Using Graphs to Characterize Nationwide Physician Referral Networks

Ding Tong
Yale University, ding.tong@yale.edu

Shu-Xia Li
Yale University, shu-xia.li@yale.edu

Isuru Ranasinghe
Yale University, isuru.ranasinghe@yale.edu

Sudhakar Nuti
Yale University, sudhakar.nuti@yale.edu

Hongyu Zhao
Yale University, hongyu.zhao@yale.edu

See next page for additional authors

Presenter/Creator Information

Ding Tong, Shu-Xia Li, Isuru Ranasinghe, Sudhakar Nuti, Hongyu Zhao, and Harlan Krumholz

Big Data @ Yale: Using Graphs to Characterize Nationwide Physician Referral Networks

Ding Tong, B.S.,[1,2] Shu-Xia Li, PhD,[2] Isuru Ranasinghe, PhD,[2] Sudhakar Nuti, B.A.,[2] Hongyu Zhao, PhD,[1] Harlan M. Krumholz, MD, SM[1,2,3]

[1]Department of Biostatistics, Yale School of Public Health; [2]Yale-New Haven Hospital Center for Outcomes Research and Evaluation; [3]Yale School of Medicine.

Use Case Background

Treatment outcomes such as quality of care vary across hospitals. The drivers of such variations are not well understood. We hypothesize that the referral pattern between providers and the hospitals, can influence outcomes such as 30-day readmission. As a first step, we sought to use graph-based approaches to characterizing nationwide physician referral networks to better understand interactions among physicians. Our ultimate goal is to assess the association between hospital outcomes and their referral network characteristics.

Key Question

How to characterize local physician referral networks across the United States?

Novelty/Innovation

1. Applied graph-based evaluation approaches on nationwide physician referral network.
2. Developed a new method to evaluate goodness of fit for the statistical network models.

Methods

1. Data Sources

Data	Source	Records #	Variables #	Variables Description
Physician Referral	CMS 2012-13 national Physician Referral data	73,071,804	5	Directional referral pair and number of occurrences from one National Provider Identifier (NPI) provider to another within 30 days of service in 2012-2013.
Physician Information	CMS	642,144	40	Physician's National Provider Identifier (NPI) and associated demographic and geographical information
Hospital Information	NPI Register	1,002,763	329	Hospital's National Provider Identifier (NPI), the demographic and geographical information, and CMS Certification Number (CCN).
Hospital Outcome	CMS	4.811	2	Hospital's CMS Certification Number (CCN) and the 30-days Hospitalwide Readmission Rates

Table.1 Data Sources and Variables.

Methods (continued)

2. Descriptive statistics of the physician referral networks at state level

For each state, we first calculated the characteristics of physicians, including total number of physicians, and numbers of physicians in different specialties, physicians' average graduation year, and gender composition. We then derived characteristics of networks, including the number of edges (referrals), average shared patients, average shared referral times, average betweenness (the number of shortest paths from all vertices to all others that pass through the node), primary care physician betweenness (average betweenness within primary care physicians) and the cluster coefficient. We visualized the primary care physician network with R software package iGraph.

3. Statistical model for local network

After deriving the overall network characteristics, we used Exponential random graph models (ERGM) to further characterize local network property and sub-networks such as a physician referral sub-network of an individual hospital.

In the ERGM model, we have:

$$P(Y=y|X)=\kappa^{-1}\exp\{\eta^t g(y,X)\},$$

where Y is considered as a random matrix taking values in the set of all n-by-n adjacency matrices G. $Y_{ij}=1$ if there is an edge from node i to node j, and $Y_{ij}=0$ otherwise. $\kappa=\kappa(\eta)$ is the normalizing constant. $g(Y,X)=\sum_{i\neq j} y_{ij}h(X_i,X_j)$, where h is a function mapping $\mathbb{R}^q\times\mathbb{R}^q$ into $\mathbb{R}^p$, is used to choose appropriate statistics.

Candidate covariates in the model include potentially clinically important factors and network property measures. Final model selection was based on statistical significance of the coefficients and akaike information criteria (AIC).

