Statistical Methods in Medical Research
XX(X):1–10

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Supplemental Online Appendix: Methodology for Supervised Optimization of the Construction of Physician Shared-Patient Networks

A. James O'Malley[1] and Yifan Zhao[2] and Carly Bobak[3] and Chuanling Qin[2] and Erika L. Moen[1] and Daniel N. Rockmore[4]

Abstract

This supplemental online appendix is to be published with the paper. The supplementary online materials supporting the paper include 5 R scripts used for data wrangling, plotting networks, computing descriptive statistics, and performing edgewise, dyadwise, and networkwise analyses to evaluate diagnostic accuracy under various specifications. These materials are available at:
https://github.com/kiwijomalley/SupervisedOptimalProjection
The GitHub page includes a link to another GitHub page that contains the Python code used to construct the weighted-directed shared-patient networks described in Section 2 of the main text of the paper.

A Additional dyadic diagnostic measures

We first recapitulate specific cases of the core dyadic diagnostic testing measures described in Section 3.2 of the main text:

$$\text{dyad sensitivity}(\theta, \theta^{'}) = \Pr((X_{ij}, X_{ji}) \in \mathcal{X}(\theta, \theta^{'}) \mid (Y_{ij}, Y_{ji}) \in \mathcal{Y})$$

[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]Program in Quantitative Biomedical Sciences, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA [3]Research Computing, Dartmouth College, Hanover and Department of Biomedical Data Science Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA [4]Department of Mathematics and Department of Computer Science, Dartmouth College, Hanover, NH 03755, USA

Corresponding author:
James O'Malley, The Department of Biomedical Data Science and The Dartmouth Institute for Health Policy and Clinical Practice, Geisel School of Medicine at Dartmouth College, Lebanon, NH 03756, USA
Email: James.OMalley@Dartmouth.edu

Prepared using sagej.cls *[Version: 2017/01/17 v1.20]*

$$\text{dyad specificity}(\theta, \theta^{'}) = \Pr((X_{ij}, X_{ji}) \notin \mathcal{X}(\theta, \theta^{'}) \mid (Y_{ij}, Y_{ji}) \notin \mathcal{Y})$$

$$\text{dyad PPV}(\theta, \theta^{'}) = \Pr((Y_{ij}, Y_{ji}) \in \mathcal{Y} \mid (X_{ij}, X_{ji}) \in \mathcal{X}(\theta, \theta^{'}))$$

$$\text{dyad PPV}(\theta, \theta^{'}) = \Pr((Y_{ij}, Y_{ji}) \notin \mathcal{Y} \mid (X_{ij}, X_{ji}) \notin \mathcal{X}(\theta, \theta^{'}))$$

The following two quantities referred to as "strong sensitivity" and "strong specificity" correspond to $\mathcal{Y} = (1, 1)$ and $\mathcal{Y} = (0, 0)$, respectively:

$$\text{dyad sensitivity}(\theta, \theta^{'}, \text{Mutual}) = \Pr(X_{ij} > \theta, X_{ji} > \theta^{'} \mid Y_{ij} = Y_{ji} = 1)$$

$$\text{dyad specificity}(\theta, \theta^{'}, \text{Null}) = \Pr(X_{ij} \leq \theta, X_{ji} \leq \theta^{'} \mid Y_{ij} = Y_{ji} = 0)$$

We may also evaluate the accompanying PPV measures:

$$\text{dyad mutual PPV}(\theta, \theta^{'}) = \Pr(Y_{ij} = Y_{ji} = 1 \mid X_{ij} > \theta, X_{ji} > \theta')$$

$$\text{directed dyad PPV}(\theta, \theta^{'}) = \Pr(Y_{ij} > Y_{ji} \mid X_{ij} > \theta, X_{ji} > \theta')$$

