What is Importance?

Stefan Conrady | stefan.conrady@bayesia.us | +1 888-386-8383

Webinar on August 28, 2019

Today’s Agenda

Motivation & Objective

- How can we quantify importance and interpret related measures?

Dimensions of Reasoning

- Prediction vs. Causation
- Theory vs. Data
- Probabilistic vs. Deterministic
- Bayesian networks as a reasoning framework

Today's Agenda (cont'd)

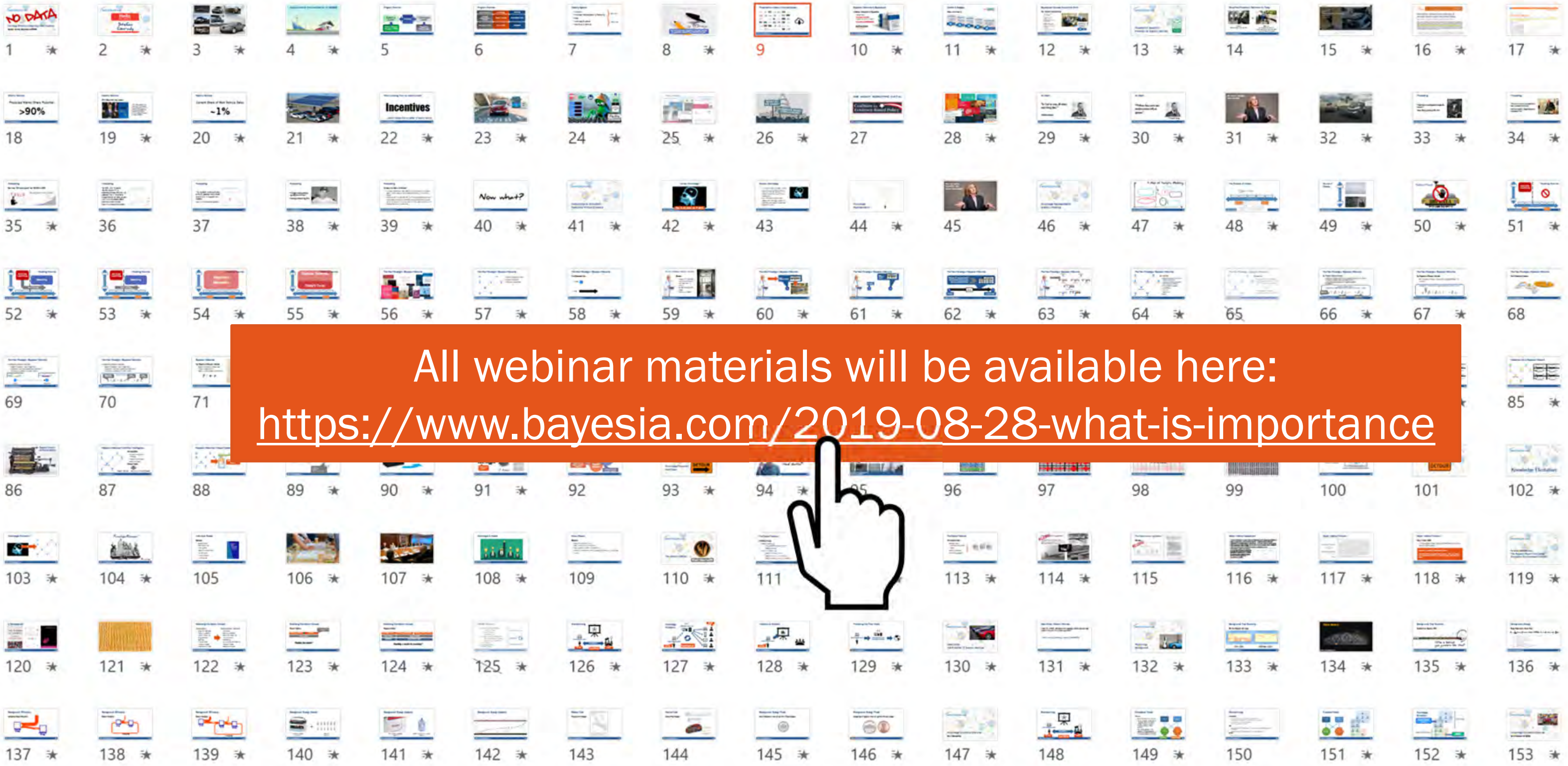
Importance in Predictive Modeling

- Total Effects
- Information Theory
    - Entropy & Mutual Information
    - Arc Force, Node Force
- Bayes Factor
- Tornado Chart

Importance in Causal Modeling

- Direct Effects
- Contributions & Synergy
- Elasticity

Slides, networks, and video will be available



All webinar materials will be available here:
https://www.bayesia.com/2019-08-28-what-is-importance

Google Scholar "the importance of"

Articles About 5,150,000 results (0.29 sec) My profile My library

Any time
Since 2019
Since 2018
Since 2015
Custom range...

Sort by relevance
Sort by date

include patents
include citations

Create alert

The methodology of focus groups: **the importance of** interaction between research participants [PDF] wiley.com
J Kitzinger - Sociology of health & illness, 1994 - Wiley Online Library
What are focus groups? How are they distinct from ordinary group discussions and what use are they anyway? This article introduces focus group methodology, explores ways of conducting such groups and examines what this technique of data collection can offer ...
Cited by 4258 Related articles All 9 versions

The importance of selenium to human health [PDF] surrey.ac.uk
MP Rayman - The lancet, 2000 - Elsevier
The essential trace mineral, selenium, is of fundamental importance to human health. As a constituent of selenoproteins, selenium has structural and enzymic roles, in the latter context being best-known as an antioxidant and catalyst for the production of active thyroid ...
Cited by 3953 Related articles All 18 versions

The maintenance of species-richness in plant communities: **the importance of** the regeneration niche [PDF] cfbiodiv.org
PJ Grubb - Biological reviews, 1977 - Wiley Online Library
SUMMARY 1 According to 'Gause's hypothesis'a corollary of the process of evolution by natural selection is that in a community at equilibrium every species must occupy a different niche. Many botanists have found this idea improbable because they have ignored the ...
Cited by 4343 Related articles All 4 versions

The importance of the ratio of omega-6/omega-3 essential fatty acids [PDF] texasgrassfedbeef.com
AP Simopoulos - Biomedicine & pharmacotherapy, 2002 - Elsevier
Several sources of information suggest that human beings evolved on a diet with a ratio of omega-6 to omega-3 essential fatty acids (EFA) of~ 1 whereas in Western diets the ratio is 15/1–16.7/1. Western diets are deficient in omega-3 fatty acids, and have excessive ...
Cited by 3192 Related articles All 18 versions

Importance

Consequence Effect Influence Priority

Influence Usefulness Value Force Significance

Weight Relevance Impact Leverage Strength

Contribution Potency Efficacy Efficiency

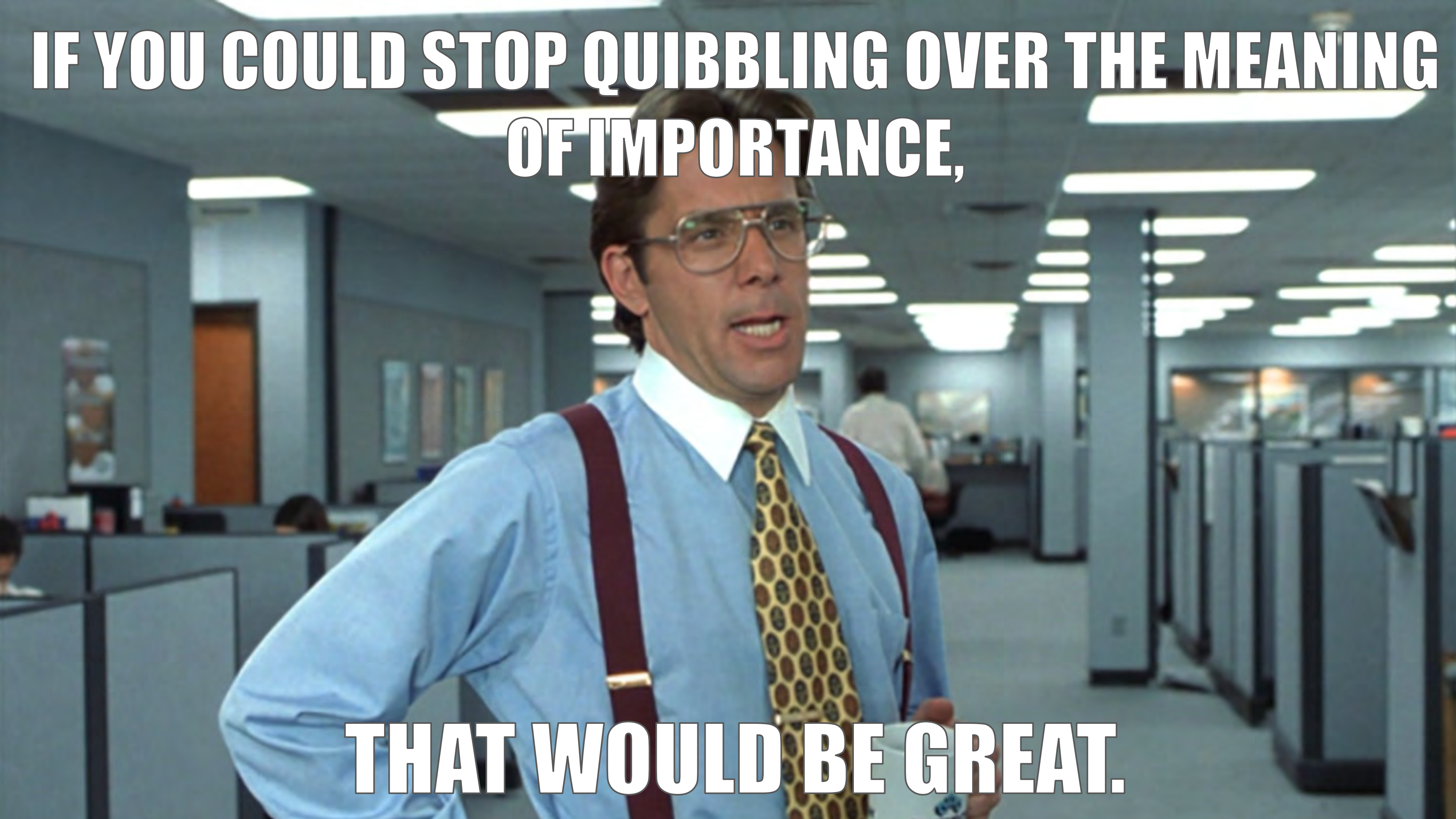
Power Capacity Magnitude Potential Gain

Effectiveness Change Support Response

Causation Connection Determinant Motive

Impulse Coefficient Parameter Propensity Bias

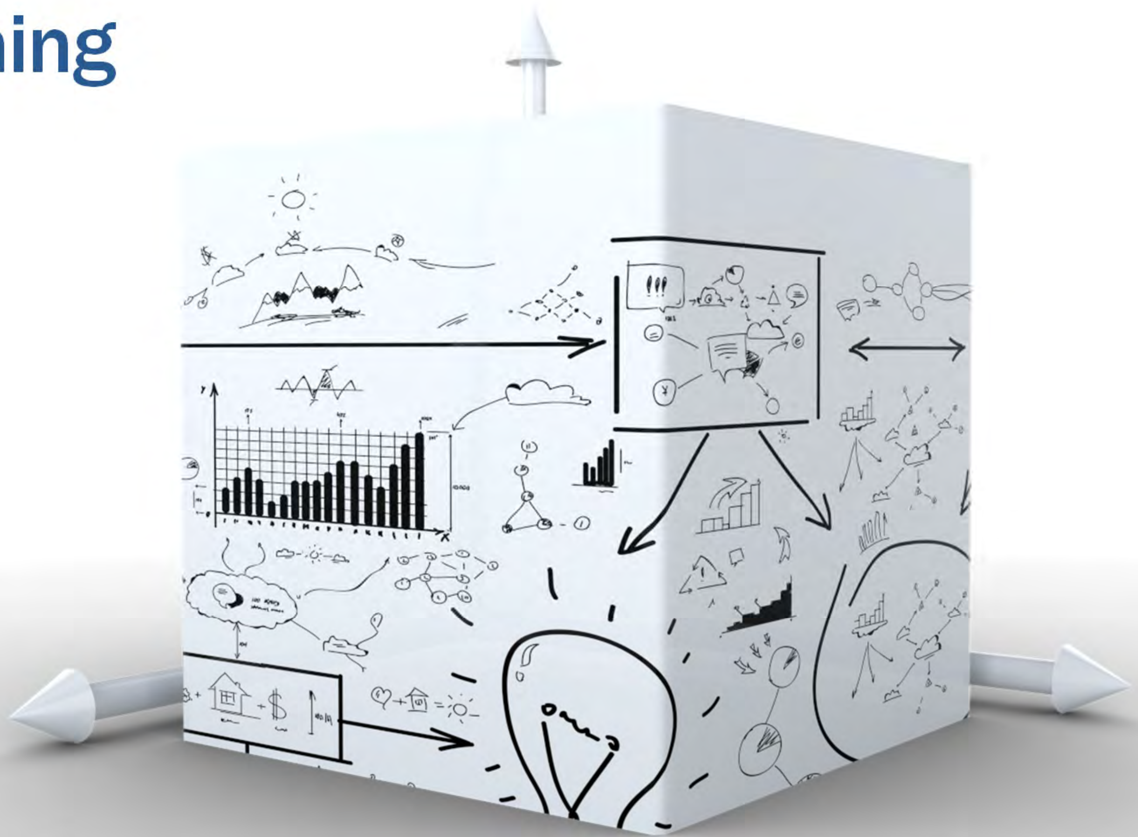
Tendency Inclination Proneness Driver Factor


IF YOU COULD STOP QUIBBLING OVER THE MEANING OF IMPORTANCE,
THAT WOULD BE GREAT.

