2-2018

Statistical Analysis of Network Change

Teresa D. Schmidt
Portland State University, tds@pdx.edu

Martin Zwick
Portland State University, zwick@pdx.edu

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

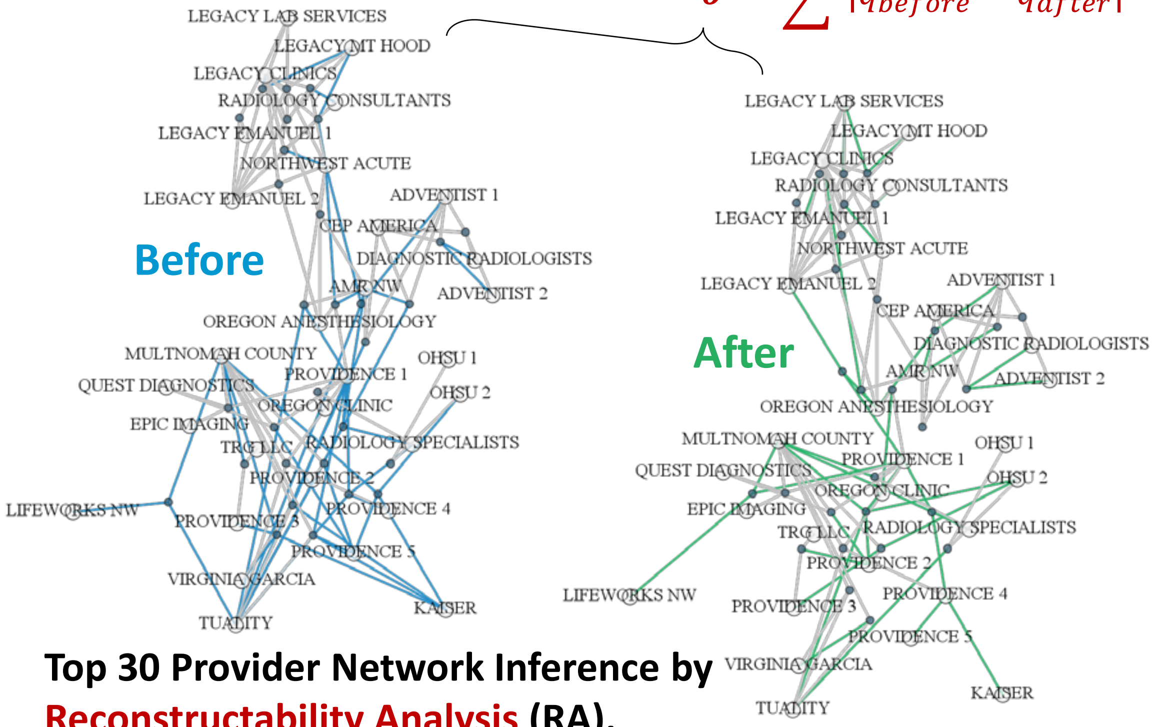
1 Inference

Infer a healthcare network from the top 30 billing providers in Medicaid claims data from before and after a legislative change.

2 Distance

For **RA**, network distance is the sum of absolute differences in calculated probability distributions.

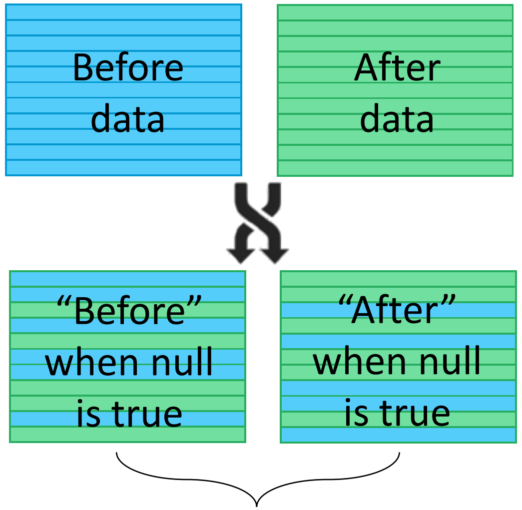
$$\hat{\theta} = \sum |q_{before} - q_{after}|$$



Top 30 Provider Network Inference by Reconstructability Analysis (RA),
a data mining methodology that uses information theory to derive relationships among variables in the form of a best model (hypergraph).

