Supplemental Appendix: Optimal Physician Shared-Patient Networks and the Diffusion of Medical Technologies

A. James O'Malley[1,*], Xin Ran[2], Chuankai An[3], and Daniel Rockmore[4]

[1]*Department of Biomedical Data Science and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA*

[2]*Department of Biomedical Data Science, The Dartmouth Institute for Health Policy and Clinical Practice, and the Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA*

[3]*Research Institute of China Investment Corporation, Beijing, 100010, China*

[4]*Department of Mathematics and Department of Computer Science, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA*

Abstract

In this supplemental appendix, we present expanded descriptions of the data wrangling and additional results that were not able to be included in the main paper due to length restrictions. The results are supported by the text that would have accompanied them in the main text had space permitted.

1 Transformations to make Network Features Have Equivalent Scales

For greater variance to imply greater discriminatory power, network features must have the same scale across the projection methods. Otherwise, the scale of the edge-weights under a projection may drive heterogeneity in a network feature as opposed to true sources of variation underlying the network. For example, under binary-valued networks, density is restricted to be between 0 and 1 but for weighted networks, especially those built using the total-ordering level of *Multiplicity*, density may attain large values and thus have a large variance across the hospitals. To alleviate this concern, we scaled density for each hospital under a given projection by the ratio of the mean density across all hospitals under the base projection (the binary-valued undirected network with *Multiplicity* $= 1$ and neither *Continuation* nor *Revisit* binding) divided by the mean under the given projection. We also made centralization scale-free by dividing by the square of the average degree of the physicians in the network. Because the size of the largest connected component, the number of components, and diameter are based on the presence or absence of edges, these were not transformed. The features that are proportions or correlations are already scale-free and so do not require transformation. Finally, to make the magnitude of the regression coefficients of the network features directly comparable to one another, the features were normalized to have a marginal variance of 1 across the hospitals and projections in the remaining study sample.

*Corresponding author. Email: James.OMalley@Dartmouth.edu.

2 Explanatory Power: Projection Factor Results

The projection factor regression coefficients in Table 1 reveal that *Continuity* and *Revisit* are generally highly significant (z-statistics almost always > 5 and often > 10) and that the direction of their estimated coefficients (positive when enforced, negative when not enforced) is in alignment with the optimal projections in Table 2 of the main text. The effects of *Multiplicity* and *Binary* are generally less pronounced. In contrast, results for density, centralization and reciprocity are notable in that *Continuity* and *Revisit* have z-statistics of smaller magnitude compared to those for other features. The dominant predictors are the third-level of *Multiplicity* (total-ordering versus existence), for density and centralization only, and *Binary* (retaining weighted edges increases heterogeneity) for reciprocity as well.

Table 1: Z-statistics of regression coefficients of projection-method factor model

		Multiplicity			Other projection factors		
Measure	TCount	TOrder	NCount	NOrder	*Cont*	*Rev*	*Bin*
Base Features							
Density	0.13	2.89	−0.04	−0.06	0.55	1.57	−3.85
Centralization	0.31	2.98	−0.02	−0.06	0.16	1.57	−2.60
Triangles (undir)*	.	.	.	.	−0.24	−0.98	.
Ave Clustering*	.	.	.	.	7.73	8.48	.
Size*	.	.	.	.	−7.01	−30.41	.
Isolates*	.	.	.	.	6.53	18.73	.
Ncomponents*	.	.	.	.	6.53	18.73	.
Diameter*	.	.	.	.	10.04	11.36	.
Directed Features (only defined on directed networks)							
Cor(In, Out Deg)	−0.08	1.47	0.48	1.18	4.29	11.08	−1.80
Assort(In,In)	0.37	0.06	−0.22	−0.53	8.84	18.54	1.91
Assort(In,Out)	0.15	−0.16	−0.04	0.00	7.50	10.74	0.30
Assort(Out,In)	−0.08	1.47	0.48	1.18	4.29	11.08	−1.80
Assort(Out,Out)	0.37	0.06	−0.22	−0.53	8.84	18.54	1.91
Reciprocity	−0.26	0.61	1.27	0.89	1.45	2.94	−7.84
Transitive Triads	0.00	0.00	0.00	−0.01	−12.35	−14.48	0.00
Cyclical Triads	0.00	0.00	0.00	−0.01	−10.54	−13.34	0.00
Network-wide (Pooled feature) regressions							
ICC (ρ_j)	0.99	−0.19	−0.58	−1.20	2.88	4.67	5.32
Ave Stand Var ($\bar{sz}_j$)	0.23	2.82	0.22	0.22	3.60	7.20	−3.74
Ave Rank ($\bar{r}_j$)	0.40	0.47	0.14	−0.46	4.52	8.80	−3.59