$$\text{directed dyad PPV} = \Pr(Y_{ij} > Y_{ji} \mid X_{ij} > X_{ji})$$

and the analogous NPV(θ, θ') measures for any given value of (θ, θ'). The special case $(\mathcal{Y} = \{(y_{ij}, y_{ji}) : y_{ij} - y_{ji} = 1\}, \theta - \theta' > 0)$ is the genesis of DDAUC while $(\mathcal{Y} = \{(y_{ij}, y_{ji}) : y_{ij} = y_{ji} = 1\}, \theta' < 0)$ yields strong edge sensitivity and $(\mathcal{Y} = \{(y_{ij}, y_{ji}) : y_{ij} = y_{ji} = 0\}, \theta' = \infty)$ yields strong edge specificity. One can define alternative AUC measures by conditioning on ranges of $(\theta, \theta^{'})$, such as $\theta - \theta^{'} > \delta$ where $\delta > 0$ represents a practically meaningful directional difference. Thus, as for sensitivity and specificity, strong edge versions of traditional PPV and NPV measures may be evaluated by conditioning on (X_{ij}, X_{ji}) exceeding (for PPV) or falling below (for NPV) specific thresholds. The (undirected) dyadic measures can be evaluated for any given shared-patient state but the most meaningful are likely to be those for which $\theta = \theta'$. In contrast, the most meaningful measures of directed dyad PPV and NPV are likely to be those for which $\theta > \theta'$ and $\theta < \theta'$, respectively.

In addition to the directed dyad diagnostic testing measures we may also evaluate "marginal" measures that focus on a single edge:

$$\text{edge sensitivity}(\theta) = \Pr(X_{ij} > \theta \mid Y_{ij} = Y_{ji} = 1)$$

$$\text{edge specificity}(\theta) = \Pr(X_{ij} \leq \theta \mid Y_{ij} = Y_{ji} = 0)$$

$$\text{edge PPV}(\theta) = \Pr(Y_{ij} = 1 \mid X_{ij} > \theta, X_{ji} > \theta')$$

$$\text{edge NPV}(\theta) = \Pr(Y_{ij} = 0 \mid X_{ij} \leq \theta, X_{ji} \leq \theta')$$

or "partial information" measures in which a subset of the possible states of a dyad are conditioned on:

$$\text{directed dyad AUC}(\theta) = \Pr(X_{ij} > X_{ji} + \theta \mid Y_{ij} > Y_{ji})$$

$$\text{directed dyad AUC} = \Pr(X_{ij} > X_{ji} \mid Y_{ij} > Y_{ji}).$$

In addition to the bivariate $(X_{ij}, X_{ji}^{'})$ edge-weights, we may also compute composite (same combined with reverse direction) edge-weights:

$$X_{ij}^{\rho} = (X_{ij} + \rho X_{ji})/(1+\rho) \text{ for } \rho \geq 0 \tag{1}$$

where ρ quantifies the extent that outbound versus inbound patient referrals inform the status of X_{ij}^{ρ} while $\rho > 0$ quantifies the level of reciprocity in the resulting shared-patient network. Because X_{ij}^{ρ} is a linear combination of X_{ij} and X_{ji}, it forms a single univariate edge-weight allowing standard measures of diagnostic accuracy to be computed, including threshold-based measures of AUC and the AUC estimated by the c-statistic of a fitted binary regression model. However, we found that even an optimally determined X_{ij}^{ρ} led to minimal improvement in diagnostic accuracy and that it was inferior to combining separately optimized network constructions of X_{ij}^{diff} and X_{ij}^{symm}.

At a specific threshold θ we may compute PPV and NPV for linear or other weighted combinations of the two shared-patient values for a dyad to compute other marginal measures of diagnostic accuracy, including:

$$\text{edge sensitivity}(\theta, \rho) = \Pr(X_{ij}^{\rho} > \theta \mid Y_{ij} = Y_{ji} = 1)$$

$$\text{edge specificity}(\theta, \rho) = \Pr(X_{ij}^{\rho} \leq \theta \mid Y_{ij} = Y_{ji} = 0)$$

$$\text{edge PPV}(\theta, \rho) = \Pr(Y_{ij} = 1 \mid X_{ij}^{\rho} > \theta)$$

$$\text{edge NPV}(\theta, \rho) = \Pr(Y_{ij} = 0 \mid X_{ij}^{\rho} \leq \theta),$$

where typically $\theta \geq \theta^{'}$ if $\rho \leq 1$ so that distinct factors related to being a null dyad do not contaminate the PPV results and factors related to being a mutual dyad do not contaminate the NPV results. Furthermore, PPV and NPV may be evaluated conditional on dyads for which the shared-patient weight of an edge surpasses a threshold θ and exceeds the level of patient-sharing of the reverse edge:

$$\text{dyad PPV-NPV}(\theta) = \Pr(Y_{ij} = 1, Y_{ji} = 0 \mid X_{ij} > \theta, X_{ji} \leq \theta)$$

However, because the likelihood of a directed edge as opposed to a mutual edge depends on the level of X_{ij} and X_{ji}, not just on whether one exceeds the other, we consider dyad PPV-NPV(θ) to be a less meaningful quantity than its marginal counterpart that averages over θ.