3 Permutation

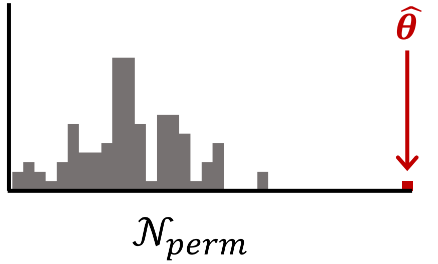
Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.



$$\hat{\theta}_{perm} = \sum |q_{before} - q_{after}|$$

4 Testing

Test null hypothesis by comparing the observed distance $\hat{\theta}$ against the reference distribution generated through permutation.



RA Results

The billing provider network changed significantly over time ($p < .01$). Connectivity consolidated into fewer relationships among providers and higher-way relationships that involved fewer providers.

Conclusion

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

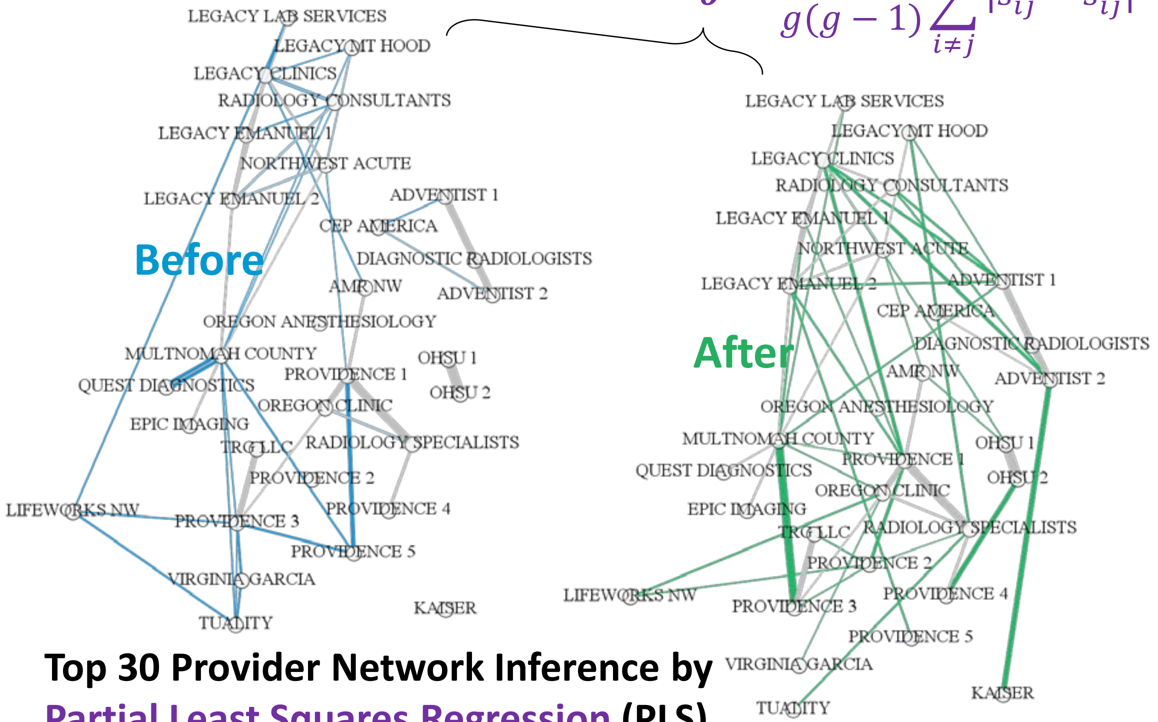
Portland State UNIVERSITY

1 Inference | **2 Distance** | **3 Permutation** | **4 Testing**

Infer a healthcare network from the top 30 billing providers in Medicaid claims data from before and after a legislative change.