9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32

Reasoning

Reasoning



Reasoning

Dimensions

Reasoning

Dimensions

Reasoning

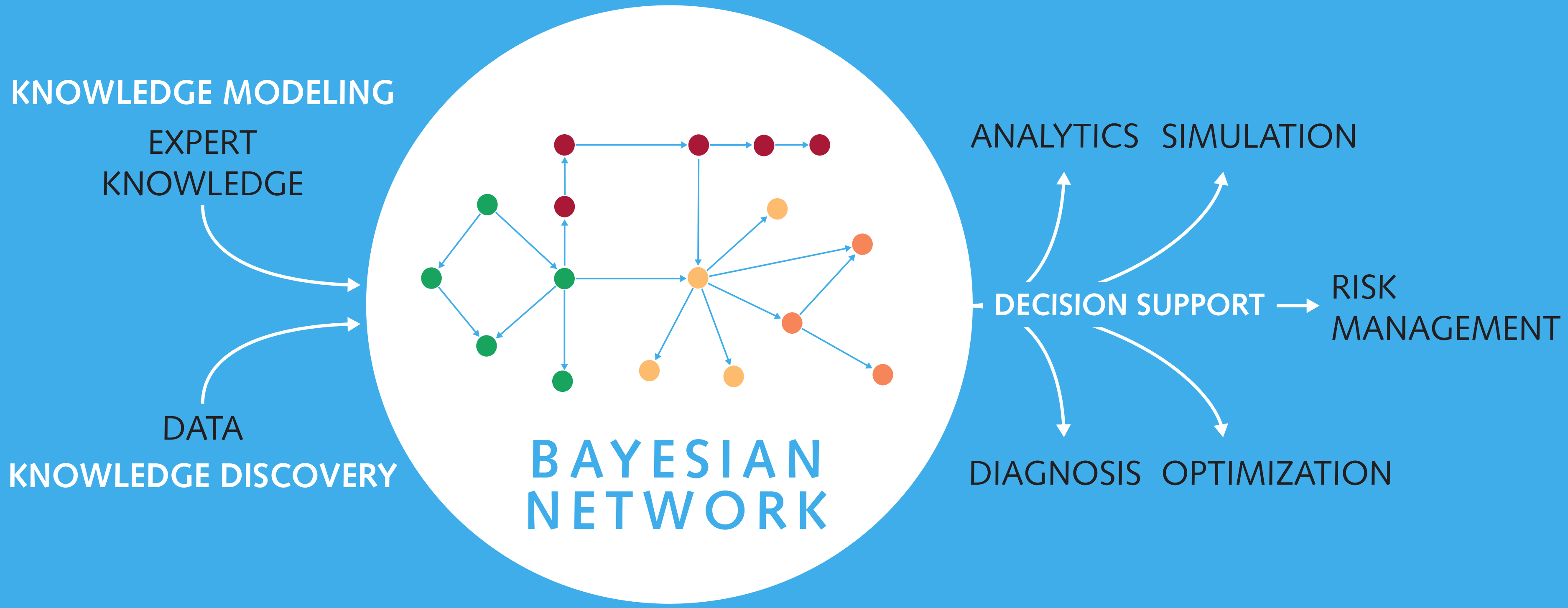
Dimensions

Reasoning

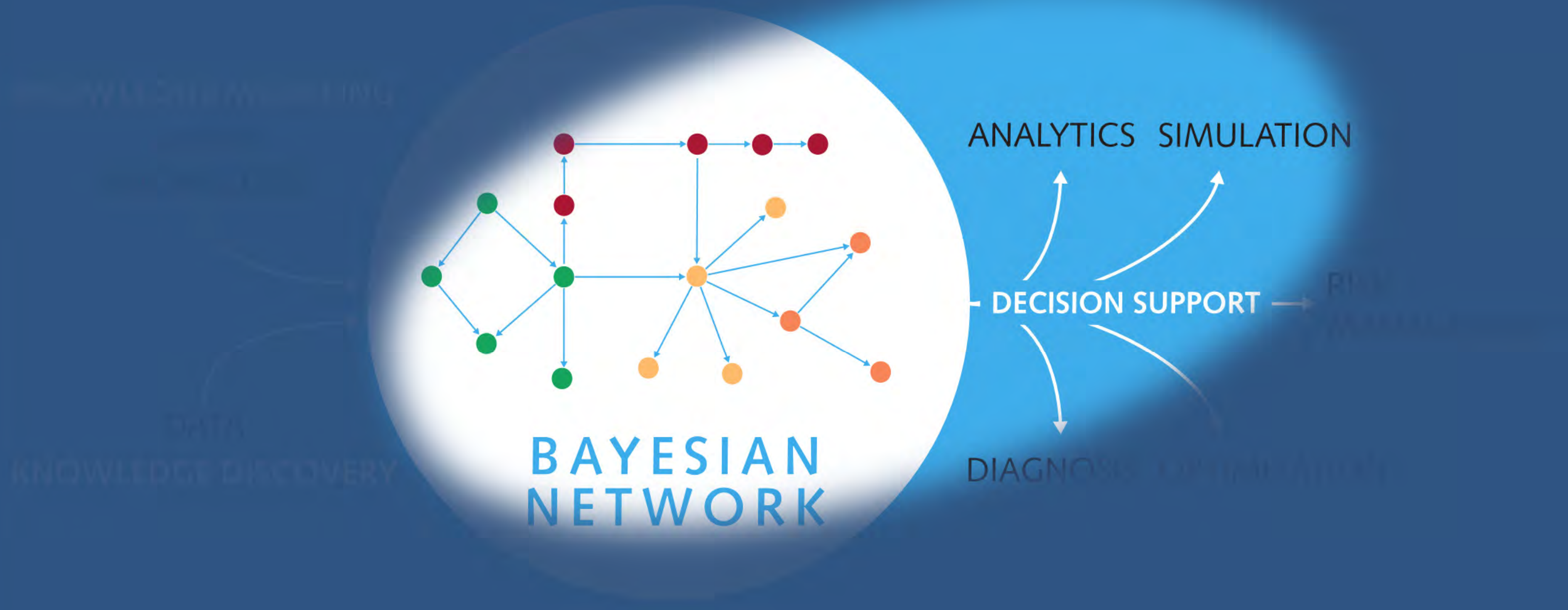
Dimensions

Bayesian Networks as Reasoning Framework

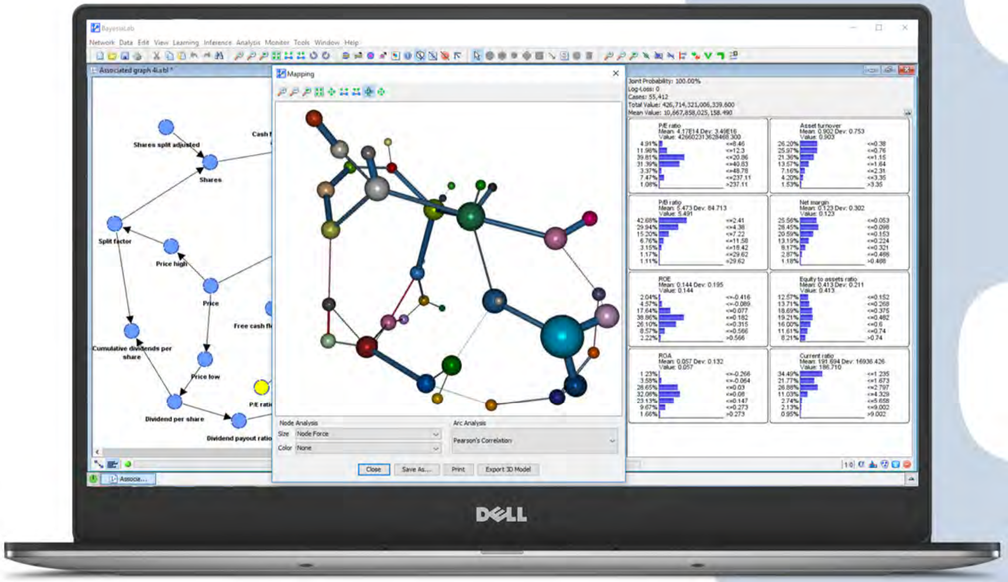
Bayesian Networks as Reasoning Framework



Bayesian Networks as Reasoning Framework



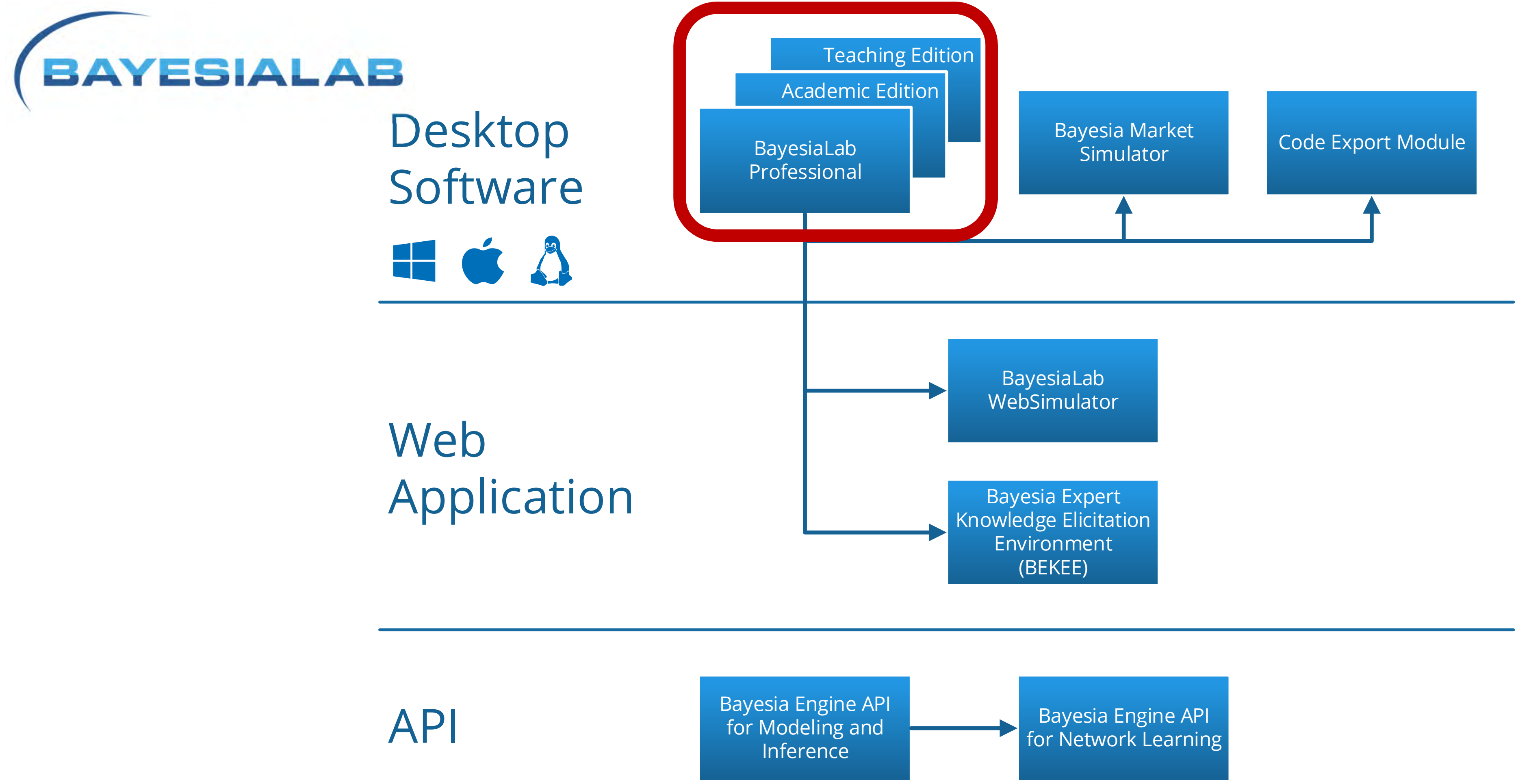
BAYESIALAB



A desktop software for:

- encoding
- learning
- editing
- performing inference
- analyzing
- simulating
- optimizing

with Bayesian networks.


