

2-2018

Statistical Analysis of Network Change

Teresa D. Schmidt
Portland State University

Martin Zwick
Portland State University, zwick@pdx.edu

Let us know how access to this document benefits you.

Follow this and additional works at: https://pdxscholar.library.pdx.edu/sysc_fac

 Part of the [Logic and Foundations Commons](#), and the [Systems Architecture Commons](#)

Citation Details

Schmidt, Teresa D. and Zwick, Martin, "Statistical Analysis of Network Change" (2018). *Systems Science Faculty Publications and Presentations*. 120.

https://pdxscholar.library.pdx.edu/sysc_fac/120

This Presentation is brought to you for free and open access. It has been accepted for inclusion in Systems Science Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick



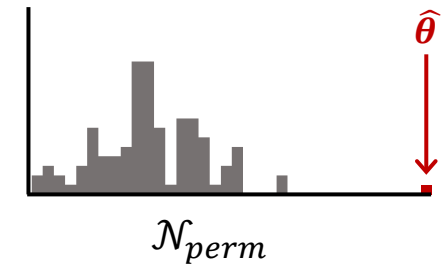
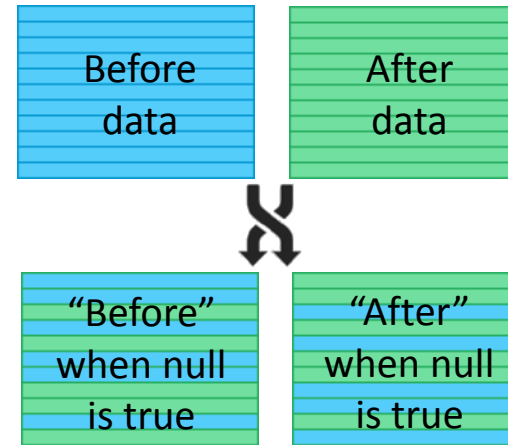
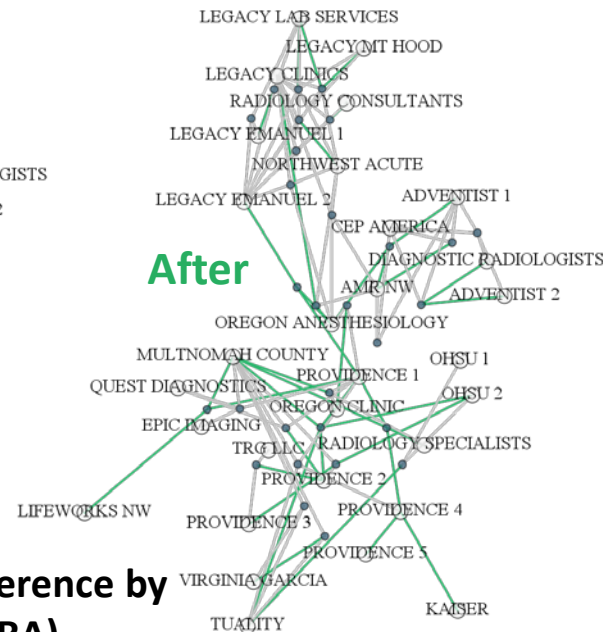
1 **Inference**
Infer a healthcare network from the top 30 billing providers in Medicaid claims data from before and after a legislative change.

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

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

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

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



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

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.

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).