A.1 Conditional dyadic analysis: Distinguishing mutual versus null dyads

To test the ability of shared-patient data to distinguish mutual versus null dyads among those that are symmetric, we may estimate the model

$$\Pr(Y_{ij} = Y_{ji} = 1 \mid Y_{ij} = Y_{ji}, X_{ij}, X_{ji}) = \text{logit}^{-1}(\beta_0 + \beta_2 X_{ji}^{\text{diff}} + \beta_2 X_{ij}^{\text{symm}}) \tag{2}$$

on the subsample of symmetric dyads (i.e., those with $Y_{ij} = Y_{ji}$). Because $Y_{ij} = Y_{ji}$ is conditioned on and each outcome observation appears twice, there are two identical observations for each dyad making it impossible to identify β_1 in (2). Therefore, we reduce (2) to:

$$\Pr(Y_{ij} = Y_{ji} = 1 \mid Y_{ij} = Y_{ji}, X_{ij}, X_{ji}) = \text{logit}^{-1}(\beta_0 + \beta_2 X_{ij}^{\text{symm}}). \tag{3}$$

The magnitude of β_2 reflects the ability of X_{ij}^{symm} to discriminate between mutual versus null dyads. Because each included dyad has two identical observations, we account for clustering by excluding observations in which $i > j$. It suffices to use maximum likelihood for estimation. Excluding asymmetric dyads and estimating the model in (3), the regression coefficient of X_{ij}^{symm} was estimated to be $\hat{\beta}_2 = 0.00997$ ($z = 0.0045$, $p = 0.0269$) implying a positive association between patient-sharing edge-weights and the likelihood of a mutual (i.e., bidirectional) survey nomination.

B Additional empirical results

B.1 Demonstration of minimal risk of overfitting

Because no model building is performed in specifying the models used in the main text (e.g., Equations 12 and 18 of the main text) to compare the projection methods, the statistical models being compared contain few predictors, and when making predictions we set $\hat{\alpha}_i = 0$ we believed that the risk of overfitting was minimal. An advantage of not splitting the data into training and test datasets, particularly for analyses involving the modestly-sized within-hospital networks, is that the level of statistical precision is greater for analyses performed on the full dataset. To confirm that there is no risk of overfitting, we conducted a simulation study on the within-hospital network dataset to compare:

(1) the differences in AUC between the optimal models from Sections 6.3 and 6.4
(2) the AUC of the optimal model in Section 6.4

between 50% randomly-split training and test datasets. Specifically, for each of 1000 iterations, we

1. Randomly split the data at the hospital-level into 50% training and test datasets
2. Fit both models on the training data
3. Use the fitted models to predict the survey responses on the test data for both models
4. Evaluate the AUC of each model and the difference in the AUC of the two models on both the training and test datasets
5. Evaluate the difference between the test and the training datasets of (1) the difference in the AUC of the models and (2) the AUC of the optimal model in Section 6.4 of the main text

The mean of the test−training dataset difference for the difference in the optimal models in Sections 6.3 and 6.4 (quantity (1) above) of the main text was 0.0000013 (SD 0.00027) and for the difference between the fit of the model in Section 6.4 (quantity (2) above) was 0.00062 (SD 0.0435) revealing minuscule bias from overfitting. For both quantities (1) and (2), the proportion of iterations for which the training data obtained a difference of greater magnitude than the test data was only 0.487 and 0.499, respectively, further emphasizing that there appears to be no concern of overfitting. Therefore, to avoid unnecessarily reducing the statistical precision of our analyses, we decided against splitting the datasets into training and test samples.

For illustration, the plots of the sampling distributions of the test−training dataset differences for quantities (1) and (2) are shown in Appendix Figure 1 and Figure 2, respectively. These emphasize the minimal differences between the in-sample assessments of predictive accuracy made on the training data and the out-of-sample external predictions (made using the model or models estimated on the training data) on the test data. Again, there is essentially no evidence of overfitting, further justifying the decision to use the full datasets in our analyses.