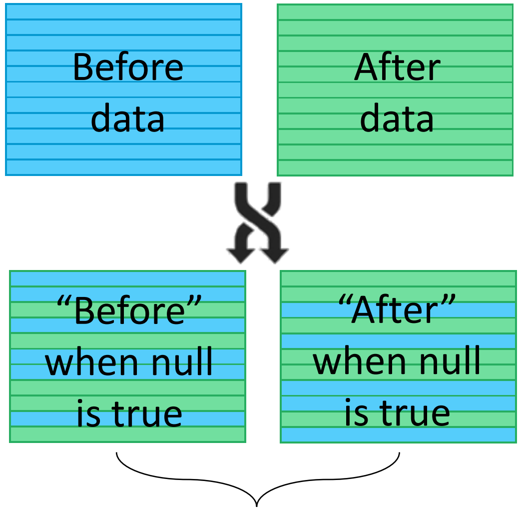
For **PLS**, network distance is the average absolute difference in the networks' connectivity matrices.

$$\hat{\boldsymbol{\theta}} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^{1} - \hat{s}_{ij}^{2}|$$



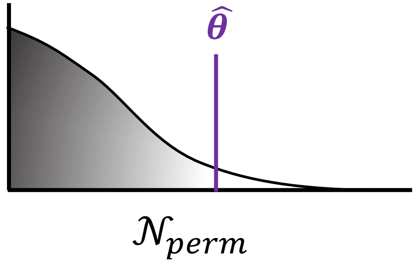
Top 30 Provider Network Inference by Partial Least Squares Regression (PLS), defines a relationship by how much one provider contributes to latent terms of covariance with other provider's billing patterns.

Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.



$$\hat{\theta}_{perm} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^{1} - \hat{s}_{ij}^{2}|$$

Test null hypothesis by comparing the observed distance $\hat{\theta}$ against the reference distribution generated through permutation.



PLS Results

The billing provider network changed significantly over time ($p < .001$). The structure of the network changed dramatically, and many strong connections appeared between healthcare organizations.

Conclusion

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.

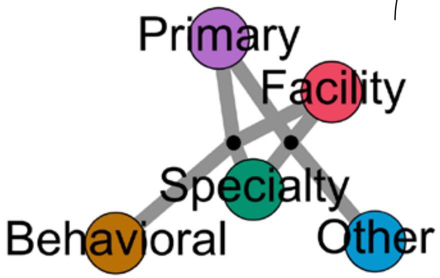
Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

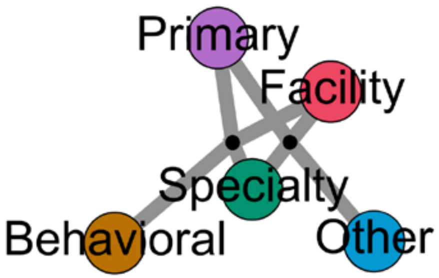
1 Inference

Infer a network of healthcare sectors* from Medicaid claims data from before and after a legislative change.

Before



After



*Sectors defined by taxonomy groups from the National Uniform Claims Committee

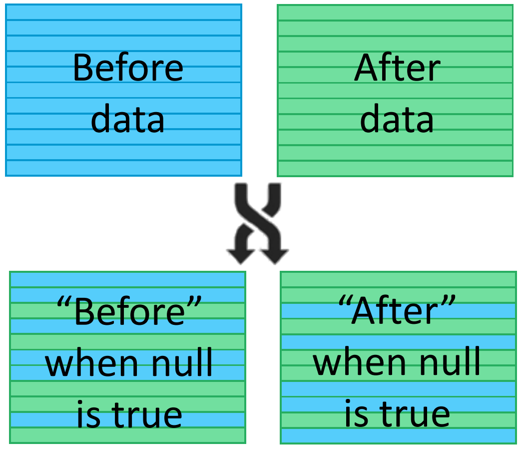
2 Distance

For **RA**, network distance is the sum of absolute differences in calculated probability distributions.