* *Feature indicates that the feature is invariant to weighted edges and so yields the same value as for the binary network. In addition, the same binary network was realized across all levels of* Multiplicity. *Note the simplifying abbreviations for the five levels of* Multiplicity *(Exist, TCount, TOrder, NCount, and NOrder) and* Continuity = Cont, Revisit = Rev, *and* Binary = Bin.

The results in the Network-wide segment of Table 1 (bottom three rows) are consistent across the three network-wide discriminatory power measures. The standout factor is *Revisit*, whose presence is associated with greater heterogeneity in the resulting hospital networks (z-statistics 4.67, 7.20 and 8.80 for $\hat{\rho}_j$, $\bar{sz}_j$ and $\bar{r}_j$). Imposing *Continuity* is also associated with increased overall heterogeneity (t-statistics 2.88, 3.60 and 4.52). However, converting the network to binary was associated with increased $\hat{\rho}_j$ but decreased $\bar{sz}_j$ and $\bar{r}_j$ reinforcing the less decisive results for *Binary* seen in Table 2 of the main text.

3 Optimal and Baseline Estimated Models of ICD-status

Table 2: Z-statistics of regression coefficients of $\Delta(c)$-optimal and baseline binary-valued model of ICD status

Term	Directed	Base (Undir)
Multiplicity	TotCount	Existence
Continuity	1	0
Revisit	0	0
Directed	1	0
Binary	1	1
Undirected Measures		
Density	−3.272	−3.259
Centralization	−1.498	−1.851
Triangles (undir)	−2.316	−1.127
Ave Clustering	0.265	2.876
Size	−1.131	1.365
Isolates	−1.600	0.422
Components	−0.186	0.162
Diameter	−0.859	−0.211
Directed Measures		
Cor(In, Out Degree)	2.290	
Assort(In,In)	−0.398	
Assort(In,Out)	0.132	
Assort(Out,In)	0.108	
Assort(Out,Out)	0.332	
Reciprocity	0.436	
Transitive triads	2.733	
Cyclical triads	−2.250	

The z-statistics are from the model in Equation (7) in the main text with the predictors evaluated according to the projections specified in the segment of the table.

The z-statistics of the network features for the model in Equation (7) in the main text are shown for the optimal and base projections to a binary network (Table 2); for parsimony, we chose to only present results for the projection method that maximized AUC when directional and base features were included as predictors (the AUC-All value in Table 4 of the main text). A z-statistic whose magnitude is greater than 2 represents a statistically significant finding. We find that triadic features played an important role in distinguishing ICD-capable and non-ICD-capable hospitals. Although a lower number of closed triads overall is associated with a hospital not being ICD-capable, having a higher proportion of closed triads that are transitive (knowing physician A refers patients to B and C makes it likely that directional patient-sharing occurs between B and C even when A is not involved) and a lower proportion that are cyclical is associated with being ICD-capable. Therefore, an organizational structure that reveals itself in the form of high transitivity (e.g., reinforced messaging or second-opinion gathering) may be a marker of a hospital that is ICD-capable. Another potential marker is a correlation between the in-degree and out-degree distribution of a hospital, which was also positively associated with ICD-capability. Therefore, balanced flows of patients to and from high degree physicians appears associated with a hospital investing in technologies for advanced cardiovascular care like ICDs.

While further work is required to evaluate the generality of these findings, the crucial point is that they were only able to be discovered by using information on directionality when constructing shared-patient physician networks. This is the first surfacing of this linkage.