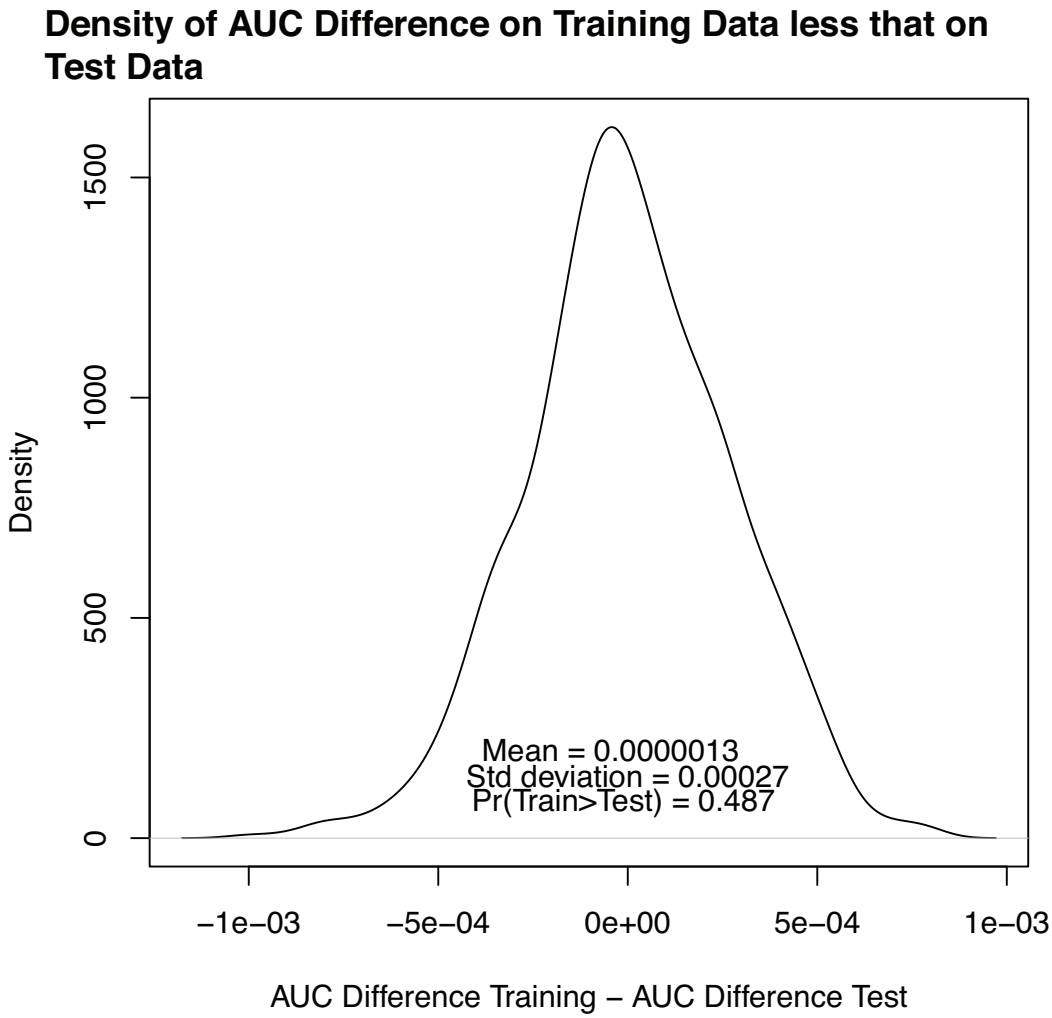

Figure 1. Sampling distribution of the test−training dataset difference of the AUC difference of the optimal models in Sections 6.3 and 6.4 of the main text

B.2 Plots of network

The plots in Figure 3 allow the within-hospital and the NPO networks to be compared with the same design and justify the observations made in the main text that within-hospital networks only have edges