$$\hat{\boldsymbol{\theta}} = \sum |q_{before} - q_{after}|$$

3 Permutation

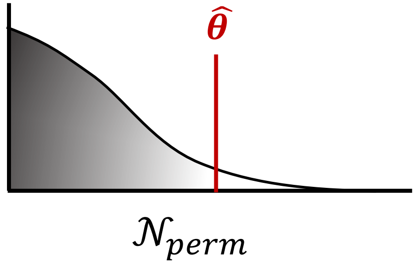
Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.



$$\hat{\theta}_{perm} = \sum |q_{before} - q_{after}|$$

4 Testing

Test null hypothesis by comparing the observed distance $\hat{\theta}$ against the reference distribution generated through permutation.



RA Results

Data from both time periods indicate two four-way relationships among sectors of care. Probability distributions changed significantly ($p < .001$), indicating an increase in patients who have only facility claims and no other type.

Healthcare Sector Network Inference by Reconstructability Analysis (RA),

a data mining methodology that uses information theory to derive relationships among variables in the form of a best model (hypergraph).

Conclusion

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.

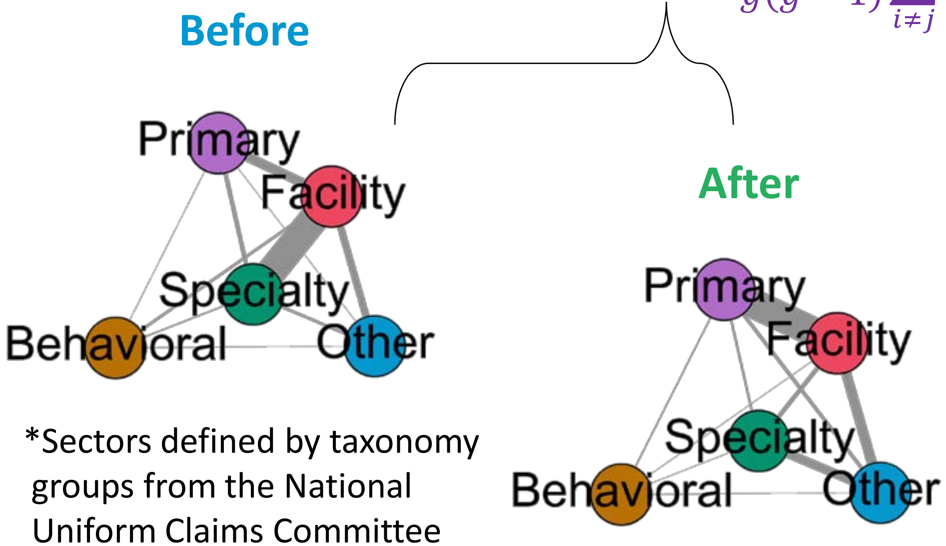
Statistical Analysis of Network Change

Portland State University

Teresa D. Schmidt and Martin Zwick

1 Inference

Infer a network of healthcare sectors* from Medicaid claims data from before and after a legislative change.



