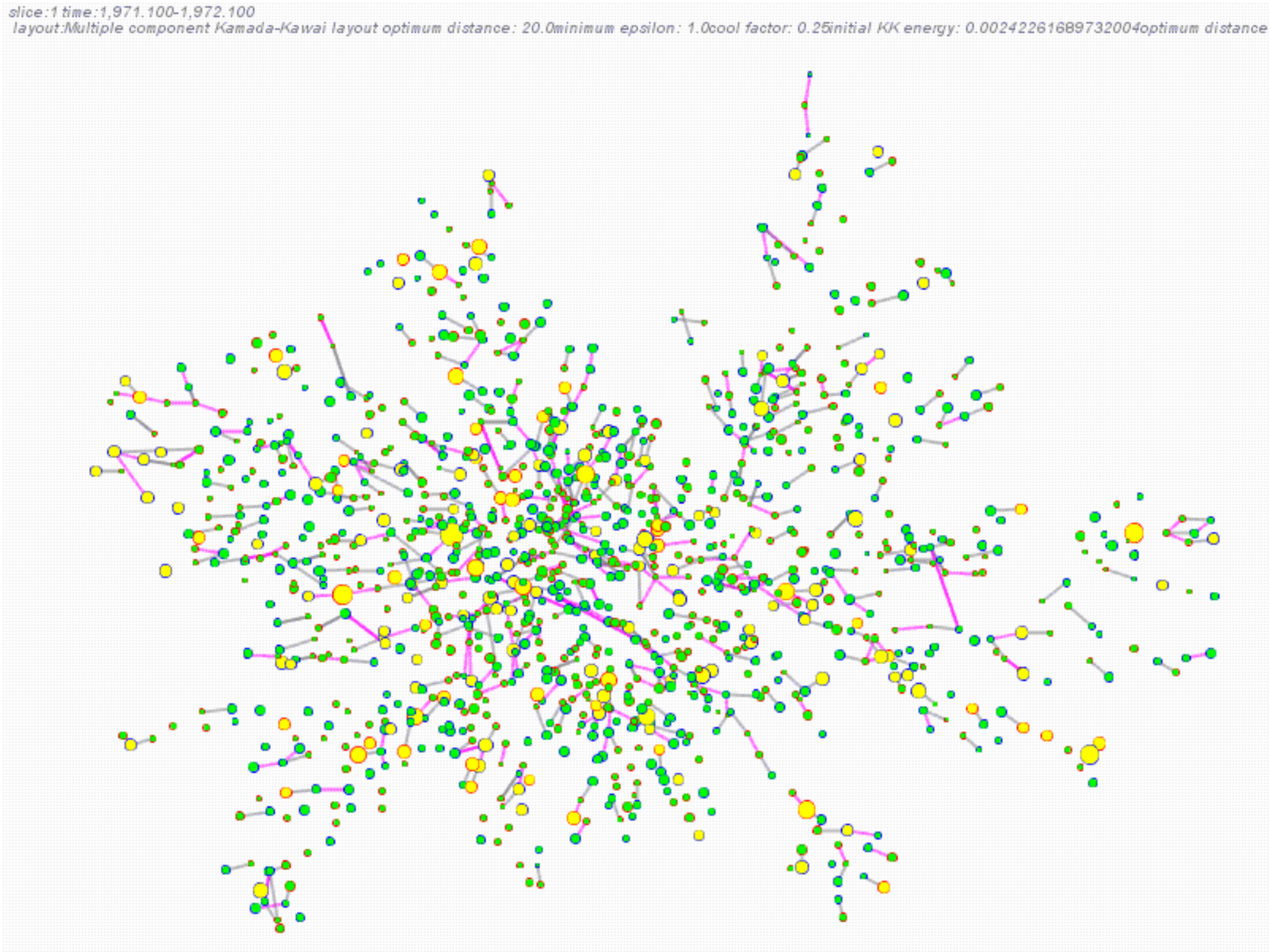
- Actors:
  - Individual persons (e.g., patients or clinicians)
  - Organizations (e.g., hospitals)
  - Health states (e.g., diseases)
  - Work products (e.g., academic papers)
- Social ties:
  - Communication
  - Influence
  - Trust or affect (e.g., friendship)
  - Affiliations (e.g., co-authors)
- Attributes:
  - Actors, relationships, or both

Layers of Networks in Medicine



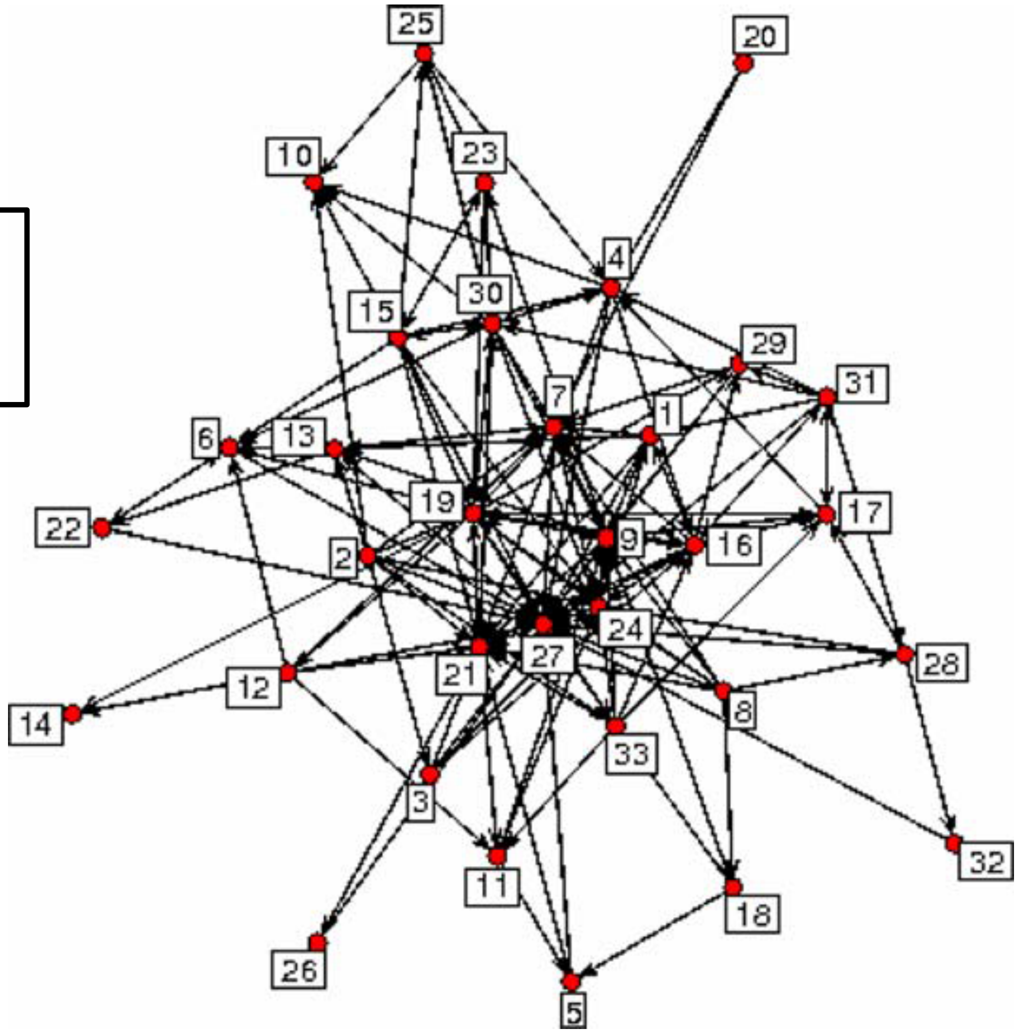
Barabási (New England Journal of Medicine 2007)

Spread of Obesity in Framingham Heart Study (Christakis and Fowler, 2007)



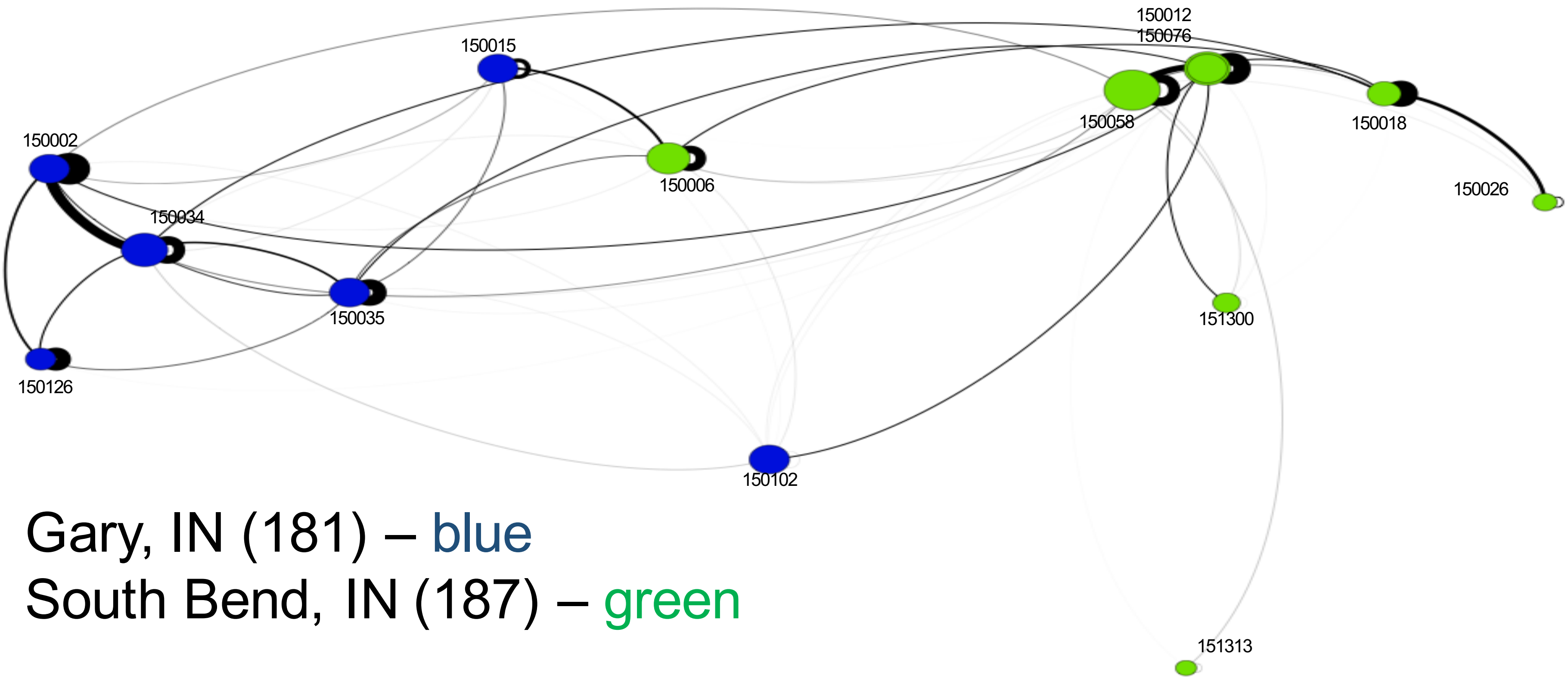
Social network of physicians in a Boston health clinic