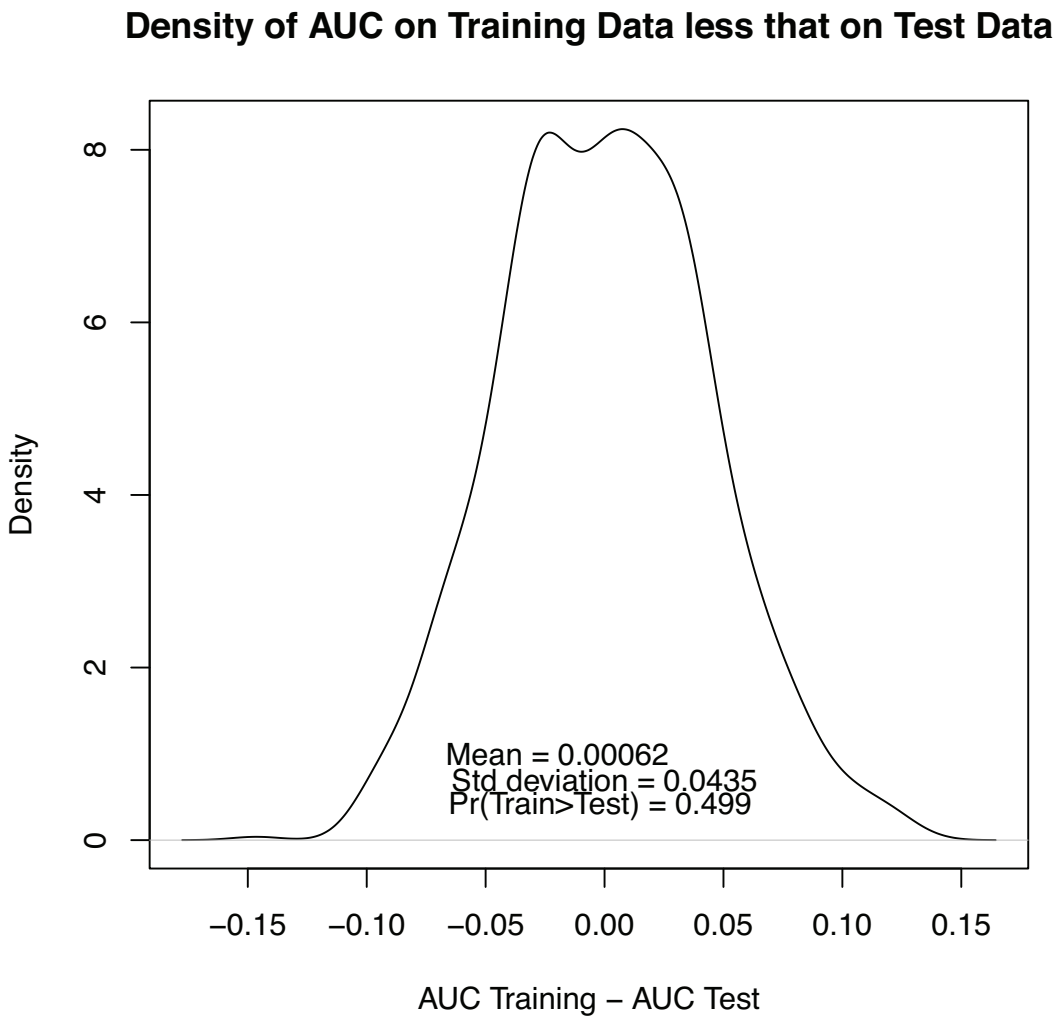

Figure 2. Sampling distribution of the test−training dataset difference of the AUC of the optimal model in Section 6.4 of the main text

between physicians at the same hospital and that imposing Revisit leads to a more highly clustered network with lower overall density compared to not enforcing Revisit.