*Sectors defined by taxonomy groups from the National Uniform Claims Committee

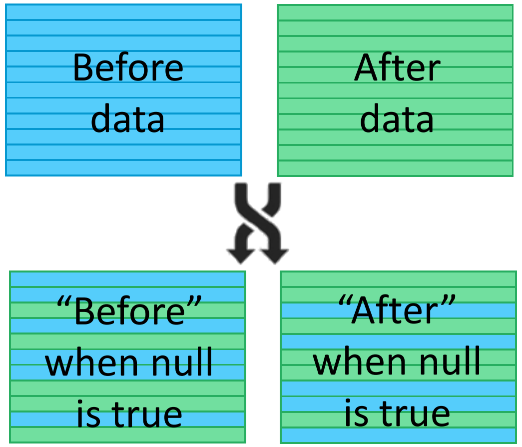
2 Distance

For **PLS**, network distance is the average absolute difference in the networks' connectivity matrices.

$$\hat{\boldsymbol{\theta}} = \frac{1}{g(g-1)}\sum_{i\neq j}\left|\hat{s}^1_{ij} - \hat{s}^2_{ij}\right|$$

3 Permutation

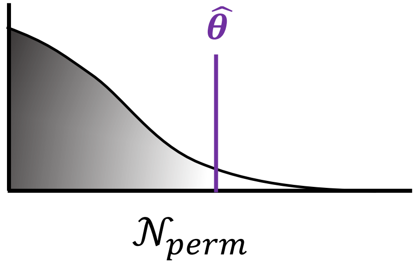
Shuffle observations between datasets, infer networks from permuted datasets, calculate distances where the null is true.



$$\hat{\theta}_{perm} = \frac{1}{g(g-1)}\sum_{i\neq j}\left|\hat{s}^1_{ij} - \hat{s}^2_{ij}\right|$$

4 Testing

Test null hypothesis by comparing the observed distance $\hat{\theta}$ against the reference distribution generated through permutation.



PLS Results

The network of healthcare sectors changed significantly ($p < .001$) over time. The facility sector became more strongly connected with primary, and primary became more connected with behavioral and less with specialty care.

Healthcare Sector Network Inference by Partial Least Squares Regression (PLS),

defines a relationship by how much one healthcare sector contributes to latent terms of covariance with other sector's billing patterns.

Conclusion

Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them.

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

Data Requirements by Reconstructability Analysis (RA)
Data must be formatted so that variables represent network 'nodes', and cases reflect observations of those nodes' behavior. RA uses nominal data, but has no distributional assumptions

Example Claims Data during Before Period

	Prov 1 (x_1)	Prov 2 (x_2)	Prov 3 (x_3)	Prov 4 (x_4)	Prov 5 (x_5)
Patient 1	0	0	1	2	0
Patient 2	1	2	0	0	0
Patient 3	2	1	0	1	0
Patient 4	0	0	0	1	2
Patient 5	2	2	0	0	0
Patient 6	0	0	0	0	2
Patient 7	2	0	0	0	0
Patient 8	0	0	1	2	0

In the billing provider network, there is a variable for each billing provider, and a record for each distinct patient. Values indicate whether a provider billed for a patient 0, 1, or 2+ times during the time period. (Other levels could also be c hosen.)

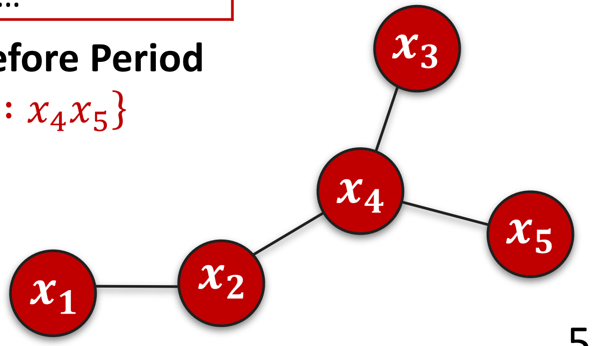
Network Inference by Reconstructability Analysis
RA produces a best model of associations (e.g., by BIC), with calculated probabilities for all combinations of variable states. This best model can be interpreted as a hypergraph with three-way and higher-way associations between variables (nodes).