BAYESIALAB
Desktop Software
Teaching Edition
Academic Edition
BayesiaLab Professional
Bayesia Market Simulator
Code Export Module
Web Application
BayesiaLab WebSimulator
Bayesia Expert Knowledge Elicitation Environment (BEKEE)
API
Bayesia Engine API for Modeling and Inference
Bayesia Engine API for Network Learning

Reasoning

Dimensions

Reasoning

Dimensions

Reasoning

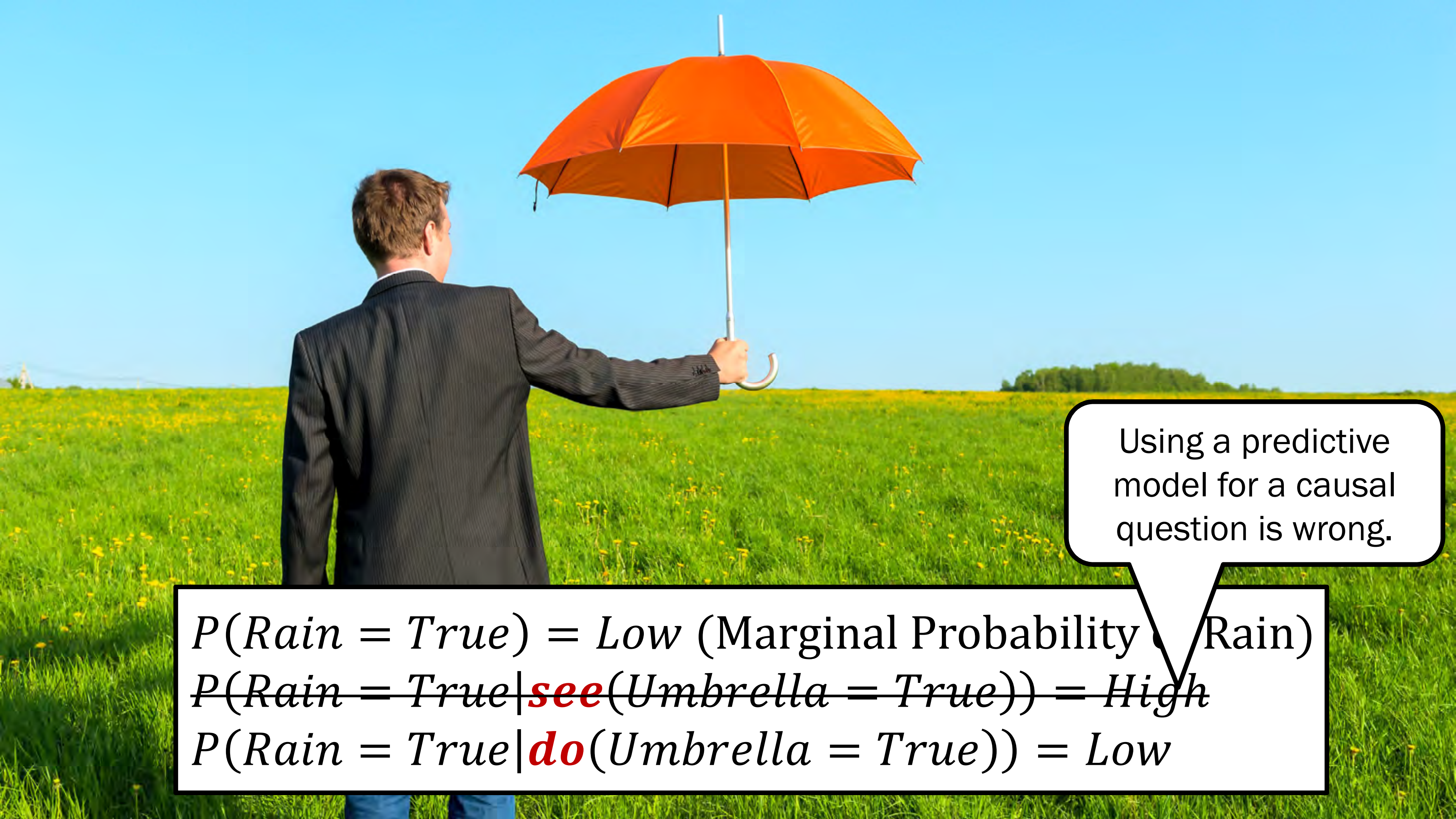
Dimensions

Reasoning

Dimensions

$$P(Rain = True) = Low \text{ (Marginal Probability of Rain)}$$

$$P(Rain = True | \textbf{see}(Umbrella = True)) = High$$



Using a predictive model for a causal question is wrong.

$P(Rain = True) = Low$ (Marginal Probability Rain)

~~$P(Rain = True|\mathbf{see}(Umbrella = True)) = High$~~

$P(Rain = True|\mathbf{do}(Umbrella = True)) = Low$

Reasoning

Predictive Modeling

Importance in Predictive Modeling

- Total Effects
- Entropy & Mutual Information
- Arc Force, Node Force
- Bayes Factor
- Tornado Chart

Note

- We are not discussing how to build or learn Bayesian network models today.
- We simply use existing models to quantify the importance of variables and their relationships.
- All of today's examples were properly introduced in other seminars, and we will provide links to those materials.

Predictive Modeling

Total Effect

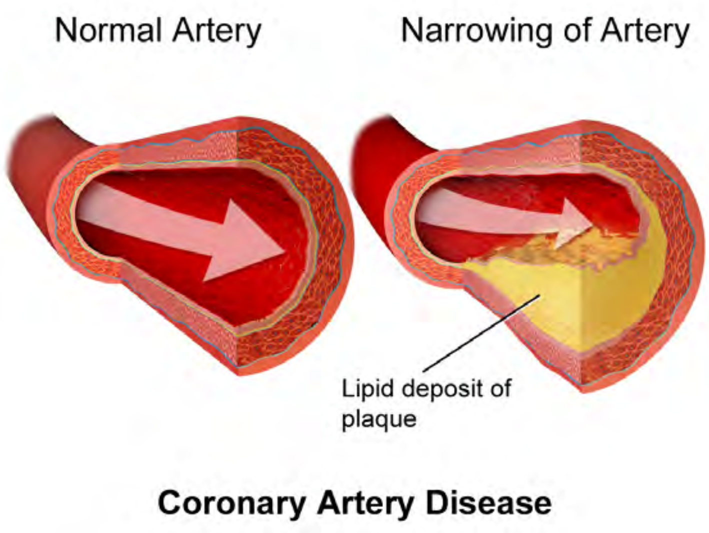
- "Given that I observe a change of one unit in variable x, how much change would I observe in variable y?"

Compare to "parameter estimates" or "coefficients" in a regression:

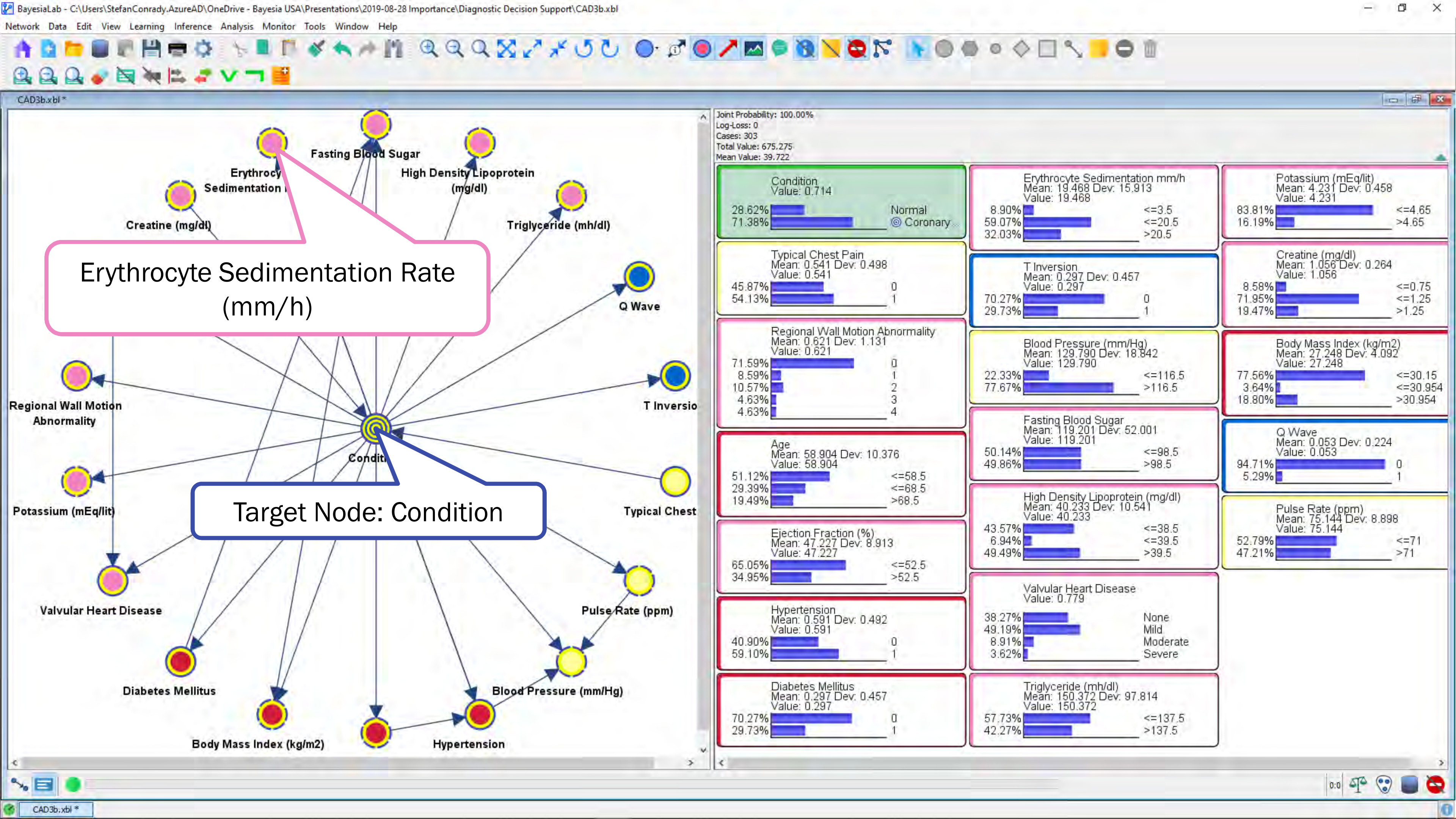
$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_n x_n$$

Predictive Modeling

Example: Diagnosing Coronary Artery Disease



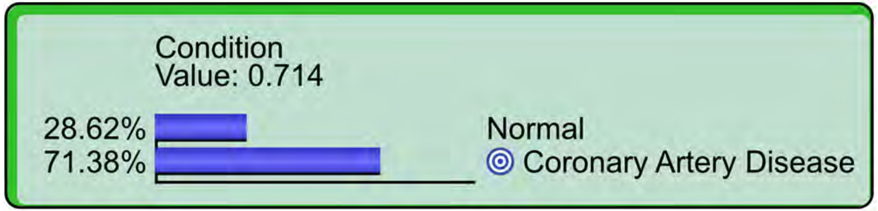
See Webinar on Diagnostic Decision Support with Bayesian Networks:
https://www.bayesia.com/webinar-diagnostic-decision-support-with-bayesian-networks


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Erythrocyte Sedimentation Rate (mm/h)
Target Node: Condition
Fasting Blood Sugar
High Density Lipoprotein (mg/dl)
Creatine (mg/dl)
Triglyceride (mh/dl)
Q Wave
T Inversio
Regional Wall Motion Abnormality
Potassium (mEq/lit)
Typical Chest
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Condition
Value: 0.714
28.62% Normal
71.38% Coronary
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Potassium (mEq/lit)
Mean: 4.231 Dev: 0.458
Value: 4.231
83.81% <=4.65
16.19% >4.65
Typical Chest Pain
Mean: 0.541 Dev: 0.498
Value: 0.541
45.87% 0
54.13% 1
T Inversion
Mean: 0.297 Dev: 0.457
Value: 0.297
70.27% 0
29.73% 1
Creatine (mg/dl)
Mean: 1.056 Dev: 0.264
Value: 1.056
8.58% <=0.75
71.95% <=1.25
19.47% >1.25
Regional Wall Motion Abnormality
Mean: 0.621 Dev: 1.131
Value: 0.621
71.59% 0
8.59% 1
10.57% 2
4.63% 3
4.63% 4
Blood Pressure (mm/Hg)
Mean: 129.790 Dev: 18.842
Value: 129.790
22.33% <=116.5
77.67% >116.5
Body Mass Index (kg/m2)
Mean: 27.248 Dev: 4.092
Value: 27.248
77.56% <=30.15
3.64% <=30.954
18.80% >30.954
Fasting Blood Sugar
Mean: 119.201 Dev: 52.001
Value: 119.201
50.14% <=98.5
49.86% >98.5
Age
Mean: 58.904 Dev: 10.376
Value: 58.904
51.12% <=58.5
29.39% <=68.5
19.49% >68.5
Q Wave
Mean: 0.053 Dev: 0.224
Value: 0.053
94.71% 0
5.29% 1
High Density Lipoprotein (mg/dl)
Mean: 40.233 Dev: 10.541
Value: 40.233
43.57% <=38.5
6.94% <=39.5
49.49% >39.5
Ejection Fraction (%)
Mean: 47.227 Dev: 8.913
Value: 47.227
65.05% <=52.5
34.95% >52.5
Pulse Rate (ppm)
Mean: 75.144 Dev: 8.898
Value: 75.144
52.79% <=71
47.21% >71
Valvular Heart Disease
Value: 0.779
38.27% None
49.19% Mild
8.91% Moderate
3.62% Severe
Hypertension
Mean: 0.591 Dev: 0.492
Value: 0.591
40.90% 0
59.10% 1
Diabetes Mellitus
Mean: 0.297 Dev: 0.457
Value: 0.297
70.27% 0
29.73% 1
Triglyceride (mh/dl)
Mean: 150.372 Dev: 97.814
Value: 150.372
57.73% <=137.5
42.27% >137.5
CAD3b.xbl *

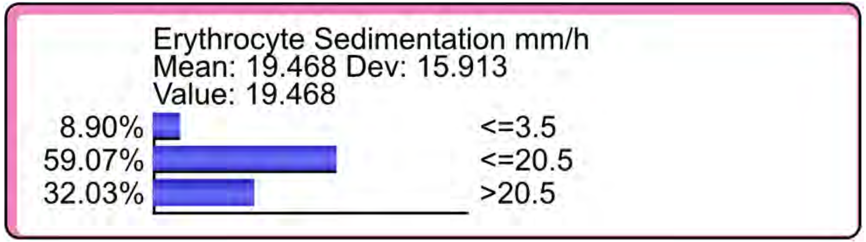
Predictive Modeling

Example: Diagnosing Coronary Artery Disease

- Target Variable: Condition (abbreviated “Cond.” or “C”)