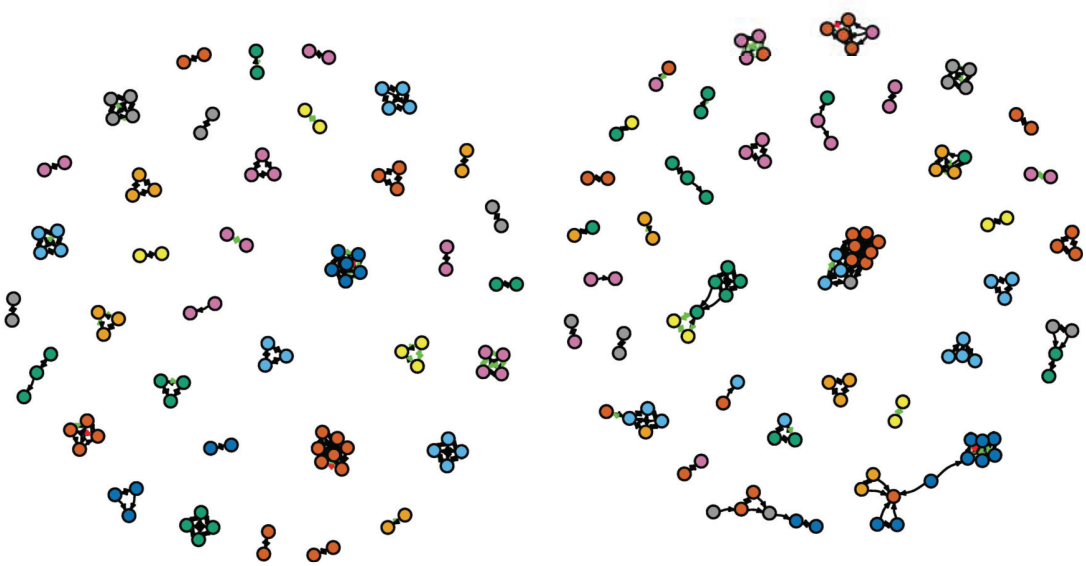

Figure 3. Directed shared-patient within-hospital (left-hand plot) and NPO (right-hand plot) physician networks constructed with Revisit binding $(= 1)$, Continuity slack $(= 0)$ and total count Multiplicity (the $(0, 1, g)$ design). To best reveal the clustering of edges by hospital in both networks and the inter-hospital edges in the NPO network, the Fruchterman-Reingold graph layout algorithm [1] is used to position the nodes. The within-hospital networks consist of the 110 physicians who made at least one within-hospital nomination while the NPO physician network consists of the 166 physicians who made at least one nomination of any type. The color of the nodes reflects the physician's primary hospital affiliation while the edges are colored green if in both the survey and the shared-patient networks, red if only in the survey network and black if only in the shared-patient network.

B.3 Impact of changing the level of Multiplicity on Diagnostic Accuracy

Because having Revisit and Continuity slack yields the best results for a single projection whereas only imposing Revisit without Continuity yields the best results for the combination of two projections, in Table 1 we only report results with Continuity slack. However, we consider each combination of Revisit and Multiplicity.

The standout finding in Table 1 is that when Revisit is binding, X_{ij}^{dir} makes a substantial contribution to the improvement in fit of Equation (12) of the main text over the SBUP with the constructions other than the total-ordering level of Multiplicity having $\Delta\chi^2 > 10$ and an AUC for the combination of measures of 0.756 or 0.757, which exceeds that of every independent projection. In each, the $z-$statistics for β_1 (the standardized coefficient of X_{ij}^{dir}) were between -3.2 and -3.5 ($p < 0.001$) whereas the z-statistics for β_2 (the coefficient of X_{ij}^{symm}) were lower than 0.6 in magnitude. When Revisit is slack, X_{ij}^{dir} is not statistically significant under any scenario whereas X_{ij}^{symm} is statistically significant or nearly so under all levels of Multiplicity other than existence, suggesting that non-directional based features from the referral paths also enhance the diagnostic accuracy of the resulting edges. These results also confirm that

Table 1. Diagnostic accuracy of shared-patient edges by Revisit and Multiplicity with respect to the same- and any-NPO-hospital name generators