Spring embedder algorithm determines positions of actors (Fruchterman and Reingold 1991)



Keating et al (2007), O'Malley and Marsden (2008). Ties reflect influential discussions among physicians

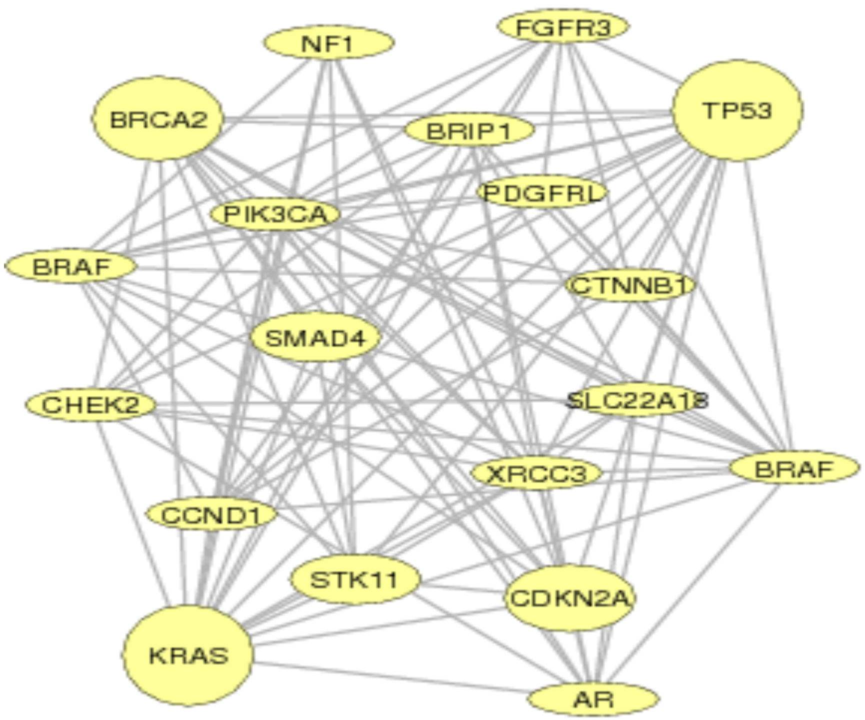
Network of Hospitals in Two Adjourning Health Referral Regions (Moen et al, 2015)



Gary, IN (181) – blue
South Bend, IN (187) – green

Edges: Thickness reflects # cardiovascular disease patients treated at both hospitals by at least one physician

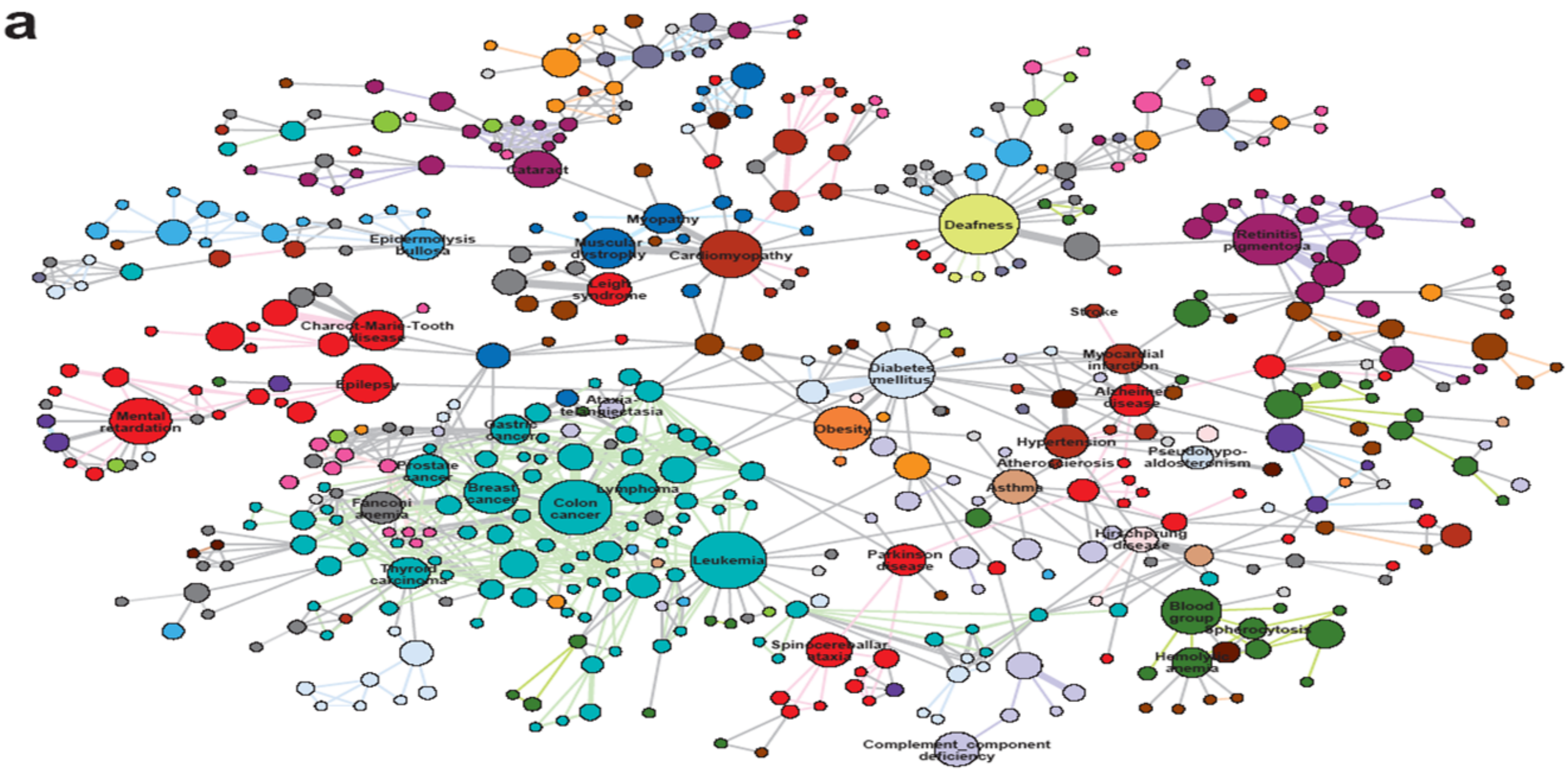
Biological network
(genes as nodes, shared proteins reflect edges)



Gene network

Adapted from: Goh, Cusick, Valle, Childs, Vidal & Barabási (PNAS 2007)

Human Disease Network (health phenotypes as nodes; edges reflect shared genes)



Adapted from: Goh, Cusick, Valle, Childs, Vidal & Barabási (PNAS 2007)

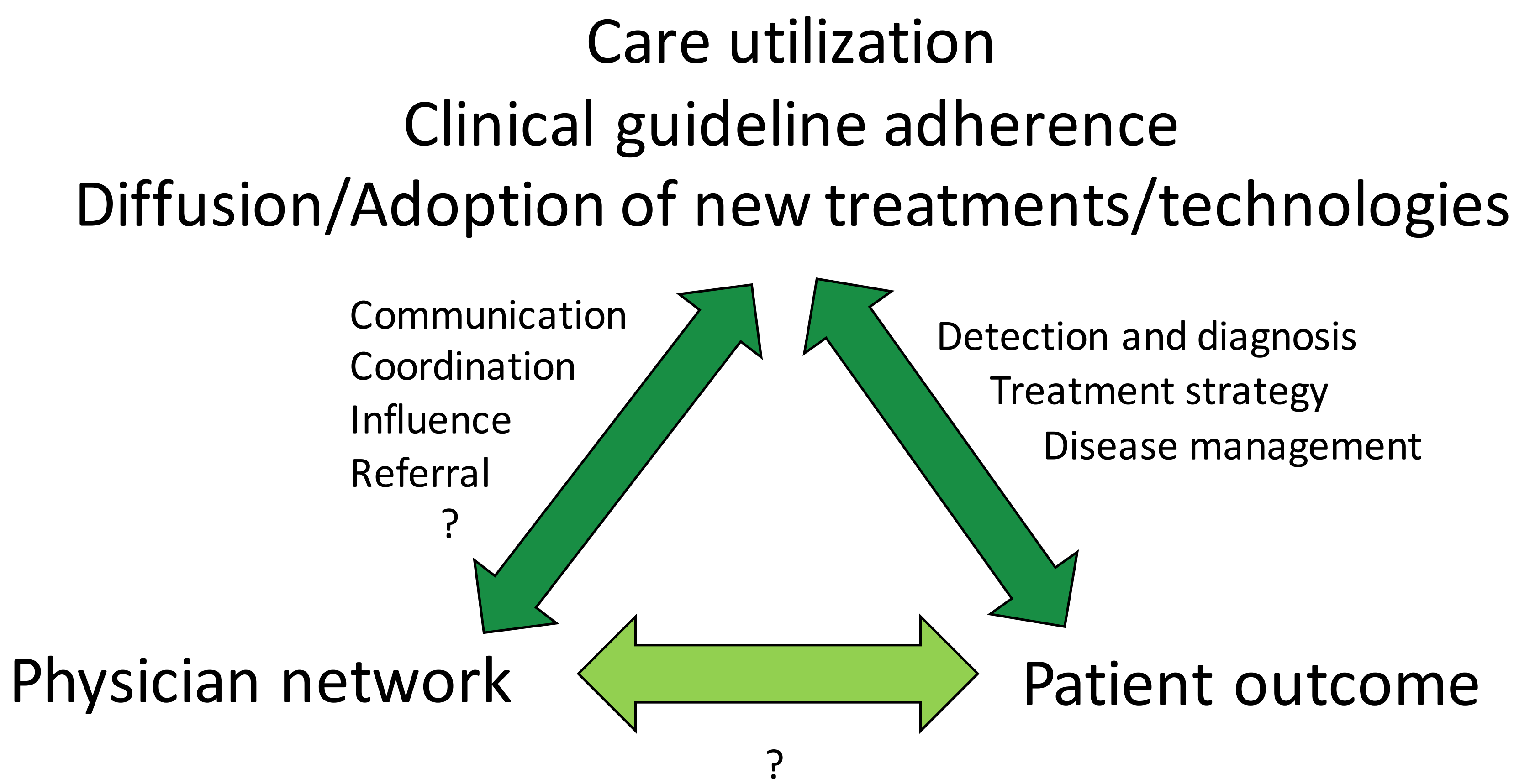
Three "Important" Types of Social Network Problems

I. Do features of social networks correlate with health outcome variables of interest?
   - Multiple networks
   - Example: do network characteristics of a health care organization and the network positions of providers caring for a patient within the system correlate with utilization, quality and cost of care?