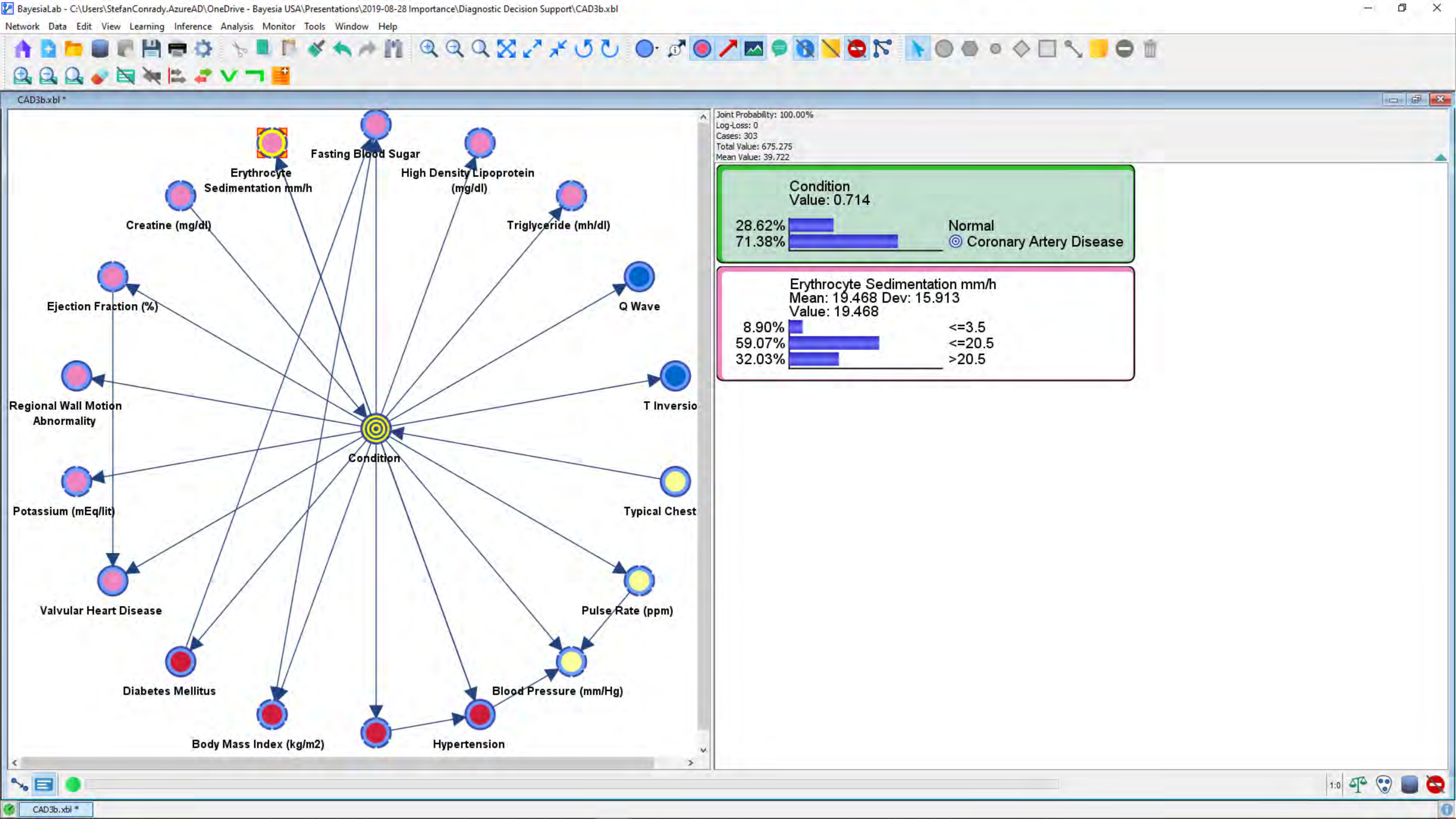
- One of 18 Predictors: Erythrocyte Sedimentation Rate (abbr. “ESR” or “E”)

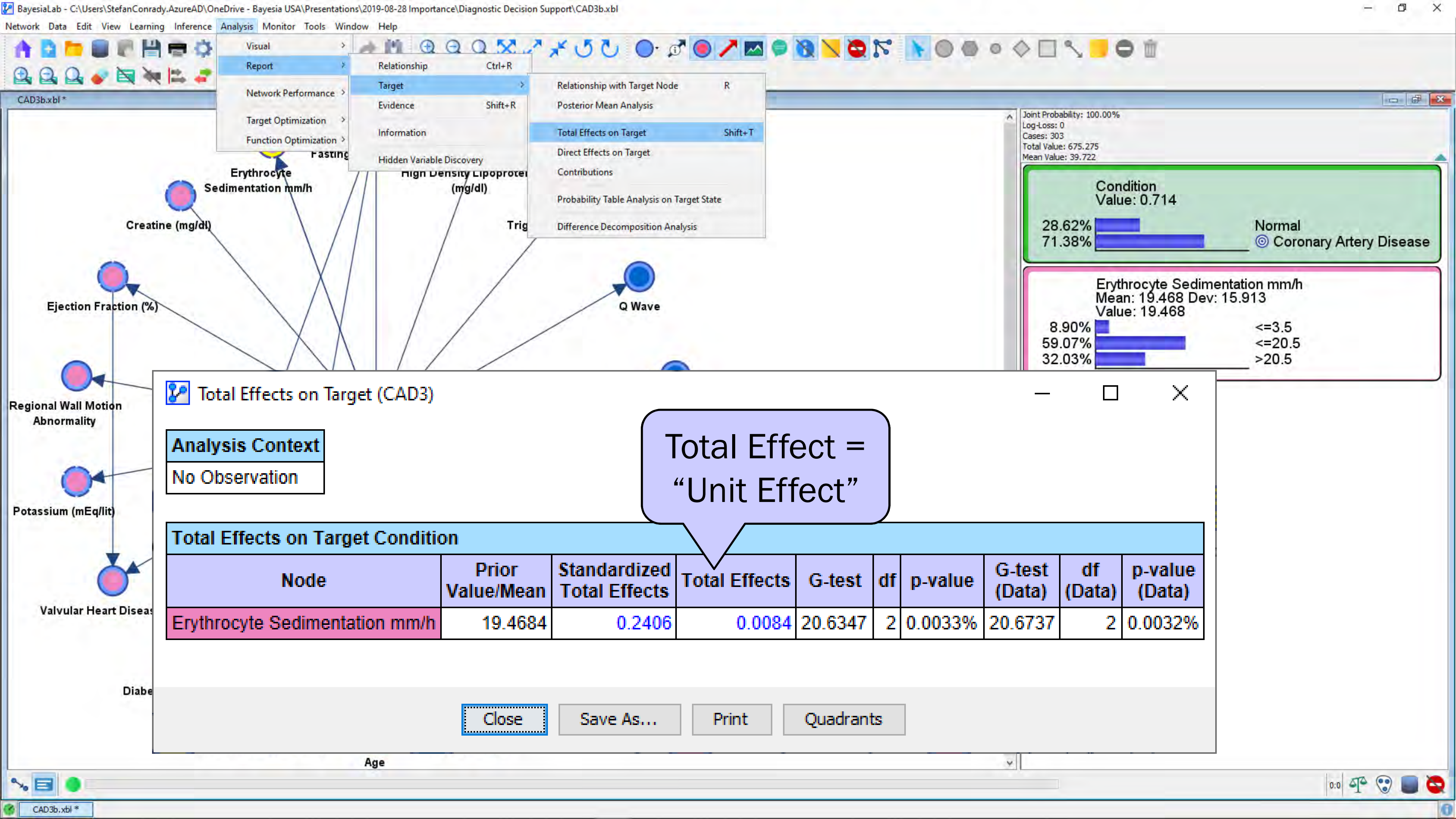


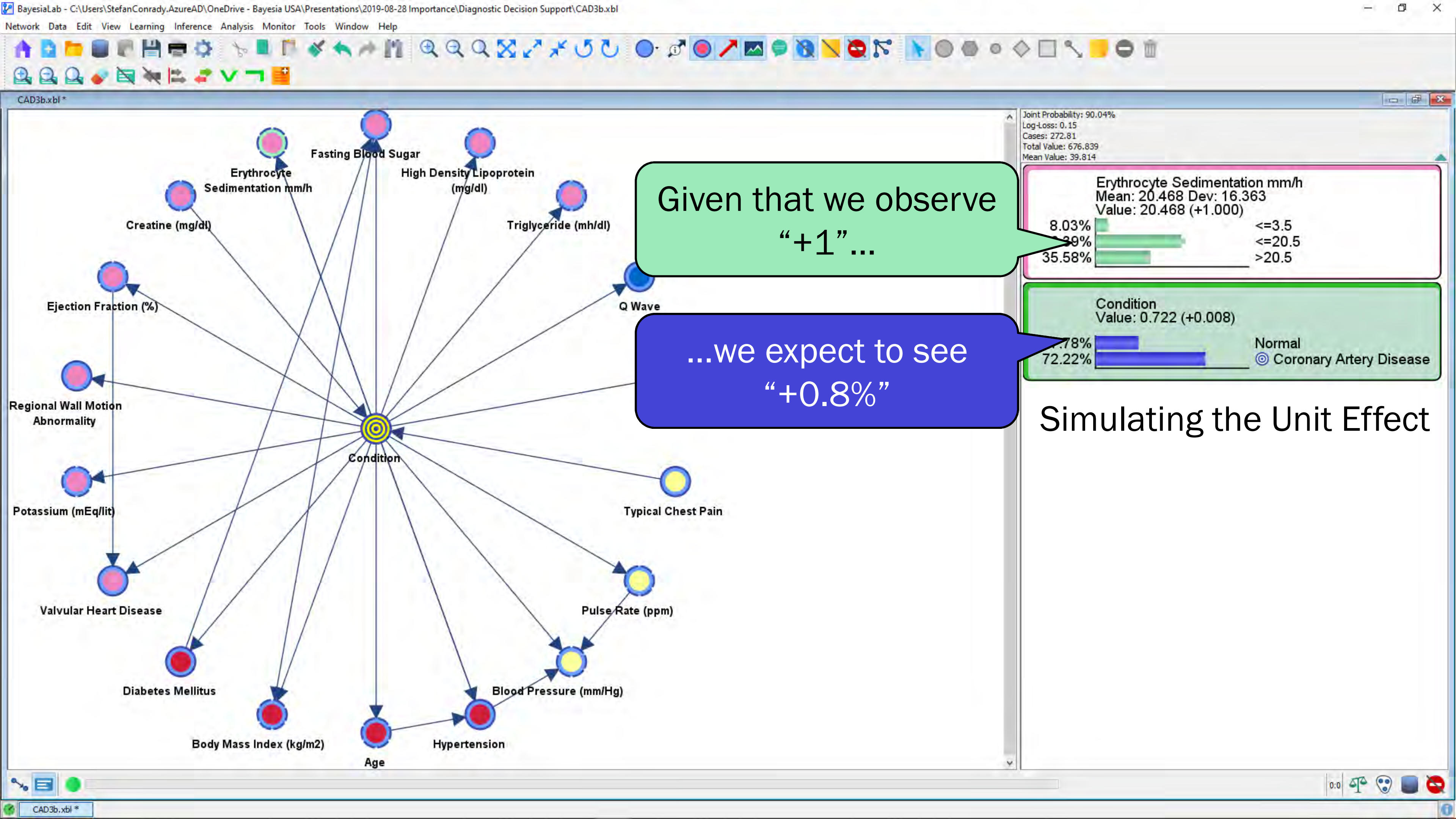
Erythrocyte Sedimentation Rate (ESR) Measurement

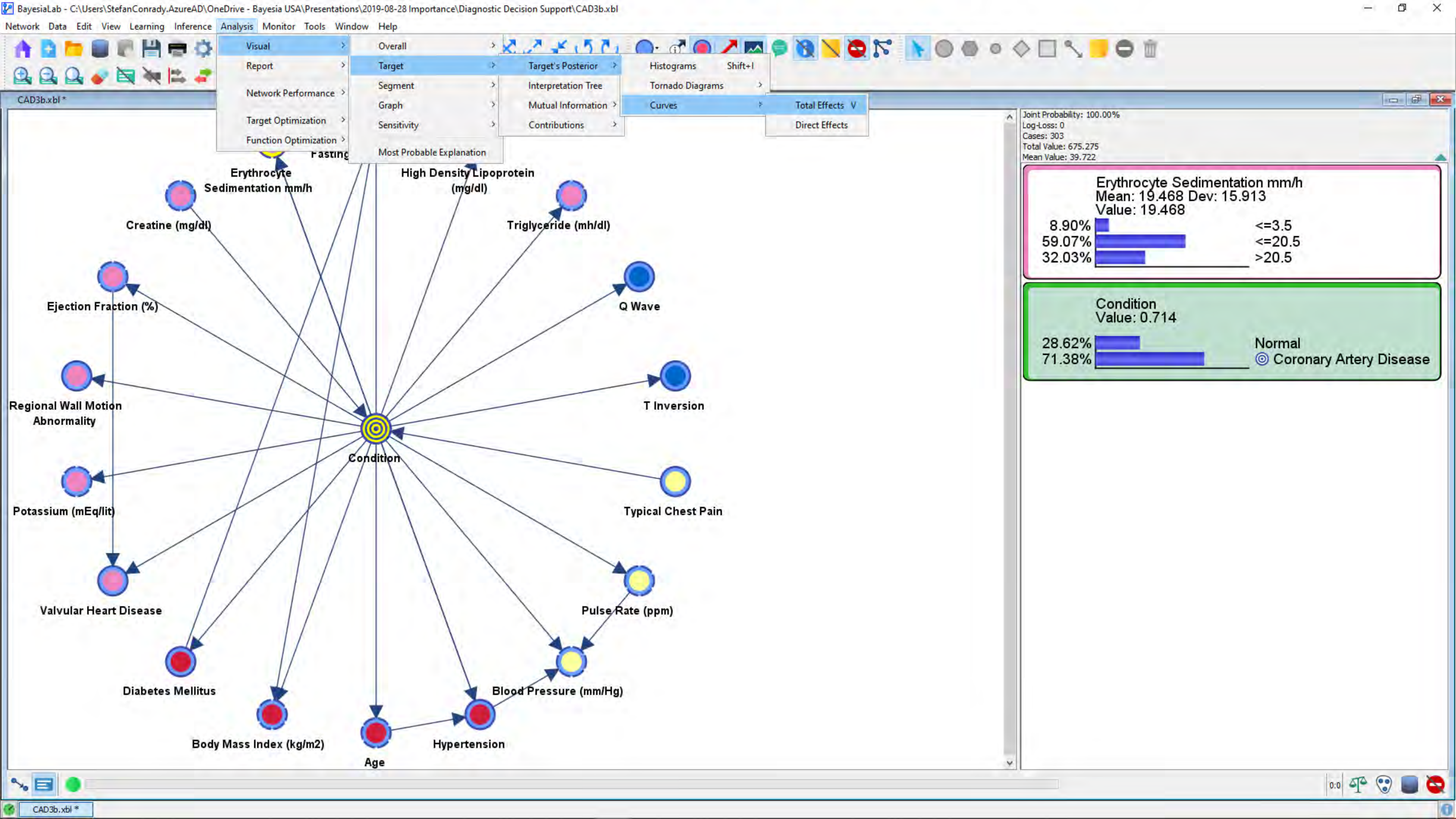
- An erythrocyte sedimentation rate (ESR) is a type of blood test that measures how quickly erythrocytes (red blood cells) settle at the bottom of a test tube that contains a blood sample.