Design Factors		Predictors						Deviance		AUC
		X_{ij}^{dir}		X_{ij}^{symm}		U_{ij}				
Revisit	Mult	z	p	z	p	z	p	$\Delta\chi^2$	%Dir	Comb
Same-hospital, within-hospital name generator										
0	1	-0.016	0.988	0.001	0.999	1.022	0.307	0.000	99.976	0.747
0	2	0.036	0.971	2.031	0.042	2.385	0.017	**4.249**	0.031	**0.750**
0	3	1.923	0.054	3.446	0.001	8.038	0.000	13.571	28.122	0.751
0	4	-0.356	0.722	-3.657	0.000	6.638	0.000	14.821	0.822	0.742
0	5	0.034	0.973	1.699	0.089	-1.699	0.089	2.570	0.046	0.744
1	1	**-3.258**	0.001	0.430	0.667	5.642	0.000	11.301	99.91	0.756
1	2	**-3.443**	0.001	0.538	0.591	6.227	0.000	**12.86**	99.51	**0.756**
1	3	-0.672	0.501	0.414	0.679	11.172	0.000	2.478	20.09	0.748
1	4	**-3.305**	0.001	0.263	0.793	5.277	0.000	11.73	100.00	0.756
1	5	**-3.234**	0.001	-0.282	0.778	6.139	0.000	12.141	93.28	0.757
Entire NPO, any physician name generator										
0	1	1.774	0.076	-8.769	0.000	11.738	0.000	85.836	11.157	0.960
0	2	1.260	0.208	-2.865	0.004	12.990	0.000	**8.911**	17.913	**0.960**
0	3	2.598	0.009	4.950	0.000	33.829	0.000	25.054	27.696	0.963
0	4	-0.267	0.789	-9.328	0.000	19.143	0.000	82.335	0.087	0.963
0	5	0.783	0.433	2.816	0.005	-2.816	0.005	7.469	8.243	0.963
1	1	**-1.980**	0.048	-7.546	0.000	27.595	0.000	**71.646**	5.165	**0.963**
1	2	**-2.301**	0.021	-5.923	0.000	26.756	0.000	68.992	8.369	0.963
1	3	-0.201	0.840	0.863	0.388	48.094	0.000	2.259	1.865	0.963
1	4	**-2.476**	0.013	-8.627	0.000	27.743	0.000	82.271	7.307	0.963
1	5	**-2.305**	0.021	-8.438	0.000	29.718	0.000	63.618	8.255	0.963

AUC denotes Area Under the Receiver Operating Characteristic (ROC) curve. The difference in the deviance of the model in Equation (12) of the main text having $(X_{ij}^{\text{dir}}, X_{ij}^{\text{symm}}, U_{ij})$ as predictors less that of the SBUP projection (undirected edge weights under the $(0, 0, e)$ design; see footnote under Table 2 of the main text) is denoted $\Delta\chi^2$ (see Equation 13, main text) while %Dir is the proportion of $\Delta\chi^2$ due to X_{ij}^{dir} alone. AUC denotes Area Under the Receiver Operating Characteristic (ROC) curve and Comb the linear predictor of the model in (12, main text). z and p denote z-statistic and p-value, respectively, for the tests of statistical significance of the estimates of the individual regression coefficients. The threshold at which $\Delta\chi^2$ attains statistical significance at the 5%-level is approximately 6; this represent a test of the joint statistical significance of $(X_{ij}^{\text{dir}}, X_{ij}^{\text{symm}})$ net of U_{ij}^{base}. Revisit and Mult are abbreviations for Revisit and Multiplicity, respectively. Revisit is coded as $1 =$ enforced and $0 =$ slack while the 5 levels of Mult are: $1 =$ existence, $2 =$ total count, $3 =$ total ordering, $4 =$ scaled total-count, $5 =$ scale total-ordering (see definitions in Section 2, main text).

in general Multiplicity has much less impact than Revisit and Continuity. Additional diagnostic measures including the single projection AUC and J are provided in Appendix Table 2.

A notable feature of Appendix Table 2 is that the construction with the highest single projection AUC (the $(0, 0, g)$ projection in both scenarios) differs from that with the highest J (the $(0, 0, h)$ projection for the same-hospital scenario and the $(0, 0, g^*)$ projection for the entire-NPO scenario). In both cases, Revisit is slack. Given its high impact on the diagnostic accuracy of the resulting network, it makes sense that Revisit was invariant across these single projection optimal designs. The optimal J of 0.380 for the same-hospital scenario is notable for being substantially greater than the corresponding values under other projections.

Table 2. Diagnostic properties of shared-patient edge-weights by Revisit and Multiplicity and both name generator