II. Do physicians influence one another, leading to diffusion of medical ideas/habits/practices?
   - Example: diffusion of use of medical treatments across physicians
   - Long history of work (Coleman, 1957, 1966)

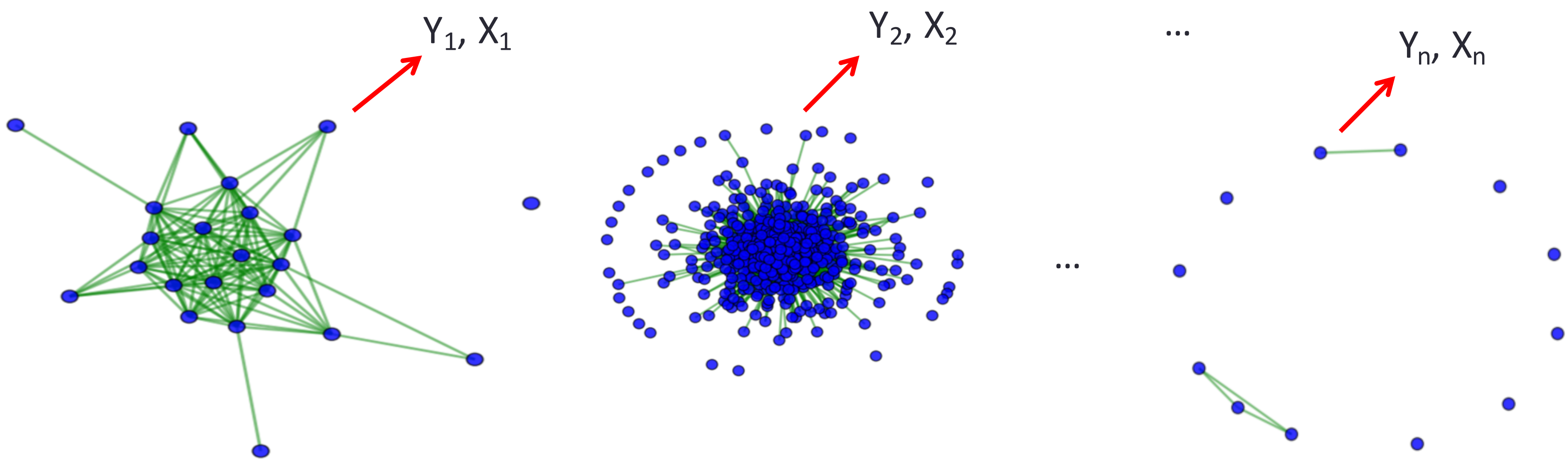
III. What factors affect the structure of a network and the formation/dissolution of relationships?
   - Examples: similarity of personal characteristics or institutional training (homophily), reinforcement of relationships (e.g., triadic closure)

Theoretical framework and scope of ongoing network research at Dartmouth



Example of Problem Type I: Multiple Networks

- Use network science and social network statistical methods to determine whether physician networks and network positions of physicians or hospitals within them are determinants of health outcome variables
- For each network or actor within, generate summary measures of their features that are used as predictors of health utilization variables

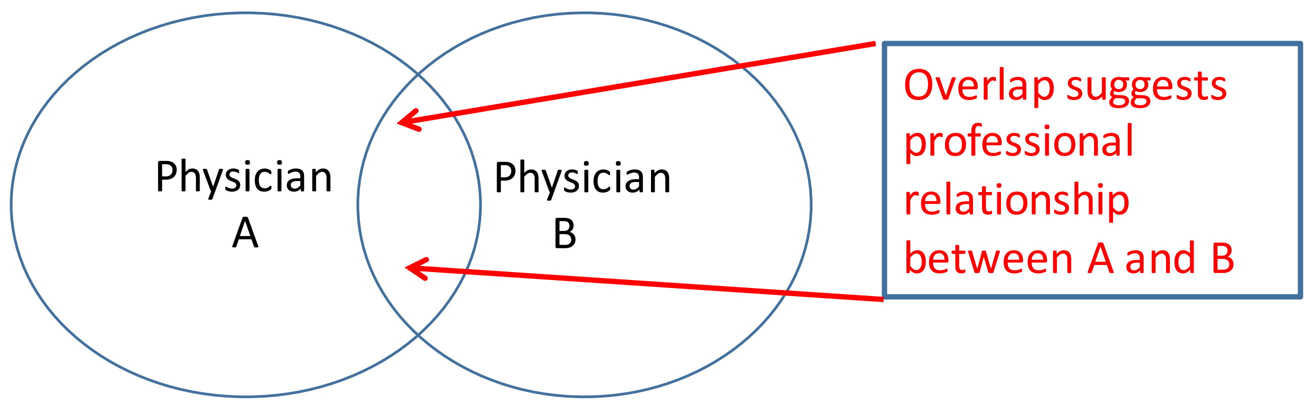


Measurement of the **physician network**

Method	Pros	Cons
Survey	Direct measure	Time intensive Generalizability issues Health topic-specific
RFID tags	Direct measure (interaction/contact)	Participation decline over study period Extensive staff commitment Battery life issues Generalizability issues
Administrative data on patient-sharing	Inclusive of all physicians Data already exist	Indirect measure of relationship

Measurement of the network of physician professional ties from claims data

Patients' medical visits during a period of time:



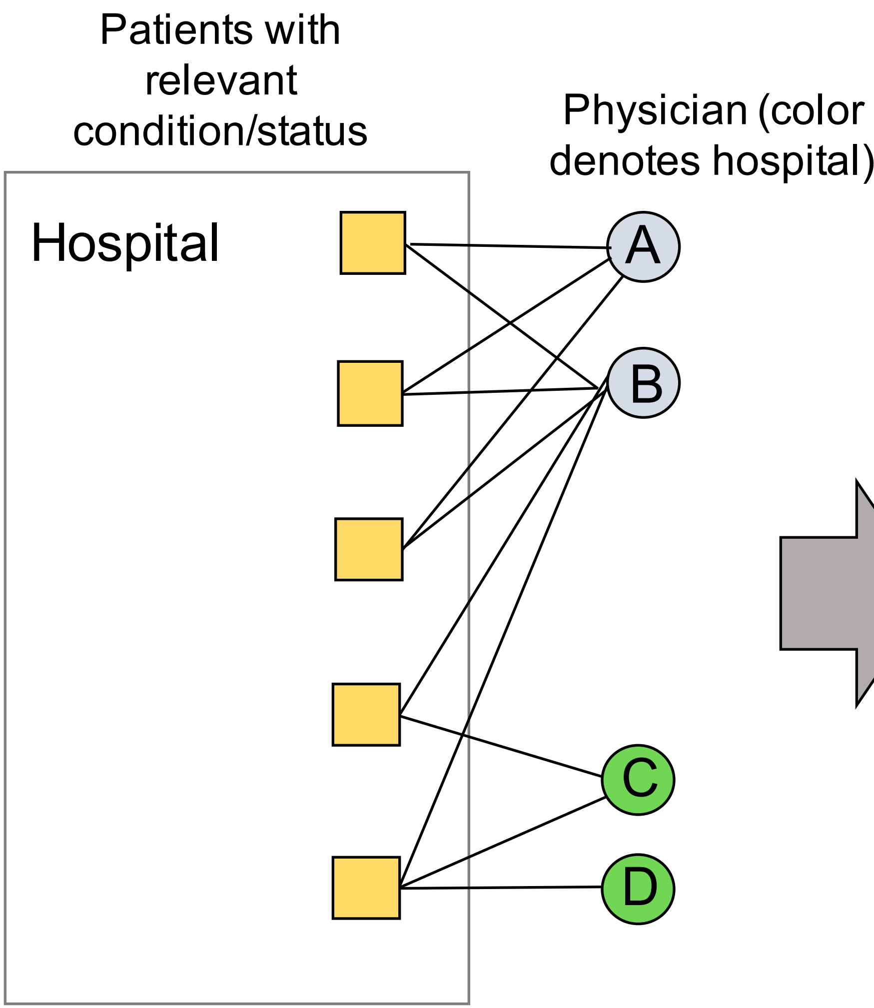
Patient – physician complexities

- Multiple encounters
- Different medical reasons for encounters
- Varying importance of encounters

Physician – physician complexities

- Multiple overlapping patients
- Different patient medical conditions
- Different level of care requirements across patients

Construction of network from Medicare claims

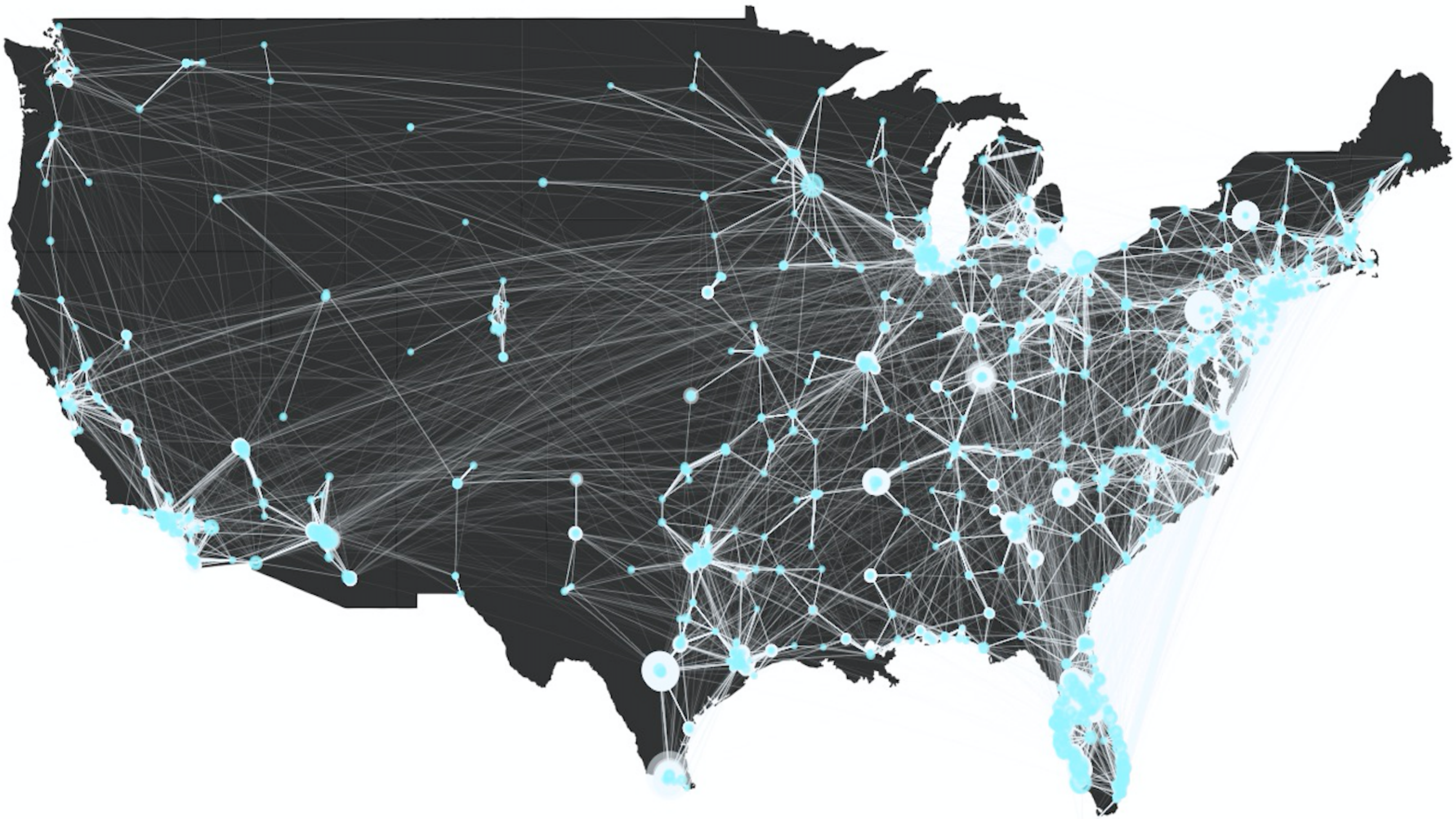


Adjacency matrix of weighted edges between physicians (shared patients)