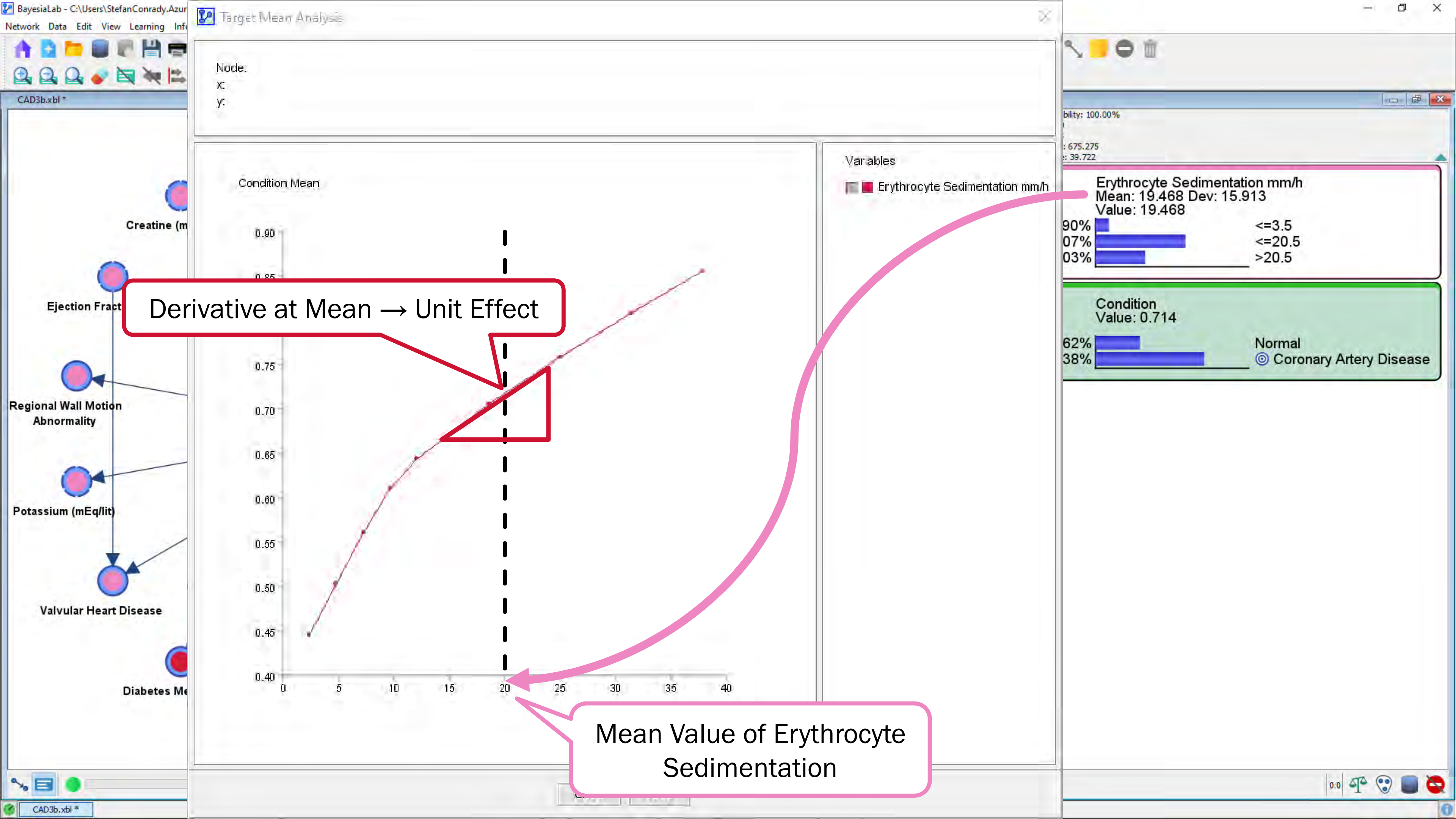
- Normally, red blood cells settle relatively slowly. A faster-than-normal rate may indicate inflammation in the body.


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Fasting Blood Sugar
Erythrocyte Sedimentation mm/h
High Density Lipoprotein (mg/dl)
Creatine (mg/dl)
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
T Inversio
Condition
Potassium (mEq/lit)
Typical Chest
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
1:0
CAD3b.xbl *

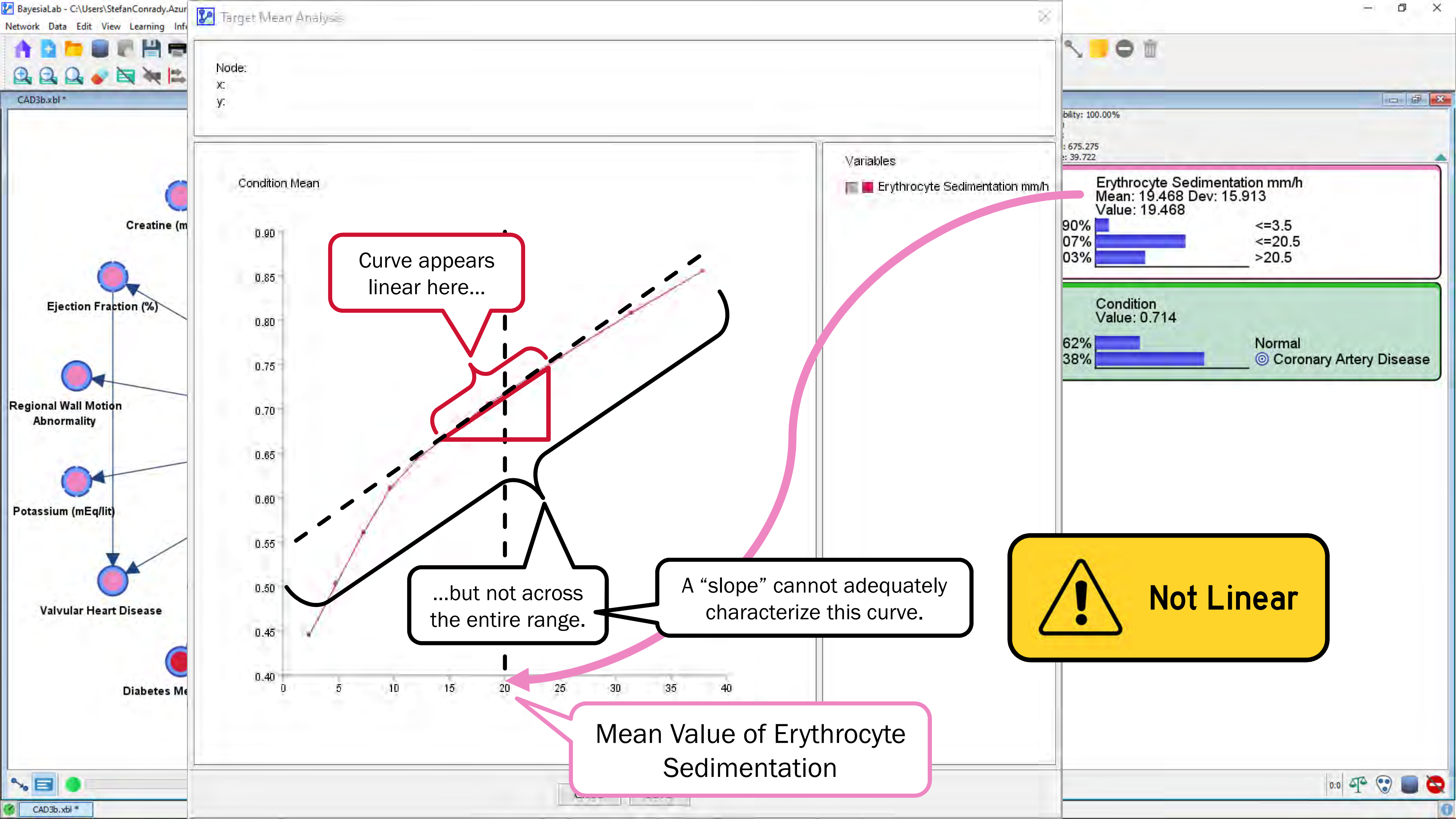

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Relationship Ctrl+R
Target
Evidence Shift+R
Information
Hidden Variable Discovery
Relationship with Target Node R
Posterior Mean Analysis
Total Effects on Target Shift+T
Direct Effects on Target
Contributions
Probability Table Analysis on Target State
Difference Decomposition Analysis
CAD3b.xbl *
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Erythrocyte Sedimentation mm/h
Creatine (mg/dl)
Fasting
High Density Lipoprotein (mg/dl)
Trig
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
Potassium (mEq/lit)
Valvular Heart Disea
Diabe
Age
Total Effects on Target (CAD3)
Analysis Context
No Observation
Total Effect =
“Unit Effect”
Total Effects on Target Condition
Node | Prior Value/Mean | Standardized Total Effects | Total Effects | G-test | df | p-value | G-test (Data) | df (Data) | p-value (Data)
Erythrocyte Sedimentation mm/h | 19.4684 | 0.2406 | 0.0084 | 20.6347 | 2 | 0.0033% | 20.6737 | 2 | 0.0032%
Close
Save As...
Print
Quadrants
CAD3b.xbl *


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Fasting Blood Sugar
Erythrocyte Sedimentation mm/h
High Density Lipoprotein (mg/dl)
Creatine (mg/dl)
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
Condition
Potassium (mEq/lit)
Typical Chest Pain
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Age
Joint Probability: 90.04%
Log-Loss: 0.15
Cases: 272.81
Total Value: 676.839
Mean Value: 39.814
Erythrocyte Sedimentation mm/h
Mean: 20.468 Dev: 16.363
Value: 20.468 (+1.000)
8.03% <=3.5
<=20.5
35.58% >20.5
Condition
Value: 0.722 (+0.008)
Normal
72.22% Coronary Artery Disease
Given that we observe "+1"…
…we expect to see "+0.8%"
Simulating the Unit Effect
CAD3b.xbl *


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Overall
Target
Segment
Graph
Sensitivity
Most Probable Explanation
Target's Posterior
Interpretation Tree
Mutual Information
Contributions
Histograms Shift+I
Tornado Diagrams
Curves
Total Effects V
Direct Effects
CAD3b.xbl *
Fasting
Erythrocyte Sedimentation mm/h
High Density Lipoprotein (mg/dl)
Creatine (mg/dl)
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
T Inversion
Condition
Potassium (mEq/lit)
Typical Chest Pain
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Age
Hypertension
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
CAD3b.xbl *



BayesiaLab - C:\Users\StefanConrady.Azur
Network Data Edit View Learning
Target Mean Analysis
Node:
x:
y:
CAD3b.xbl *
Creatine (m
Ejection Fract
Regional Wall Motion Abnormality
Potassium (mEq/lit)
Valvular Heart Disease
Diabetes Me
Condition Mean
0.90
0.85
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0
5
10
15
20
25
30
35
40
Variables
Erythrocyte Sedimentation mm/h
Derivative at Mean → Unit Effect
Mean Value of Erythrocyte Sedimentation
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
<=3.5
<=20.5
>20.5
Condition
Value: 0.714
62%
38%
Normal
Coronary Artery Disease


BayesiaLab - C:\Users\StefanConrady.Azur
Network Data Edit View Learning Inf
CAD3b.xbl *
Creatine (m
Ejection Fraction (%)
Regional Wall Motion Abnormality
Potassium (mEq/lit)
Valvular Heart Disease
Diabetes Me
Target Mean Analysis
Node:
x:
y:
Condition Mean
Variables
Erythrocyte Sedimentation mm/h
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0
5
10
15
20
25
30
35
40
Curve appears linear here...
...but not across the entire range.
A “slope” cannot adequately characterize this curve.
Mean Value of Erythrocyte Sedimentation
Not Linear
bility: 100.00%
: 675.275
: 39.722
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
90%
07%
03%
<=3.5
<=20.5
>20.5
Condition
Value: 0.714
62%
38%
Normal
Coronary Artery Disease
CAD3b.xbl *

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl

Total Effects on Target (CAD3)