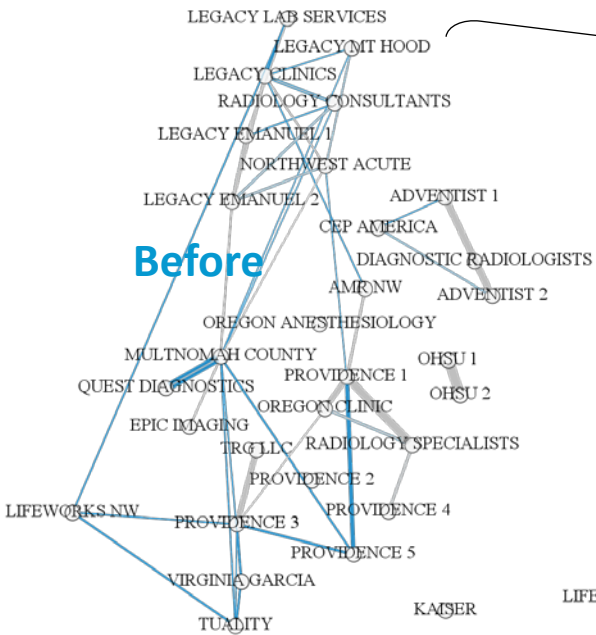
Conclusion
Networks can be tested for significant change over time when inferred from datasets of independent observations. PLS can measure change among pairs of providers, while RA identifies the higher-way relationships among them. 1

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.

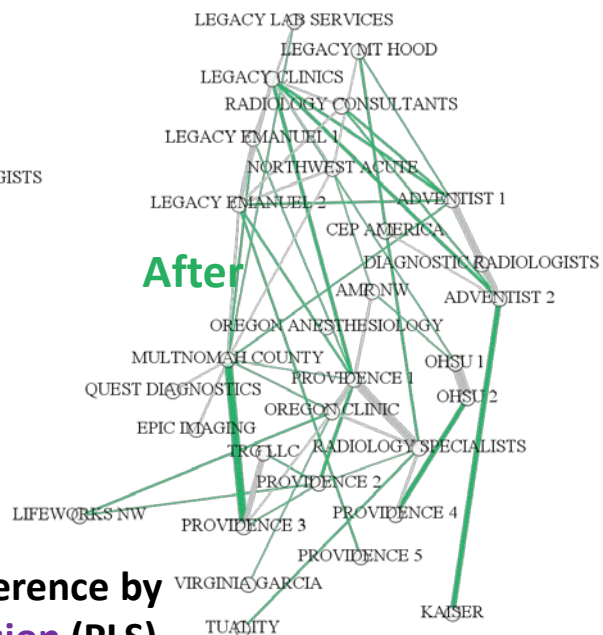


Top 30 Provider Network Inference by Partial Least Squares Regression (PLS),

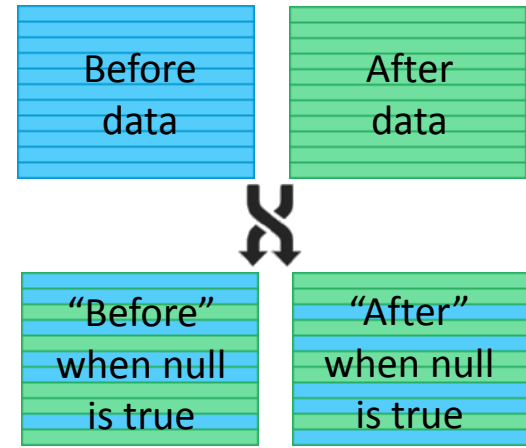
defines a relationship by how much one provider contributes to latent terms of covariance with other provider's billing patterns.

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

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



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

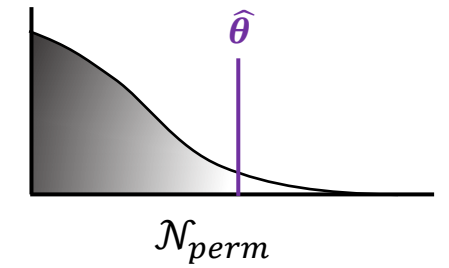


$$\hat{\theta}_{perm} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^1 - \hat{s}_{ij}^2|$$

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. 2

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



PLS Results

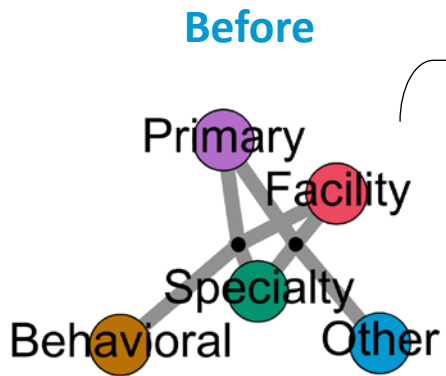
The billing provider network changed significantly over time ($p < .001$). The structure of the network changed dramatically, and many strong connections appeared between healthcare organizations.