	A	B	C	D
A	0	3	0	0
B	3	0	2	1
C	0	2	0	1
D	0	1	1	0

Sum 2×2 quadrants to get network with hospitals as nodes

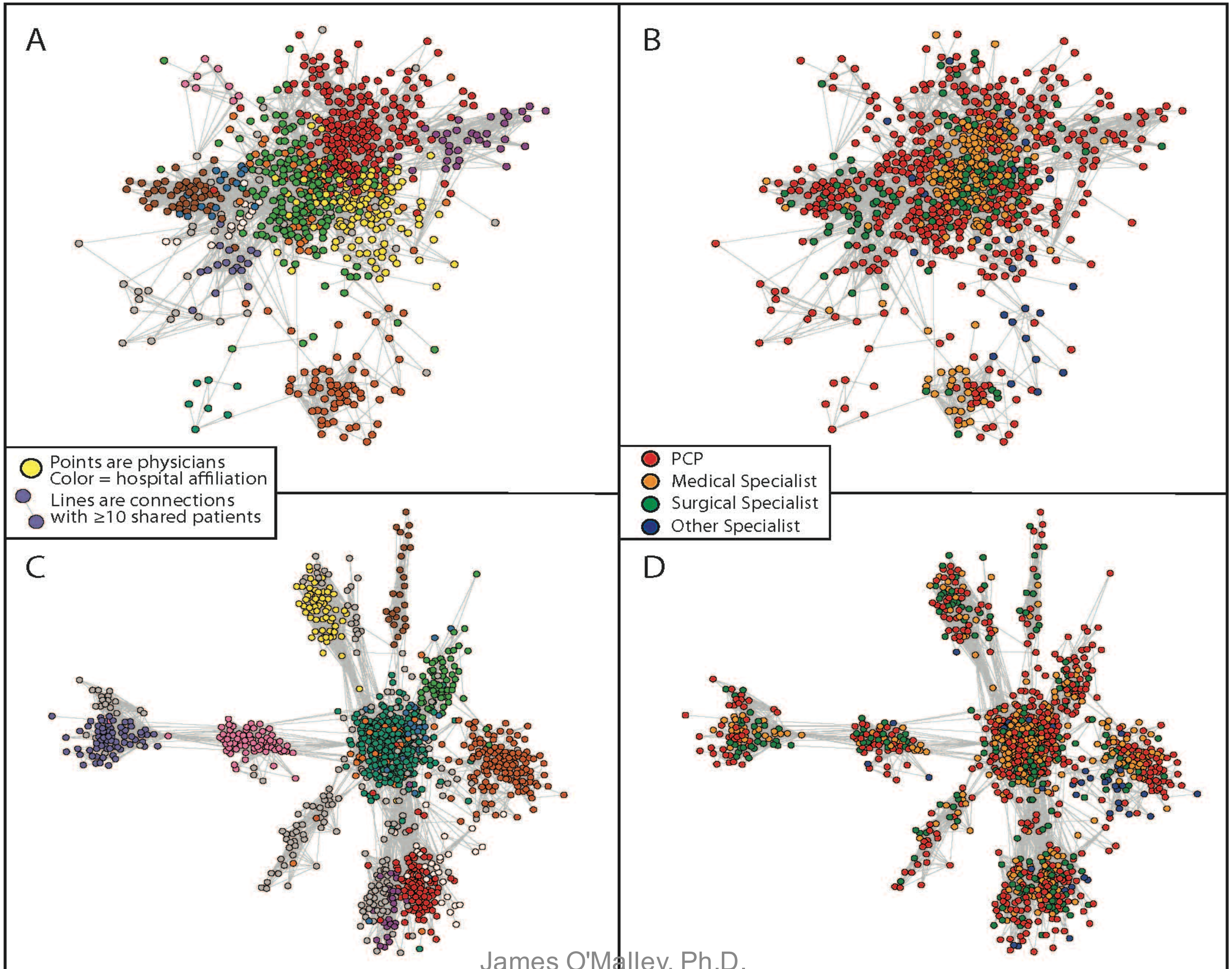
Hospital-Level Social Network in 2011 (Only Top 25% Degree Hospitals Shown)



Node size corresponds to hospital's degree
Edge thickness reflects shared patient care

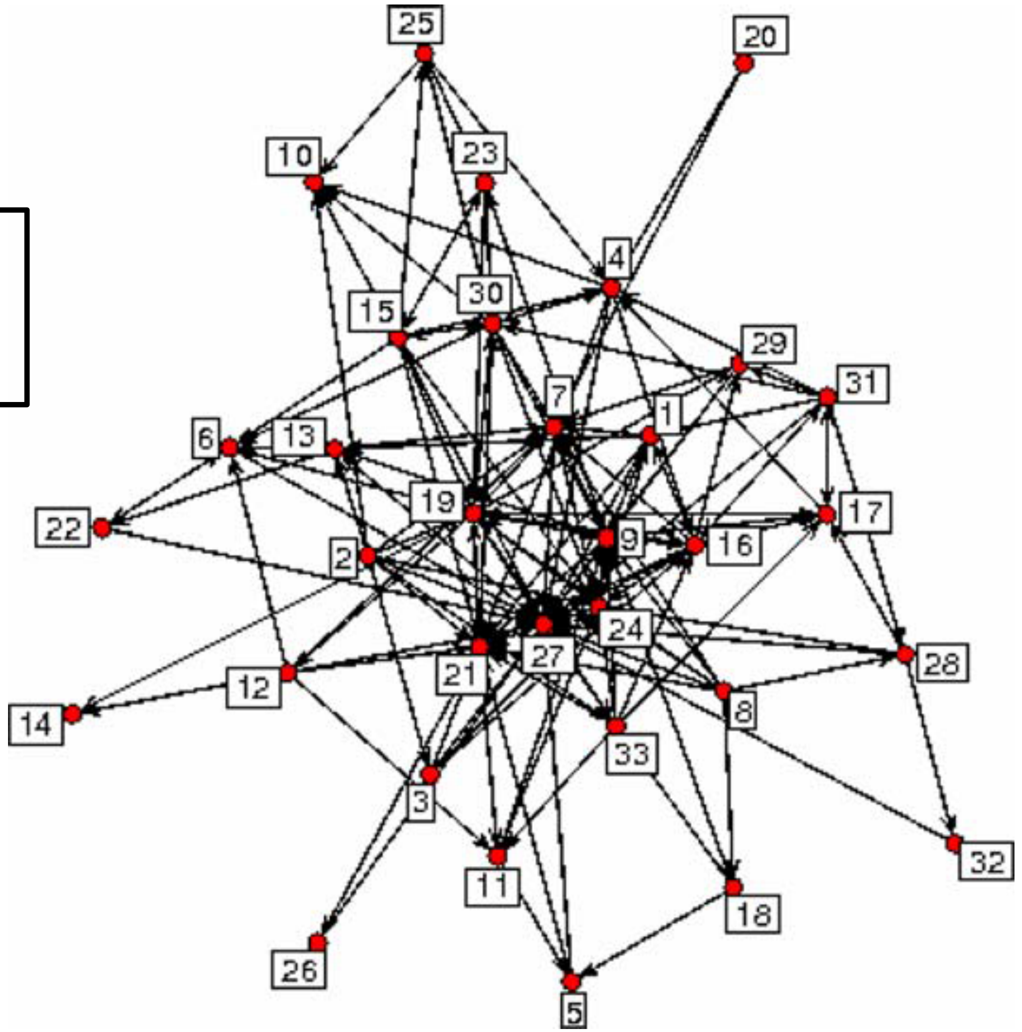
Prepared by
Erika Moen

Networks of Physicians in 4 Health Referral Regions (Landon et al 2012, JAMA)



Social network of physicians in a Boston health clinic

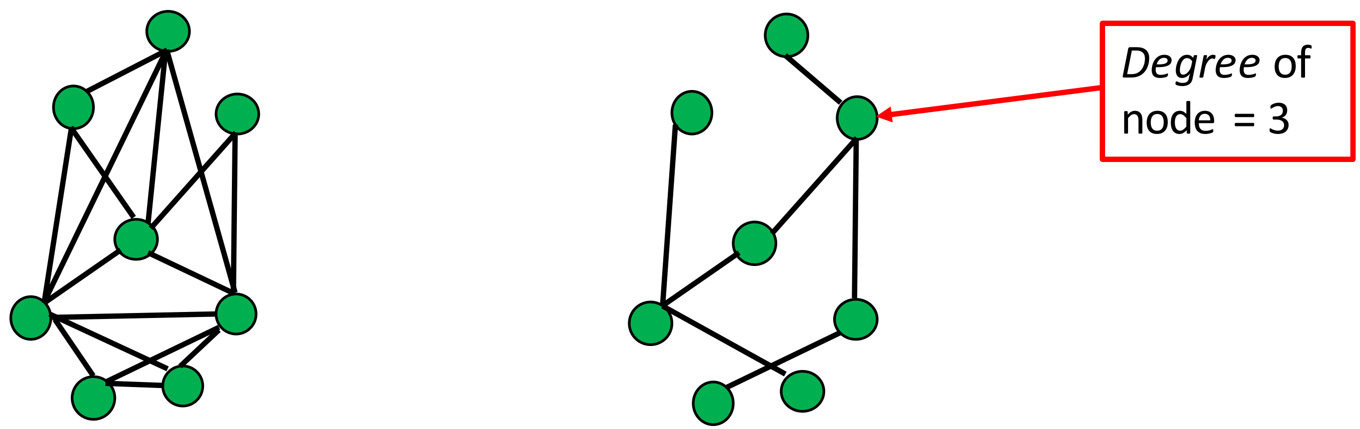
Spring embedder algorithm determines positions of actors (Fruchterman and Reingold 1991)