Analysis Context
No Observation

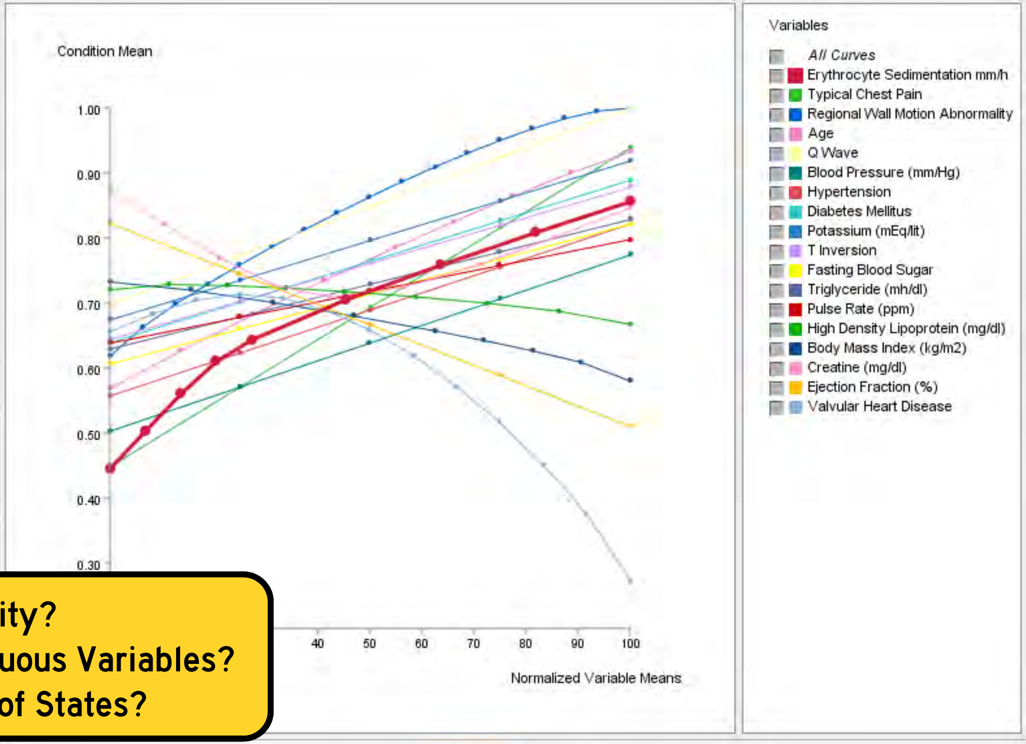
Total Effects on Target Condition

Node	Prior Value/Mean	Standardized Total Effects	Total Effects	
Typical Chest Pain	0.5413	0.5400	0.4898	9
Age	58.9041	0.3345	0.0162	3
Ejection Fraction (%)	47.2270	-0.3296	-0.0251	3
Regional Wall Motion Abnormality	0.6212	0.3136	0.1253	4
Hypertension	0.5910	0.2875	0.2643	2
Diabetes Mellitus	0.2973	0.2526	0.2498	2
Blood Pressure (mm/Hg)	129.7904	0.2497	0.0088	1
Erythrocyte Sedimentation mm/h	19.4684	0.2406	0.0084	2
Fasting Blood Sugar	119.2006	0.2384	0.0033	1
T Inversion	0.2973	0.2367	0.2340	1
Triglyceride (mh/dl)	150.3718	0.2176	0.0015	1
Potassium (mEq/lit)	4.2308	0.1992	0.2829	1
Pulse Rate (ppm)	75.1444	0.1760	0.0116	
Q Wave	0.0529	0.1496	0.3022	1
Body Mass Index (kg/m2)	27.2477			
Creatine (mg/dl)	1.0556			
High Density Lipoprotein (mg/dl)	40.2330			
Valvular Heart Disease	0.7789			

Close | Save As... | Print

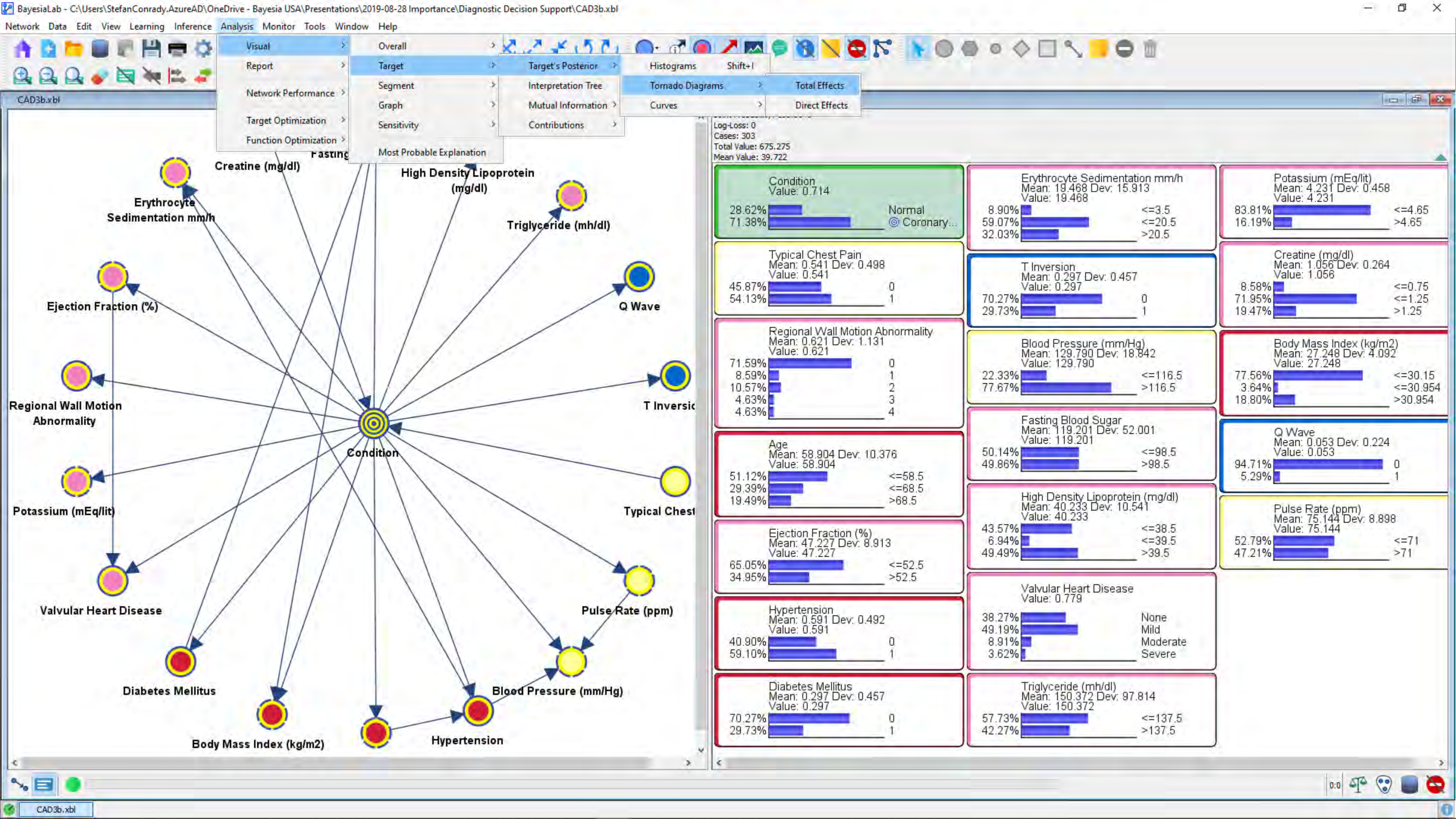
Target Mean Analysis

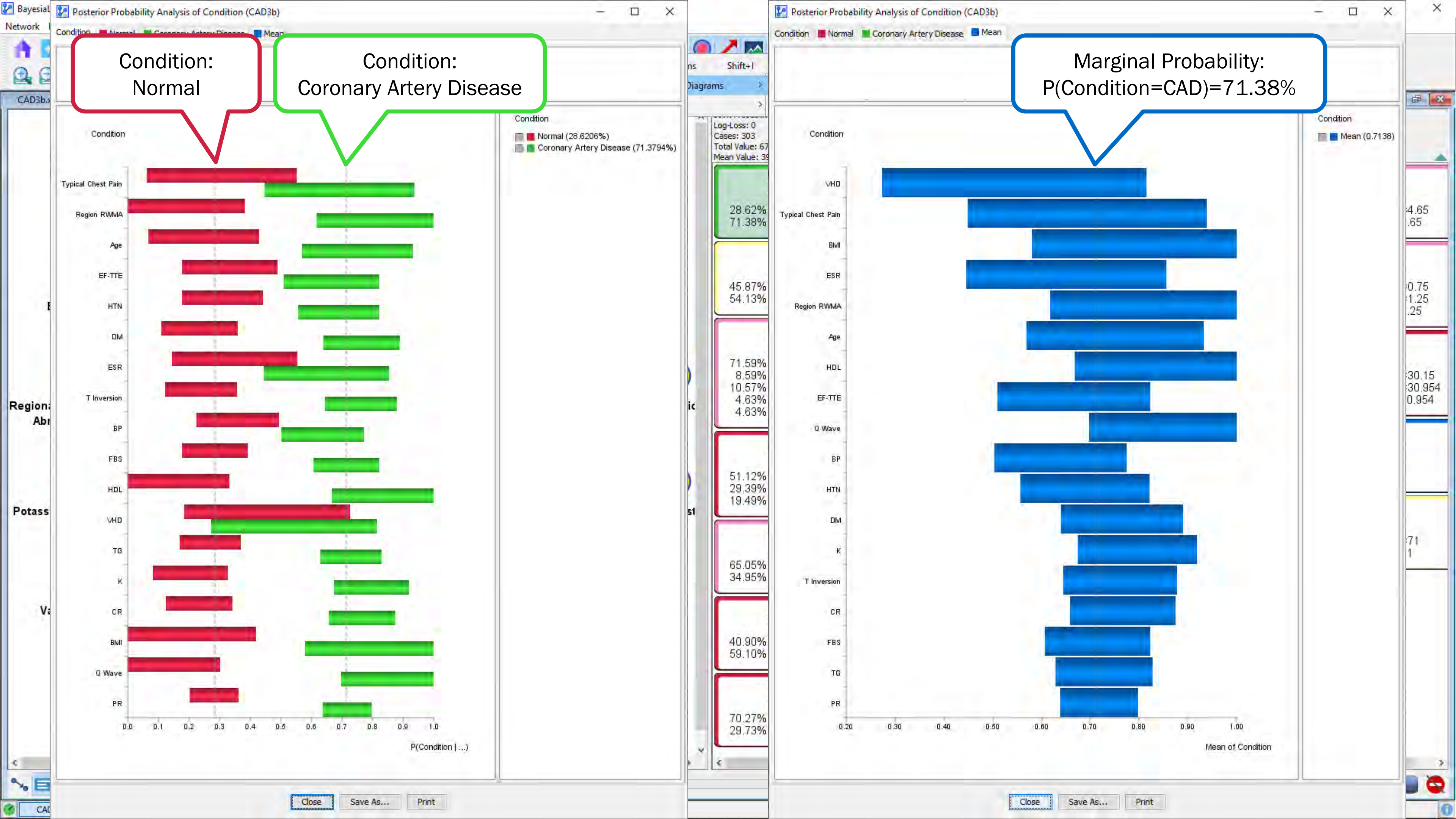
Node:
x:
y:



Close | Save

Linearity?
Continuous Variables?
Order of States?


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Overall
Target
Segment
Graph
Sensitivity
Most Probable Explanation
Target's Posterior
Interpretation Tree
Mutual Information
Contributions
Histograms Shift+I
Tornado Diagrams
Curves
Total Effects
Direct Effects
CAD3b.xbl
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Creatine (mg/dl)
Fasting
High Density Lipoprotein (mg/dl)
Erythrocyte Sedimentation mm/h
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
Condition
T Inversic
Potassium (mEq/lit)
Typical Chest
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Condition Value: 0.714
28.62% Normal
71.38% Coronary...
Erythrocyte Sedimentation mm/h Mean: 19.468 Dev: 15.913 Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Potassium (mEq/lit) Mean: 4.231 Dev: 0.458 Value: 4.231
83.81% <=4.65
16.19% >4.65
Typical Chest Pain Mean: 0.541 Dev: 0.498 Value: 0.541
45.87% 0
54.13% 1
T Inversion Mean: 0.297 Dev: 0.457 Value: 0.297
70.27% 0
29.73% 1
Creatine (mg/dl) Mean: 1.056 Dev: 0.264 Value: 1.056
8.58% <=0.75
71.95% <=1.25
19.47% >1.25
Regional Wall Motion Abnormality Mean: 0.621 Dev: 1.131 Value: 0.621
71.59% 0
8.59% 1
10.57% 2
4.63% 3
4.63% 4
Blood Pressure (mm/Hg) Mean: 129.790 Dev: 18.842 Value: 129.790
22.33% <=116.5
77.67% >116.5
Body Mass Index (kg/m2) Mean: 27.248 Dev: 4.092 Value: 27.248
77.56% <=30.15
3.64% <=30.954
18.80% >30.954
Fasting Blood Sugar Mean: 119.201 Dev: 52.001 Value: 119.201
50.14% <=98.5
49.86% >98.5
Q Wave Mean: 0.053 Dev: 0.224 Value: 0.053
94.71% 0
5.29% 1
Age Mean: 58.904 Dev: 10.376 Value: 58.904
51.12% <=58.5
29.39% <=68.5
19.49% >68.5
High Density Lipoprotein (mg/dl) Mean: 40.233 Dev: 10.541 Value: 40.233
43.57% <=38.5
6.94% <=39.5
49.49% >39.5
Pulse Rate (ppm) Mean: 75.144 Dev: 8.898 Value: 75.144
52.79% <=71
47.21% >71
Ejection Fraction (%) Mean: 47.227 Dev: 8.913 Value: 47.227
65.05% <=52.5
34.95% >52.5
Valvular Heart Disease Value: 0.779
38.27% None
49.19% Mild
8.91% Moderate
3.62% Severe
Hypertension Mean: 0.591 Dev: 0.492 Value: 0.591
40.90% 0
59.10% 1
Diabetes Mellitus Mean: 0.297 Dev: 0.457 Value: 0.297
70.27% 0
29.73% 1
Triglyceride (mh/dl) Mean: 150.372 Dev: 97.814 Value: 150.372
57.73% <=137.5
42.27% >137.5
CAD3b.xbl