Design		Diagnostic Measures						
Revisit	Mult	AUC	J	Opt-J	Sens	Spec	PPV	NPV
Same-hospital, within-hospital name generator								
0	1	0.747	0.355	18	0.787	0.569	0.085	0.981
0	2	**0.750**	0.361	26	0.761	0.600	0.089	0.98
0	3	0.742	**0.380**	209	0.713	0.667	0.099	0.979
0	4	0.739	0.361	3.7	0.865	0.496	0.081	0.986
0	5	0.745	0.365	15.3	0.822	0.543	0.084	0.984
1	1	0.700	0.318	2	0.678	0.640	0.088	0.975
1	2	**0.702**	**0.324**	3	0.609	0.715	0.098	0.973
1	3	0.689	0.316	54	0.643	0.672	0.091	0.974
1	4	0.695	0.312	2.167	0.661	0.651	0.088	0.974
1	5	0.672	0.279	0.506	0.774	0.505	0.074	0.978
Entire NPO, any physician name generator								
0	1	0.960	0.913	0.000	0.923	0.990	0.027	1.000
0	2	**0.960**	0.913	0.000	0.923	0.990	0.027	1.000
0	3	0.960	0.913	0.000	0.923	0.990	0.027	1.000
0	4	0.960	**0.914**	0.164	0.923	0.990	0.029	1.000
0	5	0.960	0.913	0.105	0.923	0.990	0.027	1.000
1	1	**0.884**	0.766	0.000	0.770	0.996	0.054	1.000
1	2	0.883	0.766	0.000	0.770	0.996	0.054	1.000
1	3	0.883	0.766	0.000	0.770	0.996	0.054	1.000
1	4	0.884	0.766	0.333	0.770	0.996	0.054	1.000
1	5	0.883	0.766	0.059	0.770	0.996	0.054	1.000

Indep (Independent) AUC is the AUC for the indicated design alone, J is the corresponding value of Youden's statistic, Opt-J is the threshold applied to the shared-patient values corresponding to the optimal J. Sens, Spec, PPV and NPV denote sensitivity, specificity, positive predictive value and negative predictive value, respectively.

B.4 Estimation of HP2 model to evaluate usefulness of within-hospital shared-patient relationships

The parameter estimates for the HP2 model in Equations (21) and (22) of Section 4.5 of the main text reveal strong average density and reciprocity effects and substantial heterogeneity across the hospitals (Appendix Table 3). However, the estimated associations between the predictors (X_{ij}^{dir} and X_{ij}^{symm}) representing the shared-patient weights and the survey nominations were far from statistically significant. A reason for the lack of significance is that the set of eligible physicians who could be named in this analysis reduced to the survey respondents at the same hospital. Prior analyses have revealed that within-hospital networks of the NPO tend to be close to complete graphs, even with a threshold for defining binary-valued ties as high as 100.[2]

References

1. Fruchterman T and Reingold E. Graph drawing by force-directed placement. *Software - Practice and Experience* 1991; 21(11): 1129–1164.

Table 3. Hierarchical P2 models estimated on the hospital networks comprising the NPO

	Base model			+ Dir patient-sharing			+ Recip patient-sharing		
Term	Estimate	z	p	Estimate	z	p	Estimate	z	p
Random-effect variance-covariance parameters									
Sender variance	7.866	0.177	0.860	2.442	1.937	0.053	3.596	0.631	0.528
Sender-receiver cov	3.010	0.116	0.908	-0.091	-0.111	0.911	0.913	0.285	0.776
Receiver variance	4.482	0.282	0.778	1.981	1.783	0.075	2.342	1.278	0.201
Density variance	0.671	0.935	0.350	0.386	1.016	0.310	0.719	0.974	0.330
Reciprocity variance	0.380	1.267	0.205	0.399	1.758	0.079	0.548	1.809	0.071
Physician covariates									
Sender hospital size				-0.519	-6.179	0.000	-0.324	-4.208	0.000
Receiver hospital size				0.065	0.591	0.555	0	0.000	1.000
Density terms									
Intercept	-3.017	-7.270	0.000	-3.100	-9.841	0.000	-2.834	-8.311	0.000
Dir shared patient				0.810	1.601	**0.109**	0.137	0.267	**0.789**
Symm shared patient				-0.246	-0.596	**0.551**	-0.002	-0.003	**0.997**
Reciprocity terms									
Intercept	1.082	1.279	0.201	1.361	1.844	0.065	0.423	2.747	0.006
Symm shared patient							-1.008	-0.824	**0.410**

Dir, recip, z and p denote directed, reciprocated, z-statistic and p-value, respectively.

2. Bobak CA, Mohan D, Murphy MA et al. Constructing within and between hospital physician social networks for modeling physician research participation. *BMC Medical Research Methodology* 2023; 23(253). DOI: 10.1186/s12874-023-02069-2.