Keating et al (2007), O'Malley and Marsden (2008). Ties reflect influential discussions among physicians

Summarization features: Density and Degree

Density – number of edges divided by the maximum possible number of edges
= *Degree* averaged across nodes



Density = 16/28 = 0.57

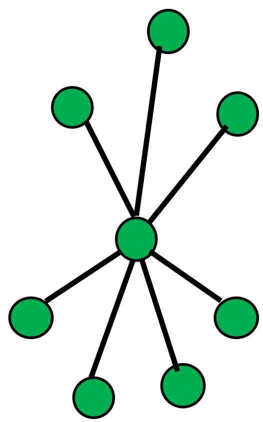
Density = 7/28 = 0.25

Hospital density and variation in care

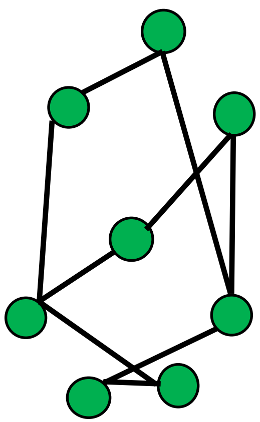
- Hospitals with a higher density of physician ties have higher costs and more intensive care (Barnett et al, 2012)

Centralization

Centralization – the extent to which there is a subgroup of highly central actors in the network



High centralization

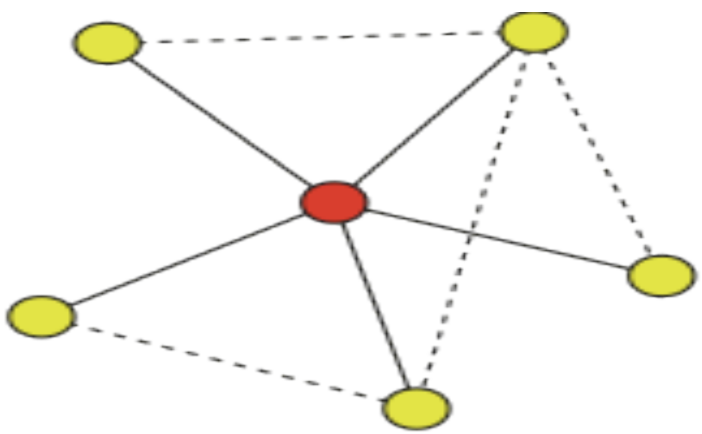


Low centralization

Hospital centralization and variation in care: We have found greater centralization associated with greater utilization

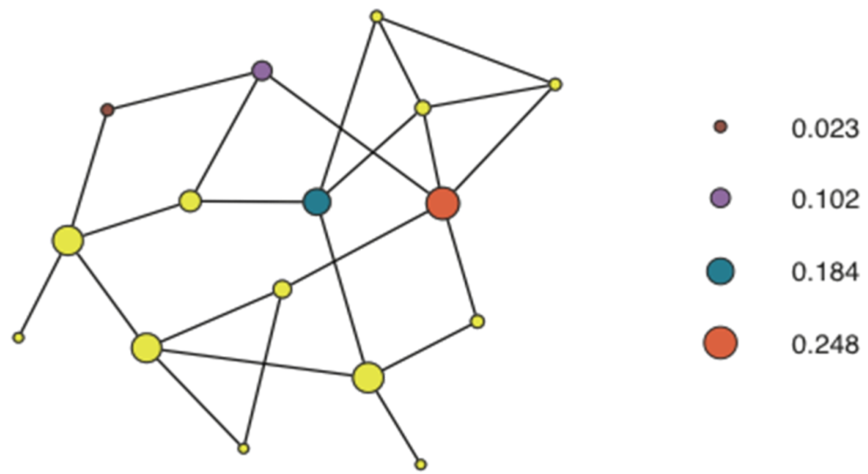
Network positional summarization examples

Clustering coefficient



Fraction of connections among neighbors of given actor (e.g., physician). In this example Clust. Coef. = 4/10

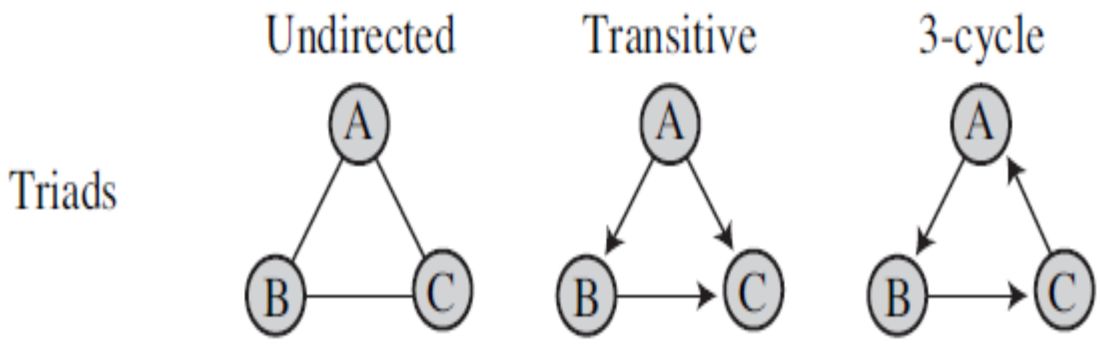
Betweenness centrality



Fraction of geodesic (shortest) paths between other actors (e.g., physicians) that pass through given actor; bigger = more central actor

Hospitals with high centrality of primary care physicians have lower costs and care intensity (Barnett et al, 2012)

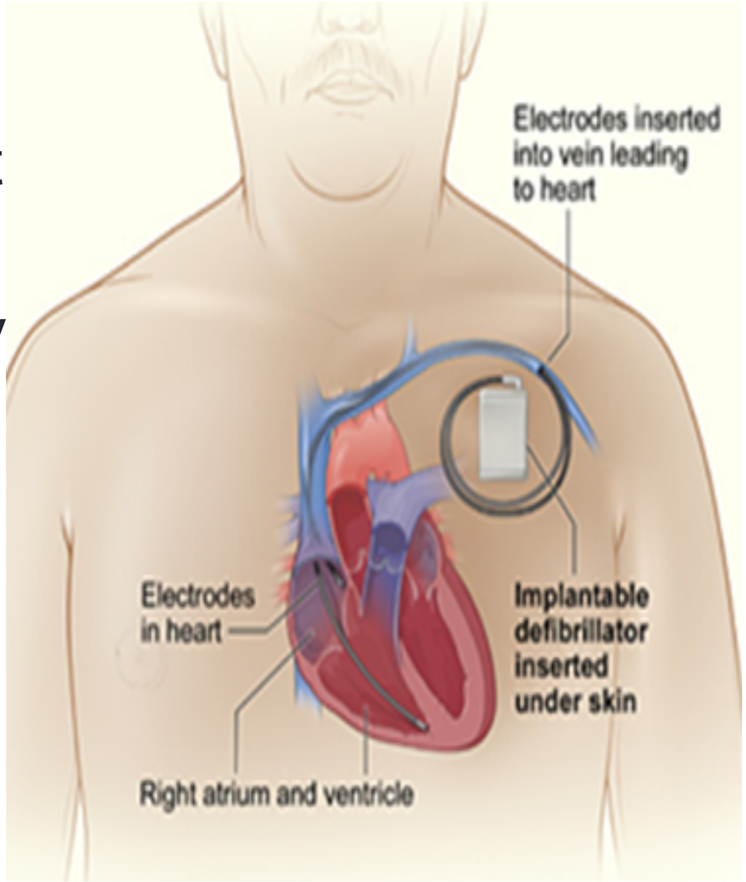
Transitivity (“A friend of a friend is a friend”)



- Sociologists → Triads are an important building block of society
- Triadic clustering is a special form of clustering
- Undirected network: the count of triangles is the basis of transitivity
- Directed network:
  - $4^3 = 64$ states of a triad
  - 16 triad groups that are non-isomorphic; embody multiple sociological constructs
  - The transitive triad (shown above for actor A) is perhaps of greatest interest

Case Study: Implantable Cardiac Defibrillators (ICDs)

- **ICDs** use electrical pulses or shocks to control potentially life-threatening ventricular arrhythmias in patients with heart failure
- Surgery is primarily performed by electrophysiologists, cardiologists, and thoracic surgeons
- Disagreement on appropriateness; therapeutic benefit versus quality of life
- Benefits depend on patient characteristics
- High cost of device



ICD therapy guidelines

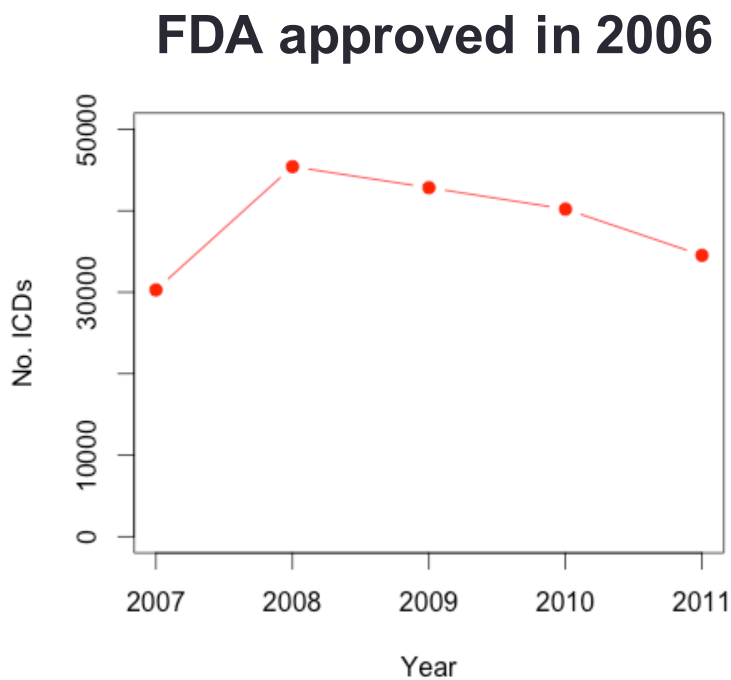
Clinical Guidelines ≠ Clinical Practice

1. Ejection fraction <=35%
2. Patient's symptoms are NYHA Class II or III
3. At least 40 days post myocardial infarction

A retrospective cohort study found that 22.5% of patients who received ICD therapy do not meet clinical guidelines.