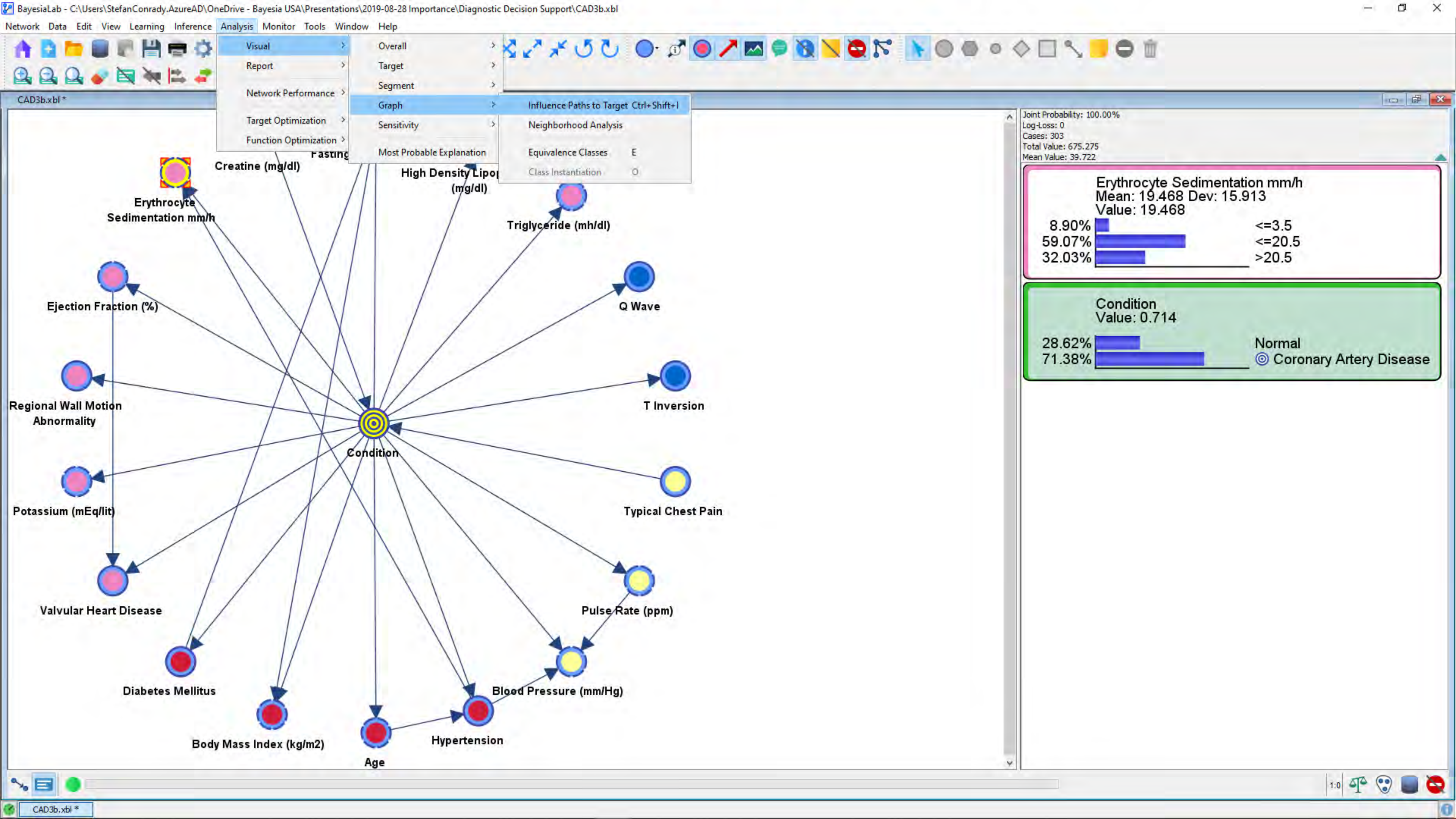


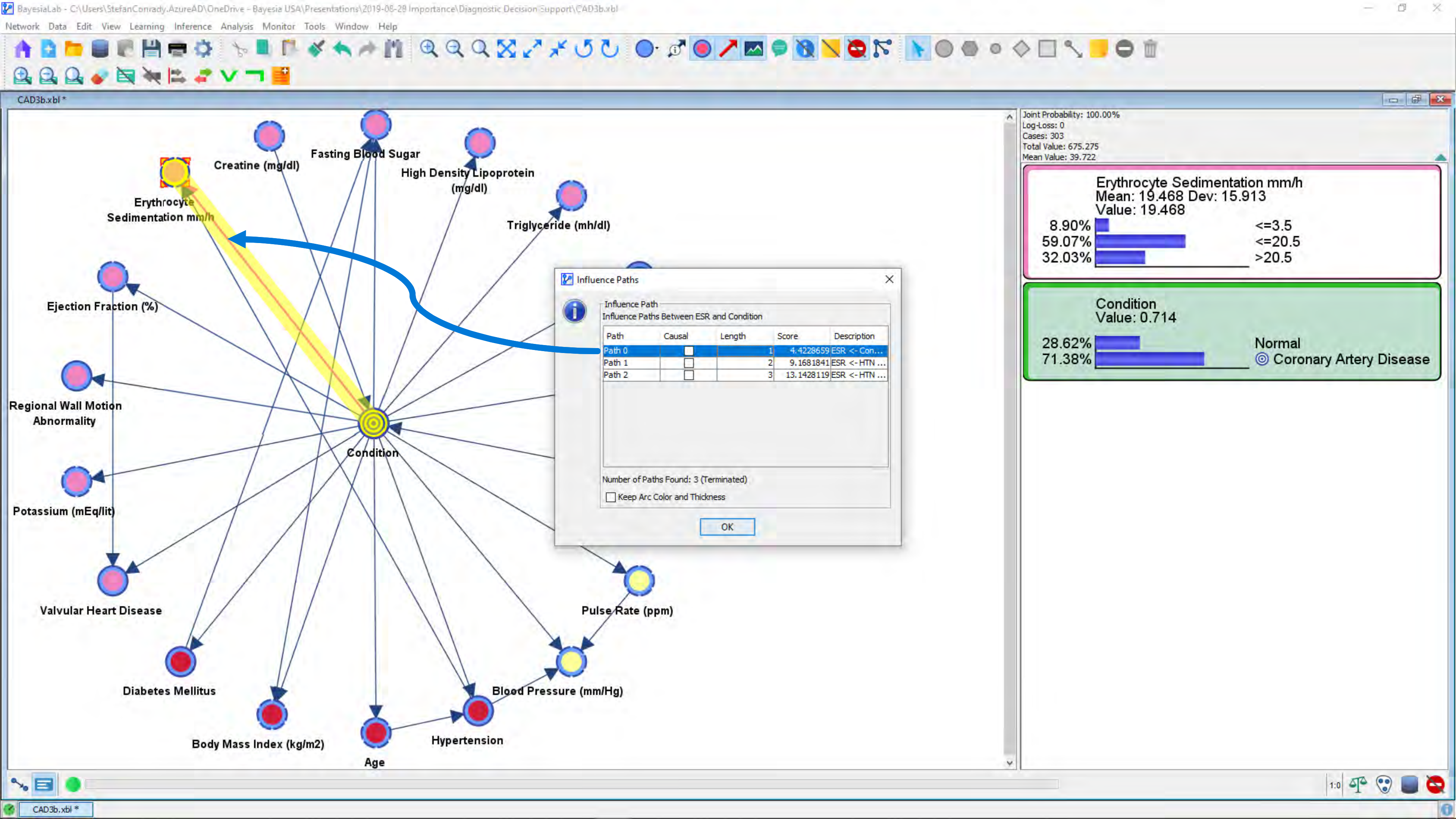
Posterior Probability Analysis of Condition (CAD3b)
Condition: Normal
Condition: Coronary Artery Disease
Condition
Normal (28.6206%)
Coronary Artery Disease (71.3794%)
Typical Chest Pain
Region RWMA
Age
EF-TTE
HTN
DM
ESR
T Inversion
BP
FBS
HDL
VHD
TG
K
CR
BMI
Q Wave
PR
P(Condition | ...)
Close
Save As...
Print
Posterior Probability Analysis of Condition (CAD3b)
Marginal Probability:
P(Condition=CAD)=71.38%
Condition
Mean (0.7138)
Mean of Condition
Close
Save As...
Print

Total Effects

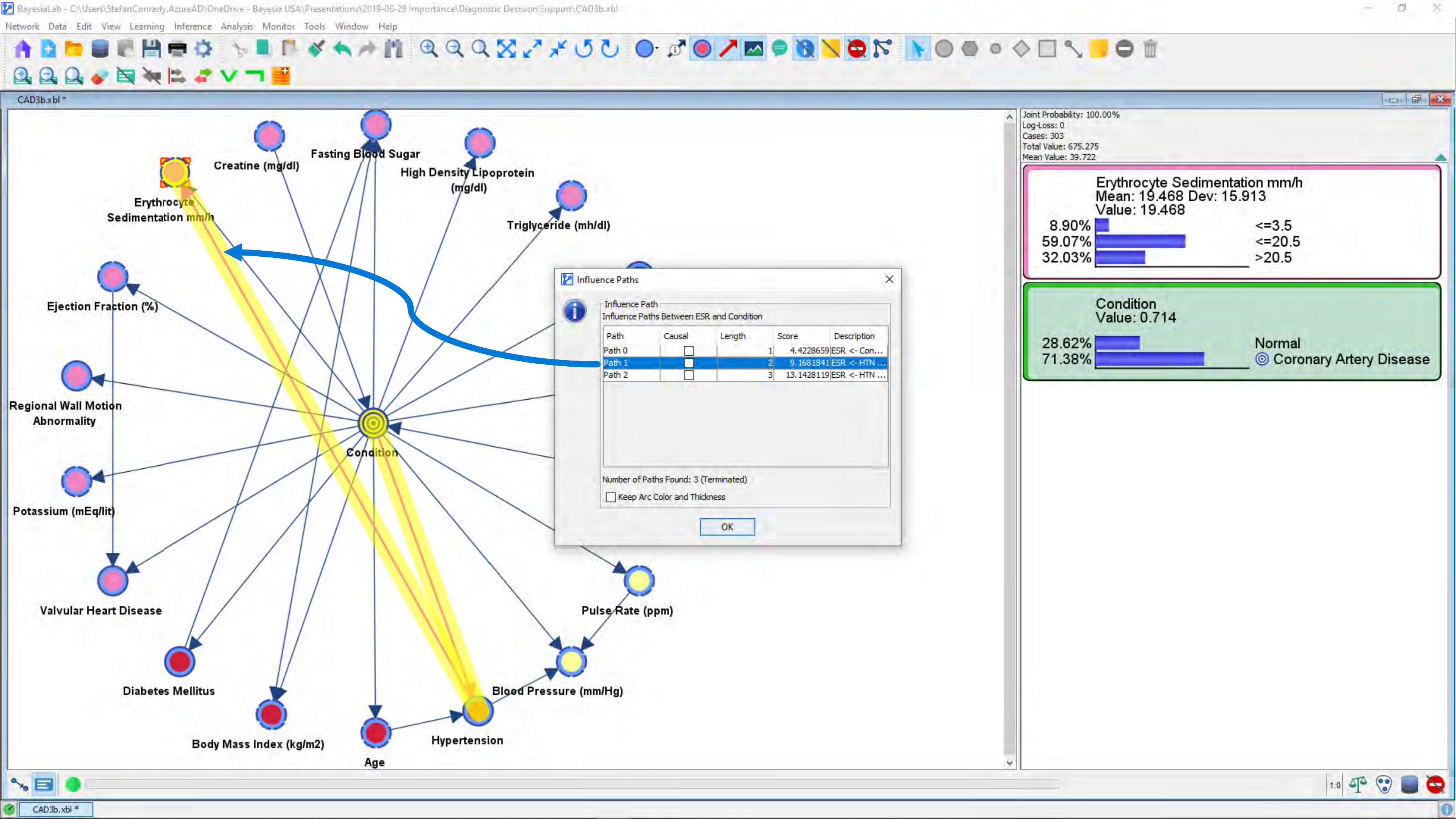
Why *"Total"* Effects?