Traditional methods for assessing goodness of fit are not applicable to ERGM. To assess the goodness of fit, we first simulated network data 1000 times using the coefficients inferred from the selected model. To assess the overall goodness of fit, we compared the overall network properties between the observed one and the simulated ones visually. To further assess the local goodness of fit, we generated the predicted network from the simulated networks by using a cutoff $p_0 : \Pr[p_{ij,\forall i\neq j}\geq p_0]=n/1000$ for every dyad, where n is the total number of edges of the observed network. We present the differences between the predicted and observed networks by heatmaps organized according to covariates to detect local lack-of-fit.

Results

1. **Descriptive measures at state level** : 1) graphs and characteristics vary substantially across geographic areas, 2) graphs and characteristics show strong spatial correlations.

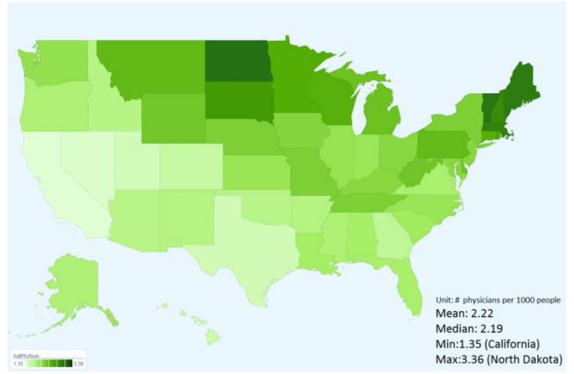

Figure.1 a) Physicians density (#physicians per 1000 people).

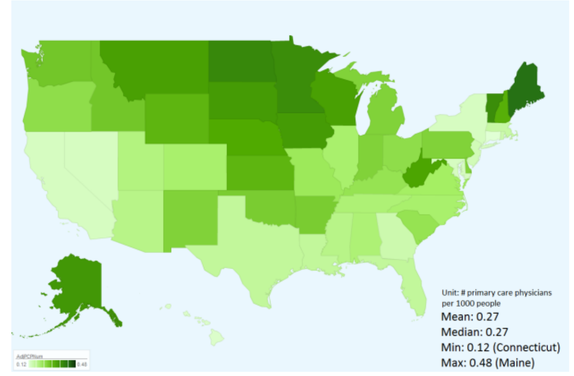

Figure.1 b) Primary health care physician density (# physicians per 1000 people)

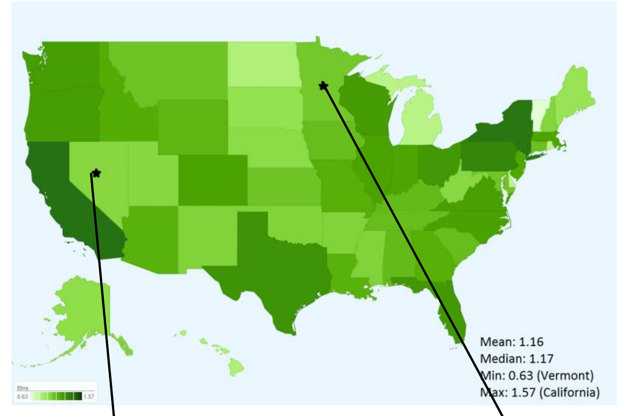

Figure.1 c) Scaled average betweeness by state (the higher the betweeness, the lighter of the network)