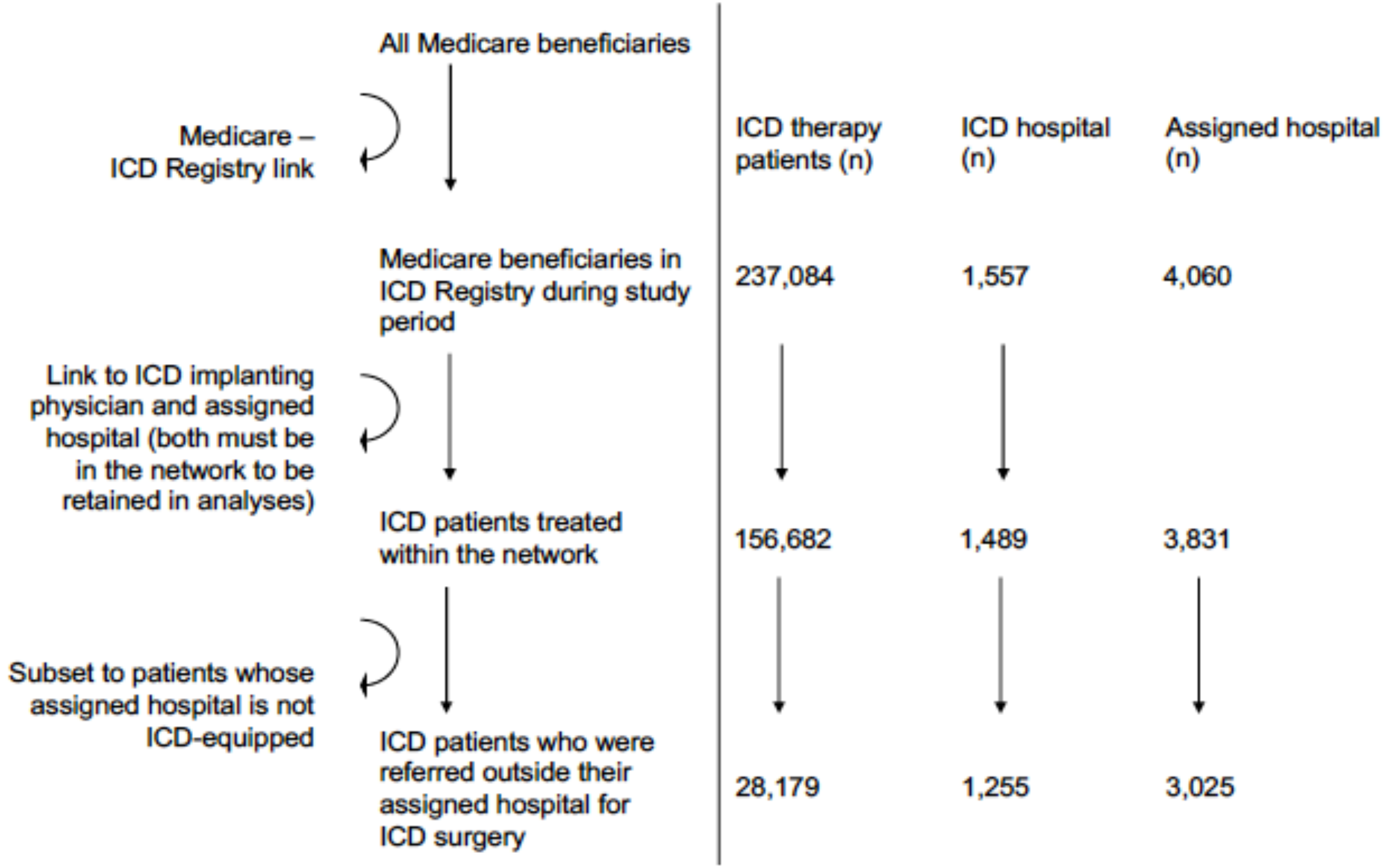
(Al-Khatib *et al*. JAMA, 2011)

Example 1: Guideline consistency of ICD utilization



- **Question**: Are the within-hospital network importance of the implanting surgeon or the importance of the referring or implanting hospitals in the US hospital network associated with ICD guideline consistency?
- Network importance measured by degree (but lots of alternatives)
- Outcome measures overuse, but not underuse, of ICD therapy
- Research focuses on patients at hospitals not equipped to perform ICDs as care more likely to be dependent on between-hospital ties for referral

Patient cohort for **outcome analyses**



Outcome data development is distinct from development of physician and hospital networks

Hierarchical model with network **positional** measures of 3 sources of clinical influence as predictors

- Patient-level logistic regression
- $i = \text{HRR}, j = \text{provider}, k = \text{patient}$
- Five types of network variables
- Lag network-based predictors by a year

$$\text{logit}(E[\text{InGuide}_{ijk}|\theta_i, \delta_{ij}]) = \beta_0 + \beta_1\text{Covariates}_{ijk} + \beta_2\text{ProvPos}_{ij}$$
$$+\beta_3\text{ReferralHospPos}_i + \beta_4\text{ReferralHospStructure}_i$$
$$+\beta_5\text{SurgeryHospPos}_i + \beta_6\text{SurgeryHospStructure}_\text{i} + \theta_i + \delta_{ij}$$
$$\text{where } \theta_i \sim \text{Normal}(0, \sigma^2) \text{ and } \delta_{ij} \sim \text{Normal}(0, \tau^2)$$

Results: ICD implanting physician

	Concurrent year analyses		Lagged year analyses	
	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
ICD implanter (physician)				
Betweenness centrality	0.81 (0.18, 3.65)	0.785	2.25 (0.39, 13.16)	0.367
Cardiologist	1.77 (1.37, 2.29)	<0.001	1.67 (1.24, 2.25)	<0.001
Clinical trial count	0.99 (0.95, 1.04)	0.685	1.01 (0.96, 1.07)	0.623
Publication count category	referent		referent	
None	1.05 (0.91, 1.22)	0.495	1.00 (0.85, 1.19)	0.981
Low (1-24)	1.09 (0.89, 1.33)	0.408	1.06 (0.96, 1.07)	0.638
High (>25)				

Moen EL, *et al*. Assessing variation in implantable cardioverter defibrillator therapy guideline adherence with physician and hospital patient-sharing networks. Medical Care, 2018; 56(4):350-7

Results: Patient's referring hospital

	Concurrent year analyses		Lagged year analyses	
	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
Referring hospital				
Degree	0.49 (0.25, 0.96)	0.037	0.61 (0.32, 1.16)	0.131
Betweenness centrality	1.14 (1.00, 1.30)	0.056	1.13 (0.97, 1.31)	0.106
Urbanicity				
Urban	referent		referent	
Large town	1.00 (0.85, 1.16)	0.961	0.98 (0.82, 1.16)	0.787
Small town	0.92 (0.76, 1.11)	0.359	0.90 (0.73, 1.11)	0.331
Teaching status				
Teaching	0.84 (0.62, 1.13)	0.245	0.99 (0.72, 1.35)	0.644

Moen EL, *et al*. Assessing variation in implantable cardioverter defibrillator therapy guideline adherence with physician and hospital patient-sharing networks. Medical Care, 2018; 56(4):350-7

Results: ICD surgery hospital

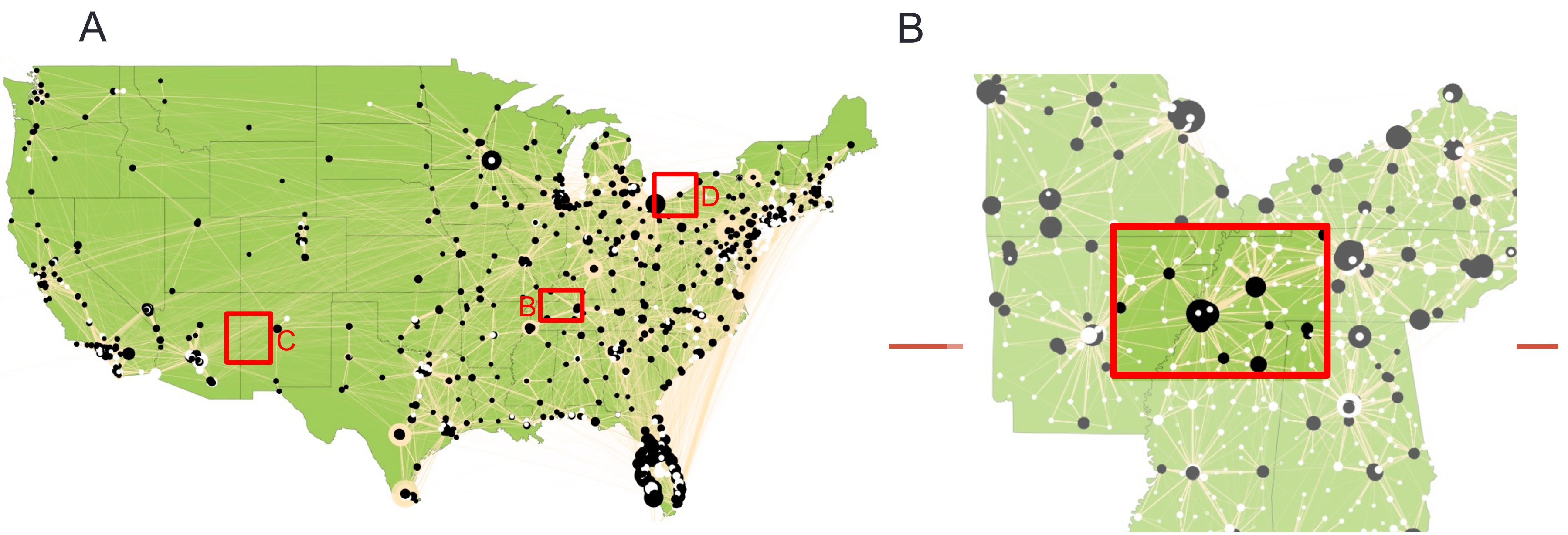
	Concurrent year analyses		Lagged year analyses	
	Odds ratio (95% CI)	p-value	Odds ratio (95% CI)	p-value
ICD surgery hospital				
Degree	1.61 (0.98, 2.64)	0.059	1.67 (0.97, 2.89)	0.064
Betweenness centrality	0.94 (0.89, 1.00)	0.067	0.93 (0.87, 1.00)	0.049
Urbanicity				
Urban	referent		referent	
Large town	1.08 (0.76, 1.55)	0.660	1.07 (0.71, 1.60)	0.756
Small town	1.44 (0.37, 5.67)	0.602	1.24 (0.30, 5.08)	0.764
Teaching status				
Teaching	1.06 (0.88, 1.28)	0.567	1.05 (0.85, 1.30)	0.644

Moen EL, *et al*. Assessing variation in implantable cardioverter defibrillator therapy guideline adherence with physician and hospital patient-sharing networks. Medical Care, 2018; 56(4):350-7

Discussion: Regionalization

- The connectedness of hospitals involved in the referral was associated with guideline adherence
  - Patients were more likely to meet guidelines if:
    - Their assigned hospital had fewer connections to other hospitals
    - Their ICD surgery hospital had more connections
- Regionalization of specialized ICD services may promote adherence to guidelines
  - If referring hospitals have fewer connections (enforcing existing information/referral paths) this could lead to more efficient relationships, improved communication/learning, and thus increased adherence to guidelines