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick



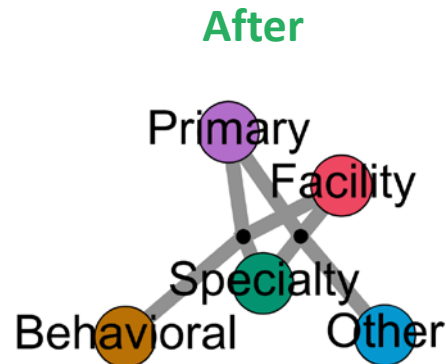
1 **Inference**
Infer a network of healthcare sectors* from Medicaid claims data from before and after a legislative change.



*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.

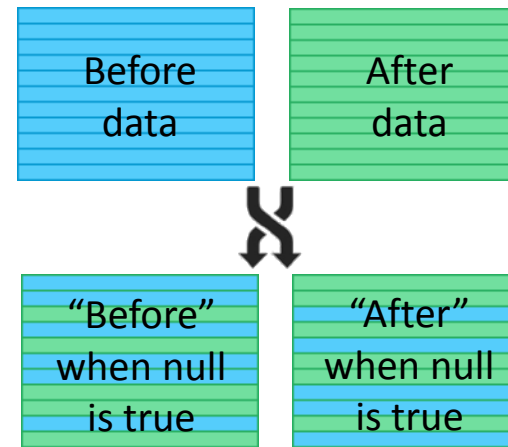
$$\hat{\theta} = \sum |q_{before} - q_{after}|$$



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).

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

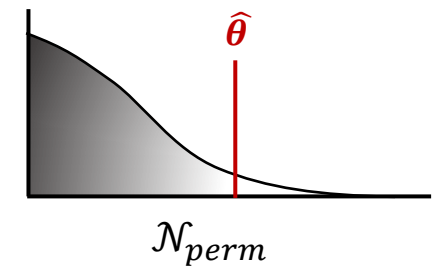


$$\hat{\theta}_{perm} = \sum |q_{before} - q_{after}|$$

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. 3

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



RA Results

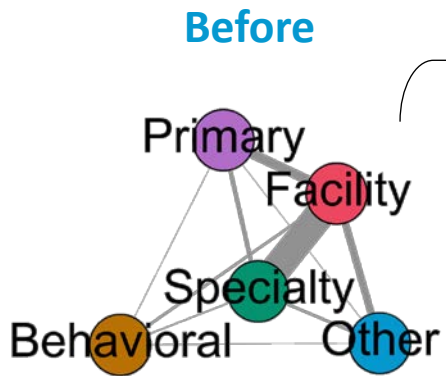
Data from both time periods indicate two four-way relationships among sectors of care. Probability distributions changed significantly ($p < .001$), indicating an increase in patients who have only facility claims and no other type.

Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

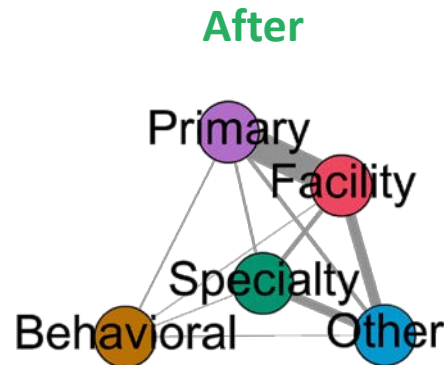


1 **Inference**
Infer a network of healthcare sectors* from Medicaid claims data from before and after a legislative change.