Example Calculated Probability Distribution from Before Period

x_1	x_2	x_3	x_4	x_5	Probability for a Patient
0	0	0	0	0	q_{0000}=0.0012
0	0	0	0	1	q_{0001}=0.0004
0	0	0	1	0	q_{0010}=0.0070
0	0	0	1	1	q_{0011}=0.0100
0	0	1	0	0	q_{0100}=0.0016
0	0	1	0	1	q_{0000}=0.0009
…					…

Example Best Model during Before Period

$\{x_1 x_2 : x_2 x_4 : x_3 x_4 : x_4 x_5\}$



Reference: Zwick, M. (2004). An overview of reconstructability analysis. *Kybernetes, 33*(5/6), 877-905.

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

Portland State University

Data Requirements by Partial Least Squares (PLS)

Data must be formatted so that variables represent network 'nodes', and cases reflect observations of those nodes' behavior. PLS uses continuous data, but has distribution assumptions.

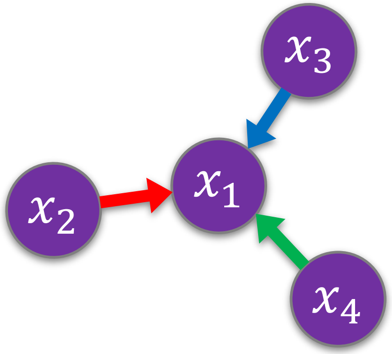
Claims Data during Before Period

	Prov 1 (x_1)	Prov 2 (x_2)	Prov 3 (x_3)	Prov 4 (x_4)	Prov 5 (x_5)
Patient 1	0	0	1	2	0
Patient 2	1	2	0	0	0
Patient 3	7	1	0	1	0
Patient 4	0	0	0	3	2
Patient 5	2	4	0	0	0
Patient 6	0	0	0	0	6
Patient 7	2	0	0	0	0
Patient 8	0	0	1	2	0

In the billing provider network, there is a variable for each billing provider, and a record for each distinct patient during a time period. Values indicate the number of claims that each provider billed for each patient during a time period.

Network Inference by Partial Least Squares

PLS determines the connectivity to any given node by the contributions of each other node to latent terms that vary with the given node's behavior.



Calculation of Connections to Example Node

x_1

$$\hat{x}_1 = \beta_0 + \boldsymbol{\beta}_1^{(1)} t_1^{(1)} + \boldsymbol{\beta}_1^{(2)} t_1^{(2)} + \cdots + \boldsymbol{\beta}_1^{(\ell)} t_1^{(\ell)}$$

$$= c_{12}^{(1)} x_2 + c_{13}^{(1)} x_3 + c_{14}^{(1)} x_4$$

$$= c_{12}^{(2)} x_2 + c_{13}^{(2)} x_3 + c_{14}^{(2)} x_4$$

$$= c_{12}^{(\ell)} x_2 + c_{13}^{(\ell)} x_3 + c_{14}^{(\ell)} x_4$$

$$\hat{s}_{21} = \boldsymbol{\beta}_1^{(1)} \boldsymbol{c}_{12}^{(1)} + \boldsymbol{\beta}_1^{(2)} \boldsymbol{c}_{12}^{(2)} + \boldsymbol{\beta}_1^{(\ell)} \boldsymbol{c}_{12}^{(\ell)}$$

$$\hat{s}_{31} = \boldsymbol{\beta}_1^{(1)} \boldsymbol{c}_{13}^{(1)} + \boldsymbol{\beta}_1^{(2)} \boldsymbol{c}_{13}^{(2)} + \boldsymbol{\beta}_1^{(\ell)} \boldsymbol{c}_{13}^{(\ell)}$$

$$\hat{s}_{41} = \boldsymbol{\beta}_1^{(1)} \boldsymbol{c}_{14}^{(1)} + \boldsymbol{\beta}_1^{(2)} \boldsymbol{c}_{14}^{(2)} + \boldsymbol{\beta}_1^{(\ell)} \boldsymbol{c}_{14}^{(\ell)}$$

Reference: Datta, S. (2001). Exploring relationships in gene expressions: a partial least squares approach. *Gene Expression, 9*(6), 249-55.