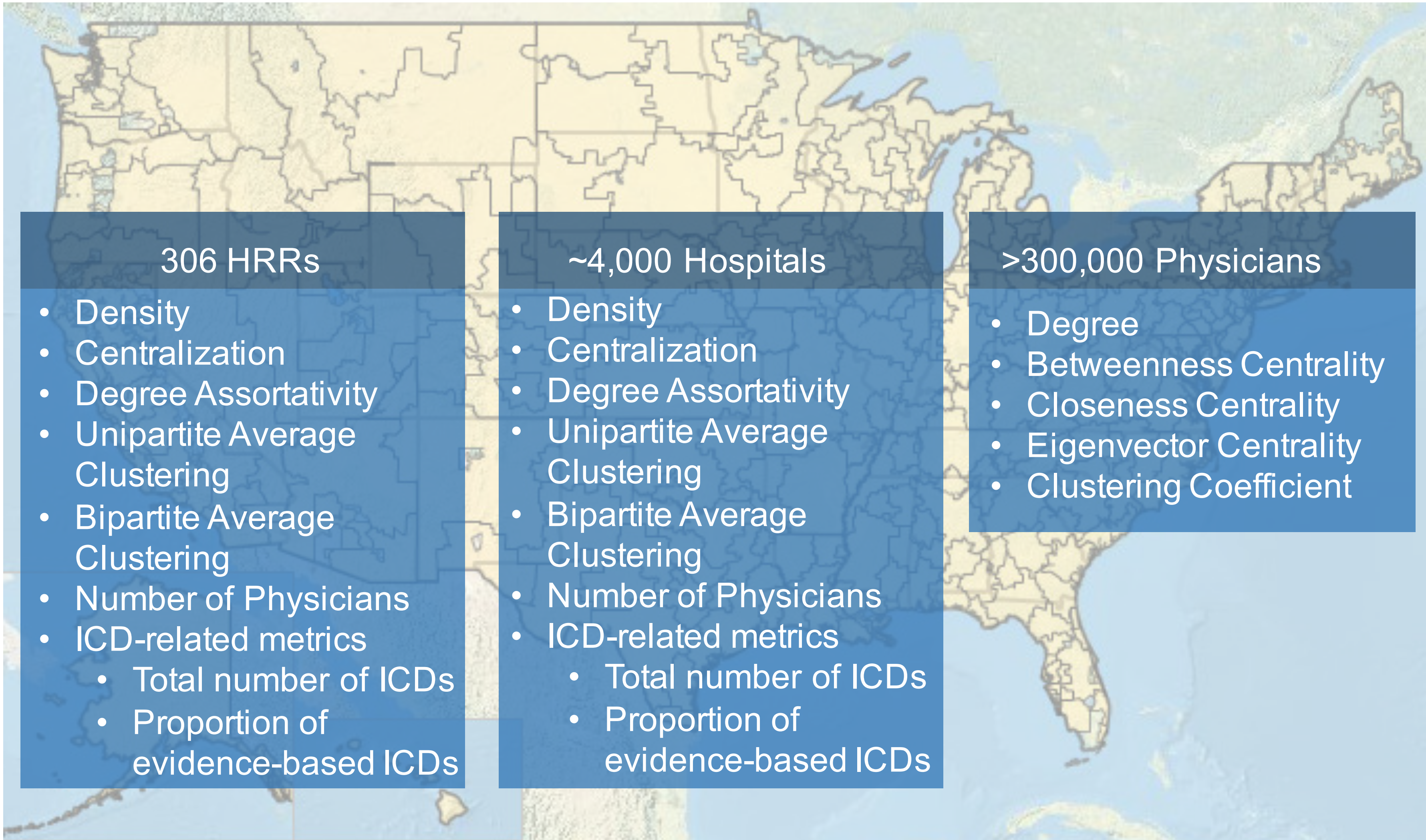
Visualization of “Regionalization”



- Referring hospitals have fewer connections!!!

Prepared by Erika Moen

Many other network measures at multiple levels of aggregation!



306 HRRs

- Density
- Centralization
- Degree Assortativity
- Unipartite Average Clustering
- Bipartite Average Clustering
- Number of Physicians
- ICD-related metrics
  - Total number of ICDs
  - Proportion of evidence-based ICDs

~4,000 Hospitals

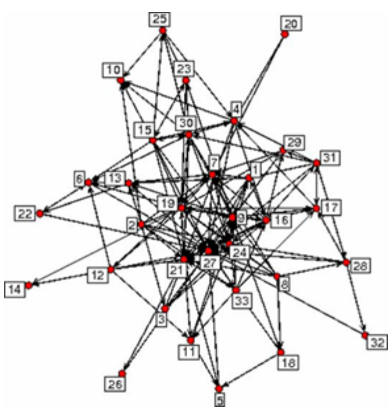
- Density
- Centralization
- Degree Assortativity
- Unipartite Average Clustering
- Bipartite Average Clustering
- Number of Physicians
- ICD-related metrics
  - Total number of ICDs
  - Proportion of evidence-based ICDs

>300,000 Physicians

- Degree
- Betweenness Centrality
- Closeness Centrality
- Eigenvector Centrality
- Clustering Coefficient

Problem type 2: Social influence analysis

- Network defines predictors
  - Do physicians or hospitals influence one another (``social influence'')
  - Example, adoption of a new medical technology
- Endogenous peer effects
  - Does the behavior of peer physician or hospital affect the focal physician's or hospital's behavior?



* $Y, X = Y_{peer}, X_{peer}$

Exogenous peer effects Does the treatment received by my peers affect my outcome (above and beyond my treatment)?

Interest in studying peer effects in networks

The NEW ENGLAND JOURNAL of MEDICINE

SPECIAL ARTICLE

The Spread of Obesity in a Large Social Network Over 32 Years

Nicholas A. Christakis, M.D., Ph.D., M.P.H., and James H. Fowler, Ph.D.

They used a unique and controversial identification strategy!

Why Estimate Peer Effects?

1. Justify budget-limited interventions to stop spread of bad practices in ICD utilization
   - Intervene (e.g., educate) a fraction of physicians or hospitals
   - Ideally, target hospitals strategically positioned to have the greatest influence on other hospitals

2. Evaluate full effect of an intervention
   - Peer effects measure extent that end up intervening on the untreated
   - Account for spillover effects (Sobel 2006)
   - “Collateral effects” (Christakis 2004)

Micro-level ("Peer-Effect") Diffusion: focal physician behavior regressed on that of their peers

- Example 1: Physician peer-to-peer influence of ICD utilization over time
  - Peer effects: core, elementary form of diffusion
  - Is there evidence of hospital-hospital influence on ICD capability adoption
  - If so, is modified by structural position in network?
    → Justify selecting certain physicians for limited-budget interventions
  - O'Malley, Moen, Bynum, Austin, Skinner (submitted)

- Example 2: Peer effect of another physician's patient having an adverse reaction following a colonoscopy (beyond effect of adverse reactions within own patient cohort)
  - With Keating, Landon, and Onnela (Submitted)
  - Not discussed today

Regression of ICD equipped status

- Let y_{it} denote ICD status (1 = equipped, 0 = not-equipped) at time t
- Key predictor is the prior year weighted average of y_{it} over the peer hospitals of hospital *i*
- We used the network strength (number of shared patients) of the edges between the hospitals as weights, W
- Thus, model has the form

$$y_{it} \mid y_{i(t-1)} = j \sim Bernoulli(p_{it}(j))$$

where $logit(p_{it}(j))$

$$= \theta_{ij} + \beta_{1j}\, x_{i(t-1)} + \beta_{2j} [W_{t-1} Y_{t-1}]_i$$

and $\theta_{ij} \sim Normal(\beta_{0j}, \tau_j^2)$ is a random effect for hospital

and x is a vector of control predictors

Regression of ICD equipped status: Add Network Positional Variables and geographic control

- Full model interacts the weighted average *WY* with hospital i's network strength
- Model given by: $logit(p_{it}(j))$
$$= \theta_{ij} + \beta_{1j}\, x_{i(t-1)} + \beta_{2j}[W_{t-1}Y_{t-1}]_i + \beta_{3j}[G_{t-1}Y_{t-1}]_i$$
$$+ \ (\beta_{4j}[W_{t-1}Y_{t-1}]_i + \beta_{5j}[G_{t-1}Y_{t-1}]_i)d_{i(t-1)}$$
where $d_{i(t-1)}$ is the network weighted degree (strength) of physician i at time *t-1*
- G is a weight matrix based on geodesic distances
- In future work, we may add additional ego and peer variables for number of implants and referrals to account for inertia and the extent of the implanting or referring

ICD Adoption of Equipped Status

ICD Adoption: 306 hrrs, 3720 hospitals, 12716 observations			
Term	Estimate	z-value	p-value
Lag network strength	-1.593	-2.67	0.008
Lag peer equipped	-0.391	-1.29	0.198
Lag peer equipped*network strength	2.295	2.81	0.005
Lag peer referral	0.268	0.58	0.564
Lag peer implant	-0.146	-0.58	0.561
Lag geographic equipped	22.750	4.67	0.000
Lag geographic referral	-0.607	-4.30	0.000
Lag geographic implant	-0.059	-1.27	0.204
Var(hospital, HRR)	1.15 +/- 1.07, 0.49 +/- 0.70		

A non-capable hospital with strong connections to peer hospitals who have the capability to implant ICDs is more likely to acquire the capability to implant ICDs

ICD Equipped Continuation

ICD De-adoption: 305 hrrs, 1410 hospitals, 4418 observations			
Term	Estimate	z-value	p-value
Lag network strength	-1.670	-3.01	0.003
Lag peer equipped	-1.456	-3.27	0.001
Lag peer equipped*network strength	2.216	2.75	0.006
Lag peer referral	-0.055	-0.07	0.943
Lag peer implant	-0.322	-1.14	0.254
Lag geographic equipped	-4.753	-1.15	0.248
Lag geographic referral	-0.042	-0.29	0.773
Lag geographic implant	-0.029	-0.78	0.433
Var(hospital, HRR)	0.84 +/- 0.92, 0.00 +/- 0.00		

An ICD capable hospital with strong connections to to peer hospitals who have the capability to implant ICDs is more likely to remain ICD capable

Causality Concerns

- Homophily: “Birds of a feather flock together”
  - Individuals with similar behaviors more likely to become friends
  - Physicians who train together have similar treatment preferences and more likely to subsequently work together?
  - Tie-dissolution due to diverging viewpoints or attitudes over time

- Unmeasured common causes
  - Unknown peer physicians
  - Regional activities (e.g., marketing campaign)
  - Exposure to marketing or the same supplier of free medical products

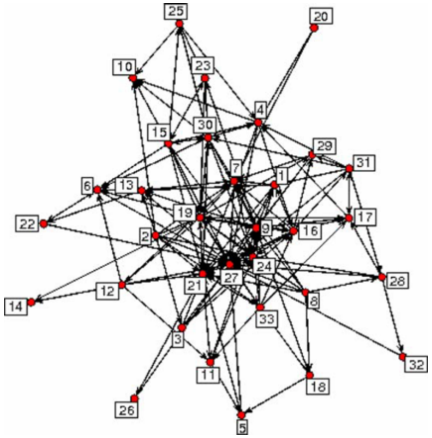
Social influence analysis causal challenges

- Overlapping groups of individuals yield the predictor(s) of individuals' outcomes!
  - Reflection problem (Manski, 1993)
- Statistical analysis challenging if seek causal claim when network not formed at random!
  - Complicated simultaneous equations model can be used but makes strong assumptions
- Longitudinal data helps with identification of causal effects
  - Reverse causality, simultaneity, …
  - Avoids reliance on strong parametric assumptions

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

Problem Type 3: Analysis of network structure “Extra Material”

- Observed network is the outcome
- Often only observe network once (cross-sectional data)
  - Longitudinal data now becoming more common
- Are global network properties explained by local configurations or sub-networks?
  - Closed dyads: reciprocity
  - Closed triads: transitivity, 3-cycles, …
- Are individuals with particular characteristics more likely to form ties (homophily, assortative mixing, social selection)?
- Do (latent) communities underlie the network?
- Non-standard and challenging statistical analyses required!