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

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

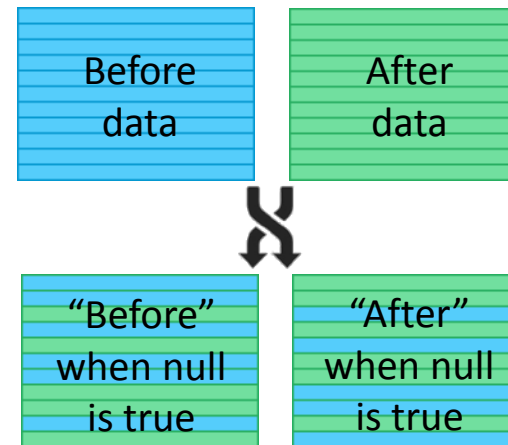


*Sectors defined by taxonomy groups from the National Uniform Claims Committee

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.

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

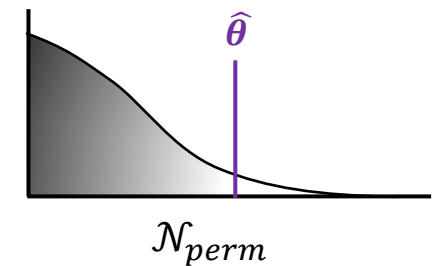


$$\hat{\theta}_{perm} = \frac{1}{g(g-1)} \sum_{i \neq j} |\hat{s}_{ij}^1 - \hat{s}_{ij}^2|$$

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. 4

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.

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

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 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

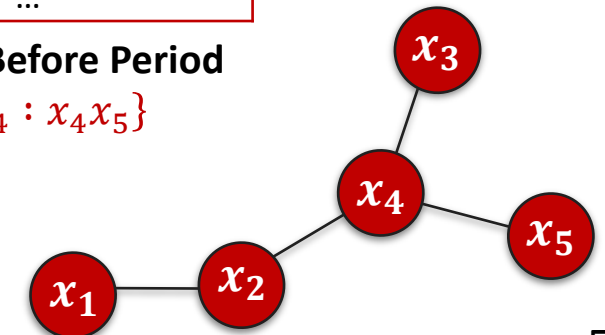
In the billing provider network, there is a variable for each billing provider, and a record for each distinct patient. Values indicate whether a provider billed for a patient 0, 1, or 2+ times during the time period. (Other levels could also be chosen.)

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	$q_{00000}=0.0012$
0	0	0	0	1	$q_{00001}=0.0004$
0	0	0	1	0	$q_{00010}=0.0070$
0	0	0	1	1	$q_{00011}=0.0100$
0	0	1	0	0	$q_{00100}=0.0016$
0	0	1	0	1	$q_{00101}=0.0009$
...					...

Example Best Model during Before Period

$$\{x_1x_2 : x_2x_4 : x_3x_4 : x_4x_5\}$$



Statistical Analysis of Network Change

Teresa D. Schmidt and Martin Zwick

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.

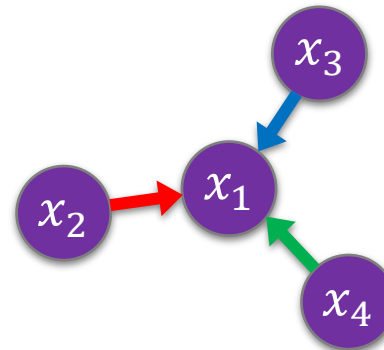
Network Inference by Partial Least Squares

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.

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.



Calculation of Connections to Example Node

x_1

$$\hat{x}_1 = \beta_0 + \beta_1^{(1)} t_1^{(1)} + \beta_1^{(2)} t_1^{(2)} + \dots + \beta_1^{(\ell)} t_1^{(\ell)}$$

$$\hat{s}_{21} = \beta_1^{(1)} c_{12}^{(1)} + \beta_1^{(2)} c_{12}^{(2)} + \beta_1^{(\ell)} c_{12}^{(\ell)}$$

$$\hat{s}_{31} = \beta_1^{(1)} c_{13}^{(1)} + \beta_1^{(2)} c_{13}^{(2)} + \beta_1^{(\ell)} c_{13}^{(\ell)}$$

$$\hat{s}_{41} = \beta_1^{(1)} c_{14}^{(1)} + \beta_1^{(2)} c_{14}^{(2)} + \beta_1^{(\ell)} c_{14}^{(\ell)}$$