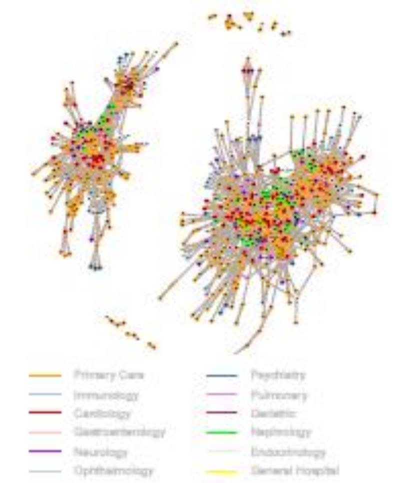

Figure.1 d)Primary Care Physician Referral Network in Nevada

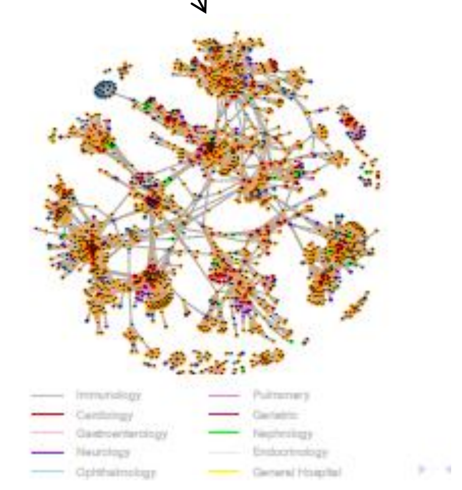

Figure.1 e)Primary Care Physician Referral Network in Minnesota

2. **Statistical model for local network:** The ERGM model shows that, in Griffin Hospital, physicians in cardiology, diagnostic radiology and geriatric medicine are more likely to send and receive referrals than physicians in primary care and internal medicine.

Variable	Estimated log-odds	p-value
Cardiologists Refer Being Referred	8.74	<1e-04
Primary Care Physician Being Referred	-1.46	0.22
Cardiologists Refer Referring Others	9.18	<1e-04
Primary Care Physician Referring Others	-2.12	0.07
Physicians within a Same City	1.78	0.0002

Table 2. Selected ERGM model output for a 27-node sub-network of Griffin Hospital

Results (continued)

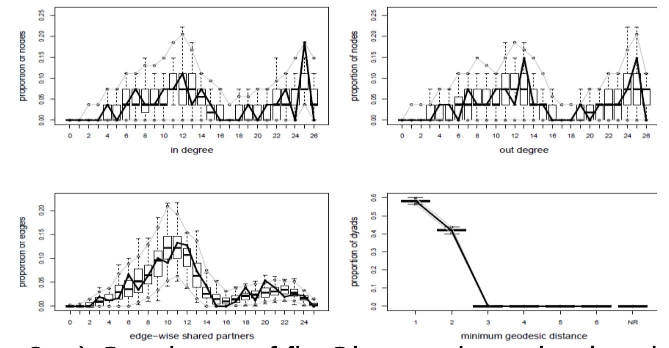

Figure.3 a) Goodness of fit: Observed vs. simulated networks regarding proportions of nodes vs. in-degree, out-degree, edge-wise shared partner and minimum geodesic distance

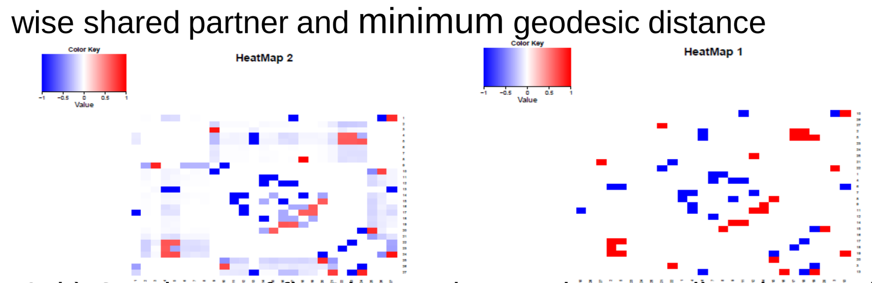

Figure.3 b) Goodness of fit: Observed network vs. predicted network by simulation; left panel, white indicates existence of edges agreed between observed and simulated, blue indicates the edges exist in observed but not simulated, red indicates otherwise; right panel shows the observed edges (value of 1 or 0) minus the empirical probability of the existence of such edges.

The goodness of fit plots show no evidence for lack-of-fit.

Discussion

Limitations

1. The descriptive network measures lack statistical rigor.
2. ERGM model can only use binary data instead of count data, and can not be easily adapted to deal with big sub-network.
3. The overall descriptive measures are at the state level but there is a need for studying smaller geographic regions.

Conclusions

Use of graph-based approaches has the ability to provide more insights in describing and evaluating the nationwide physician referral networks.

In the future, we will refine the network descriptive measures, refine algorithms of defining cluster or module of network, and overcome the computational challenge of the ERGM and further adapting it for counting outcomes. We will further study the association between these network characteristics and health care outcomes.

References

[1] Landon BE and et al. Variation in patient-sharing networks of physicians across the united states. JAMA , 308(3):265-273,2012.

[2] Hunter DR and et al. Goodness of fit of social network models. JASA, 103(481):248-258,2008