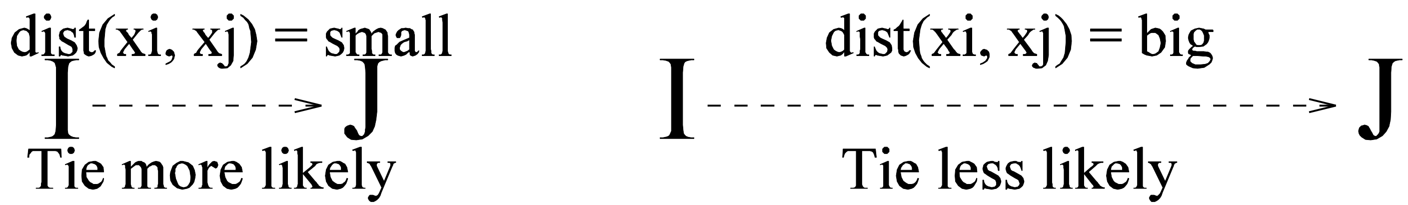


Why Model Relationships?

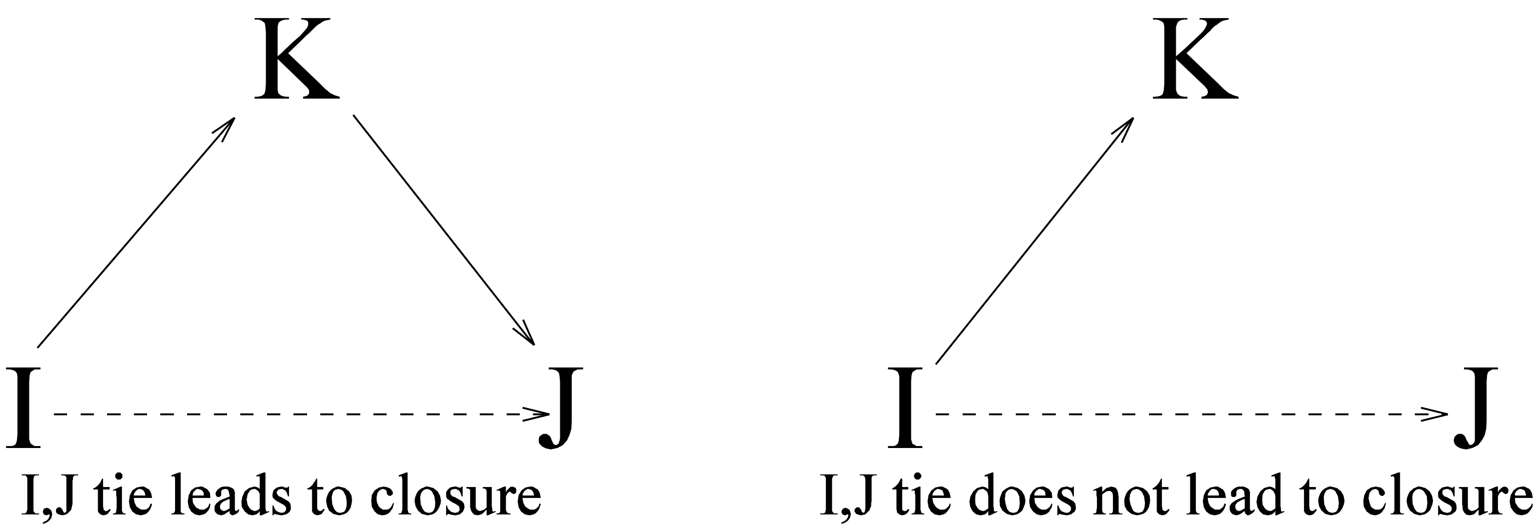
- Recipe for manipulating the influences to which an individual is exposed
  - Determine factors that reinforce relationships
- Find optimal position in the network to identify actors for which intervention will have maximal impact
  - Optimize interventions on physicians, hospitals, health systems, regions
- Gain insight in how to manipulate health organization into more favorable forms
  - Identify key elements of network structure of the best Accountable Care Organization (ACO) and replicate them!

Some Key Sociological Relationships

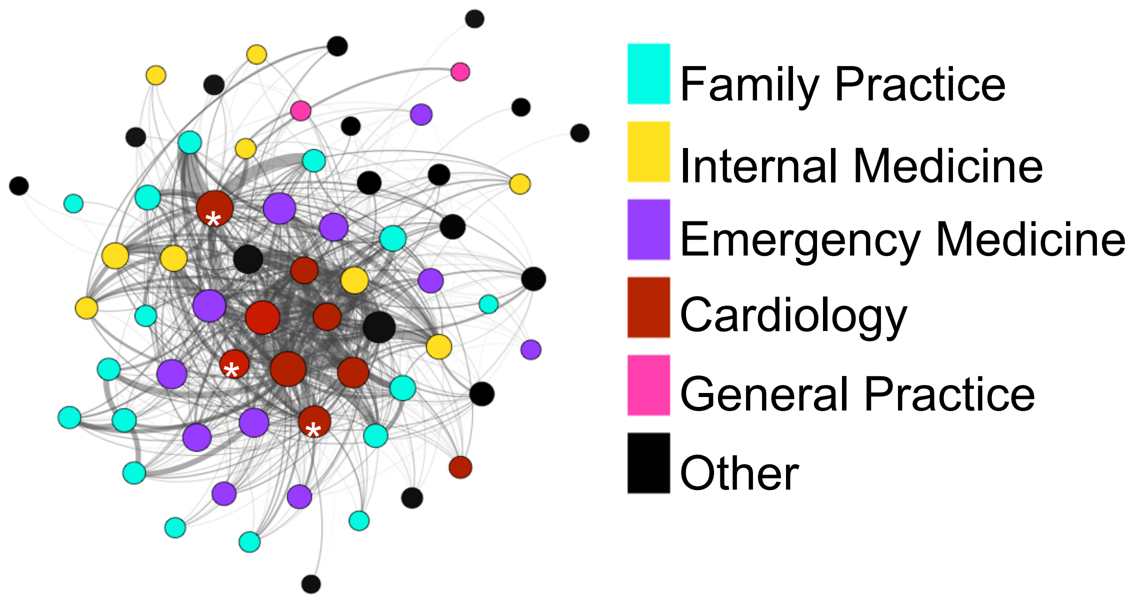
Homophily



Triadic Closure (Transitivity)



Simple Example: Estimate effects of physician homophily on the relationships between physicians within a hospital



*ICD Provider

Exponential family random graph model (ERGM) of hospital network:

Term	Coefficient
Edge (overall density of ties)	-0.4238
Homophily by specialty	
Cardiology	3.42
Family Practice	-1.39
Internal Medicine	0.20

Change in log-odds of the tie if the physicians are both cardiologists compared to if they have different specialties, conditional on the rest of the network

Beyond Dyadic Independence Models

- Assuming dyadic independence (as in prior slide) allows model for the network to be generated from the model for the dyad
  - Allows logistic regression estimation to be used!
- Dependence between dyads arises whenever the state of one dyad depends on the state of another dyad over and above actor-specific effects
  - Triadic dependence: an edge is more (or less) likely to form if its actors have a common third actor
  - Cannot multiply probability distributions of dyads to generate model for the network!

  → Forced to model whole network simultaneously!

Dartmouth Group References

- 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(3), 132-153
- Moen EL, Bynum JP, Austin AM, Skinner JS, Chakraborti G, O'Malley AJ. Assessing variation in implantable cardioverter defibrillator therapy guideline adherence with physician and hospital patient-sharing networks. Accepted for publication in Medical Care
- An C, Rockmore D, O'Malley AJ: Nationwide Physician Networks determined from CMS's publically available Medicare data on physicians' overlap of care for the same patients. Accepted for publication in Statistics in Medicine
- Modeling Peer Effect Modification by Network Structural Prominence: The Diffusion of Implantable Cardioverter Defibrillators. O'Malley AJ, Moen EL, Bynum JPW, Austin A, Skinner JS. Under review

Harvard Group References

- Landon BE, Onnela JP, Keating NL, Barnett ML, Paul S, O'Malley AJ, Keegan T, Christakis NA. Using Administrative Data to Identify Naturally Occurring Networks of Physicians. Medical Care, 2013, 51 (8), 715-721. PMID: 23807593.
- Landon BE, Keating NL, Barnett ML, Onnela JP, Paul S, O'Malley AJ, Keegan T, Christakis NA. Variation in Patient-Sharing Networks of Physicians Across the United States. Journal of the American Medical Association, 2012, 308, 265-73. PMID: 22797644
- Barnett ML, Christakis NA, O'Malley AJ, Onnela J-P, Keating NL, Landon BE. Physician Patient-Sharing Networks and the Cost and Intensity of Care in US Hospitals. Medical Care 2012, 50, 152-60. PMID: 22249922.
- Barnett ML, Landon BE, O'Malley AJ, Keating NL, Christakis NA. Mapping physician networks with self-reported and administrative data. Health Services Research 2011, 46, 1592-1609. PMID: 21521213

Methodological References

- O'Malley AJ, Marsden PV. The Analysis of Social Networks. Health Services and Outcomes Research Methodology 2008, 8, 222-269
- 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
- O'Malley AJ, Paul S (2014). Using Retrospective Sampling to Estimate Models of Relationship Status in Large Longitudinal Social Networks. Computational Statistics and Data Analysis, 82, 35–46.
- Paul S, O'Malley A.J. (2013). Hierarchical longitudinal models of relationships in social networks. Journal of the Royal Statistical Society, Series C.
- 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
- Under revision: Viles WD, O'Malley AJ. External Information in Community Detection