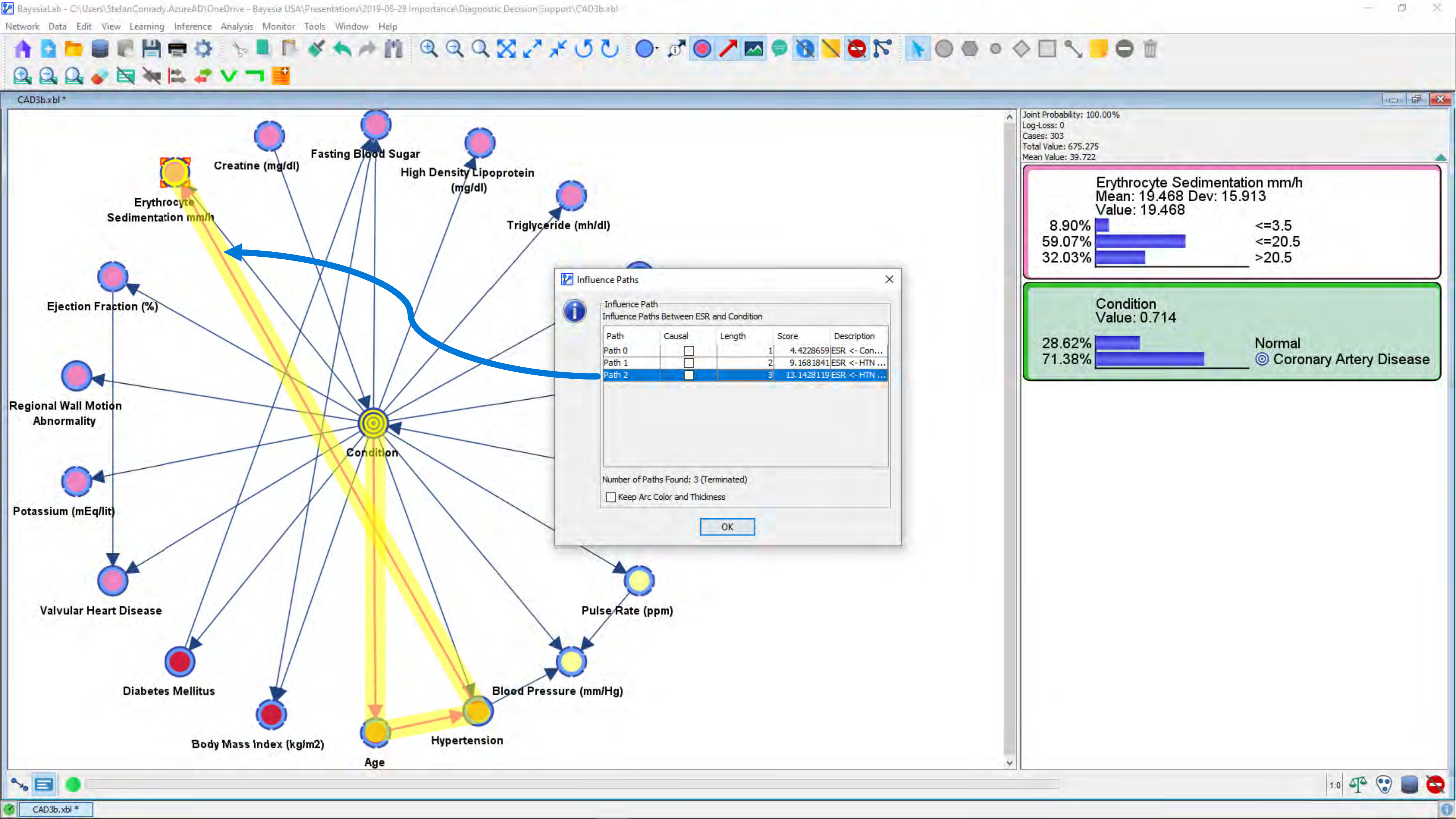
- In a Bayesian network, inference is performed in all directions, regardless of the arc direction.

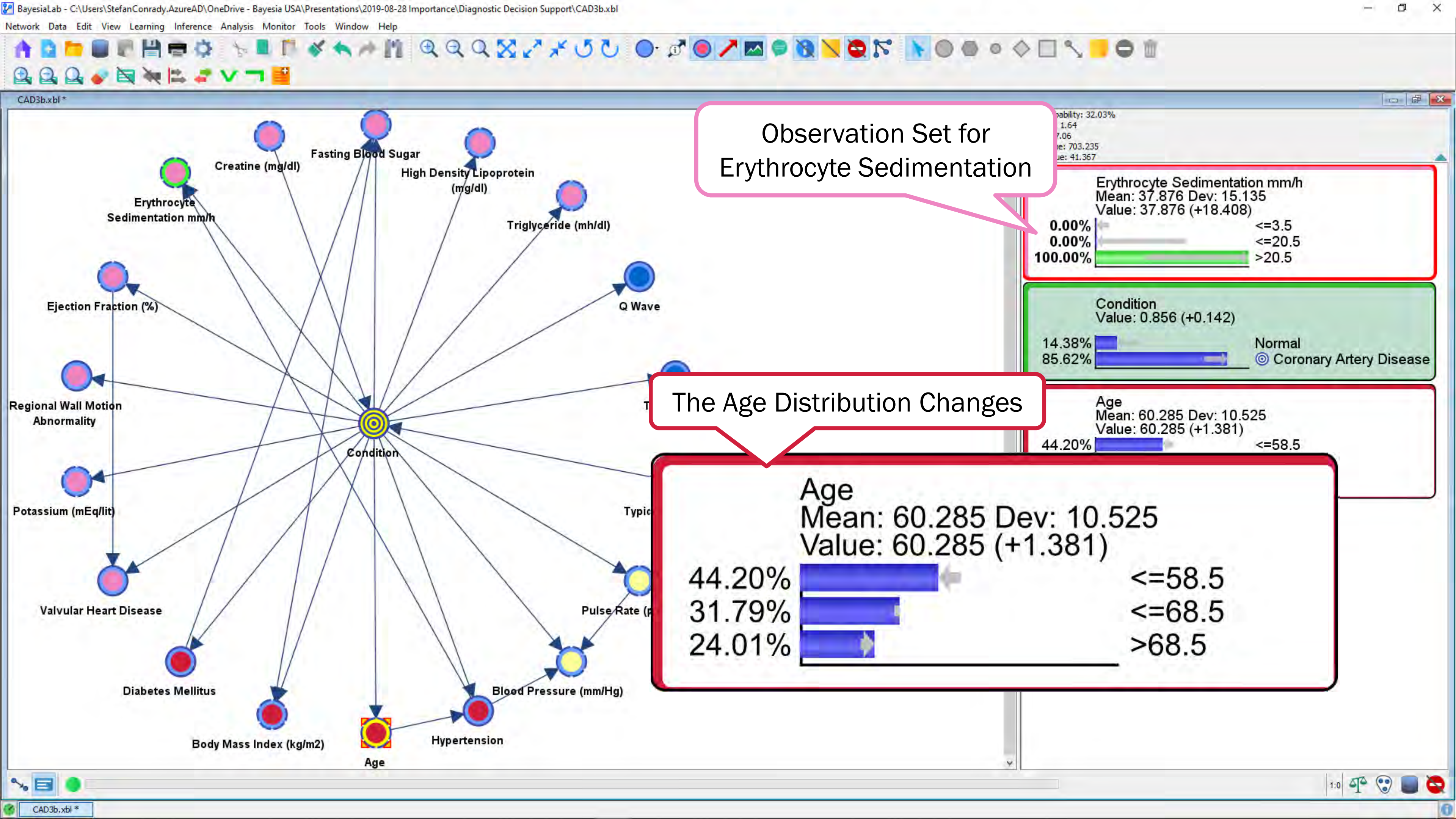

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Overall
Target
Segment
Graph
Sensitivity
Most Probable Explanation
Influence Paths to Target Ctrl+Shift+I
Neighborhood Analysis
Equivalence Classes E
Class Instantiation O
CAD3b.xbl *
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
Creatine (mg/dl)
Fasting
High Density Lipop (mg/dl)
Erythrocyte Sedimentation mm/h
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
T Inversion
Condition
Potassium (mEq/lit)
Typical Chest Pain
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Age
CAD3b.xbl *



BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-29 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Creatine (mg/dl)
Fasting Blood Sugar
High Density Lipoprotein (mg/dl)
Triglyceride (mh/dl)
Erythrocyte Sedimentation mm/h
Ejection Fraction (%)
Regional Wall Motion Abnormality
Potassium (mEq/lit)
Valvular Heart Disease
Condition
Diabetes Mellitus
Body Mass Index (kg/m2)
Age
Hypertension
Blood Pressure (mm/Hg)
Pulse Rate (ppm)
Influence Paths
Influence Path
Influence Paths Between ESR and Condition
Path | Causal | Length | Score | Description
Path 0 | | 1 | 4.4228659 | ESR <- Con...
Path 1 | | 2 | 9.1681841 | ESR <- HTN ...
Path 2 | | 3 | 13.1428119 | ESR <- HTN ...
Number of Paths Found: 3 (Terminated)
Keep Arc Color and Thickness
OK
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
CAD3b.xbl *


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-29 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Creatine (mg/dl)
Fasting Blood Sugar
High Density Lipoprotein (mg/dl)
Triglyceride (mh/dl)
Erythrocyte Sedimentation mm/h
Ejection Fraction (%)
Regional Wall Motion Abnormality
Condition
Potassium (mEq/lit)
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Age
Influence Paths
Influence Path
Influence Paths Between ESR and Condition
Path | Causal | Length | Score | Description
Path 0 | | 1 | 4.4228659 | ESR <- Con...
Path 1 | | 2 | 9.1681841 | ESR <- HTN ...
Path 2 | | 3 | 13.1428119 | ESR <- HTN ...
Number of Paths Found: 3 (Terminated)
Keep Arc Color and Thickness
OK
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
CAD3b.xbl *


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-29 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Creatine (mg/dl)
Fasting Blood Sugar
High Density Lipoprotein (mg/dl)
Triglyceride (mh/dl)
Erythrocyte Sedimentation mm/h
Ejection Fraction (%)
Regional Wall Motion Abnormality
Potassium (mEq/lit)
Valvular Heart Disease
Diabetes Mellitus
Body Mass Index (kg/m2)
Age
Hypertension
Blood Pressure (mm/Hg)
Pulse Rate (ppm)
Condition
Influence Paths
Influence Path
Influence Paths Between ESR and Condition
Path | Causal | Length | Score | Description
Path 0 | | 1 | 4.4228659 | ESR <- Con...
Path 1 | | 2 | 9.1681841 | ESR <- HTN ...
Path 2 | | 3 | 13.1428119 | ESR <- HTN ...
Number of Paths Found: 3 (Terminated)
Keep Arc Color and Thickness
OK
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 675.275
Mean Value: 39.722
Erythrocyte Sedimentation mm/h
Mean: 19.468 Dev: 15.913
Value: 19.468
8.90% <=3.5
59.07% <=20.5
32.03% >20.5
Condition
Value: 0.714
28.62% Normal
71.38% Coronary Artery Disease
CAD3b.xbl *


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD3b.xbl *
Observation Set for Erythrocyte Sedimentation
The Age Distribution Changes
Creatine (mg/dl)
Fasting Blood Sugar
High Density Lipoprotein (mg/dl)
Erythrocyte Sedimentation mm/h
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
Condition
Potassium (mEq/lit)
Valvular Heart Disease
Pulse Rate
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Age
Erythrocyte Sedimentation mm/h
Mean: 37.876 Dev: 15.135
Value: 37.876 (+18.408)
0.00% <=3.5
0.00% <=20.5
100.00% >20.5
Condition
Value: 0.856 (+0.142)
14.38% Normal
85.62% Coronary Artery Disease
Age
Mean: 60.285 Dev: 10.525
Value: 60.285 (+1.381)
44.20% <=58.5
31.79% <=68.5
24.01% >68.5
CAD3b.xbl *

Predictive Modeling

So far:

- We inferred the expected change in the mean value of a target variable given that we observed a change in a predictor variable.

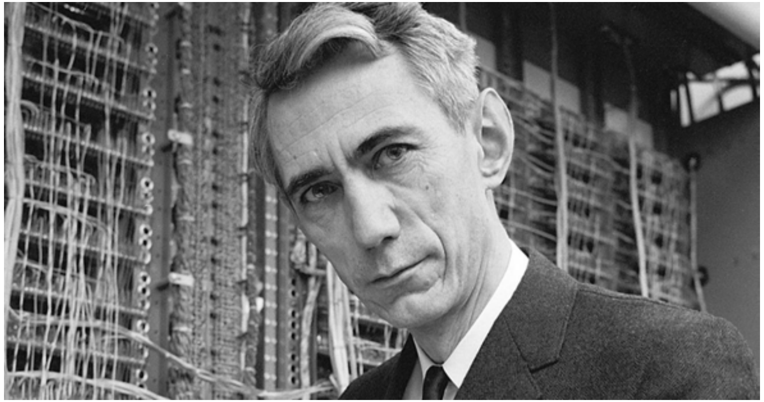
Next:

- Can we say anything about the uncertainty of one variable given another variable?

Information Theory

"Information is the resolution of uncertainty."

Claude Shannon, 1948

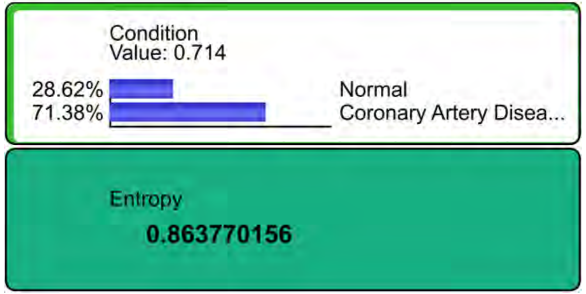


Claude Shannon (1916-2001)

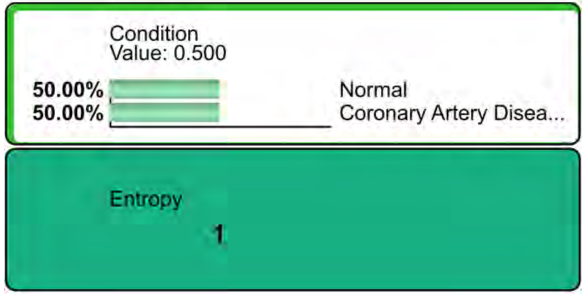
Information Theory

Entropy, a Measure of "Uncertainty"

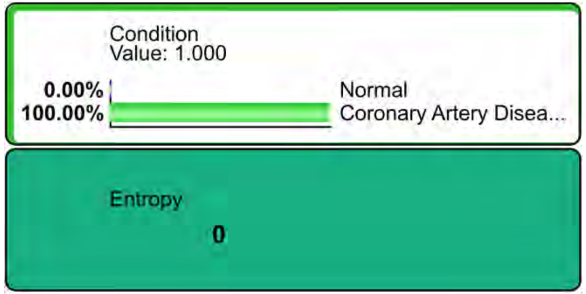
$$H(X) = -\sum_{x \in X} p(x) \times \log_2 p(x)$$



Marginal Entropy



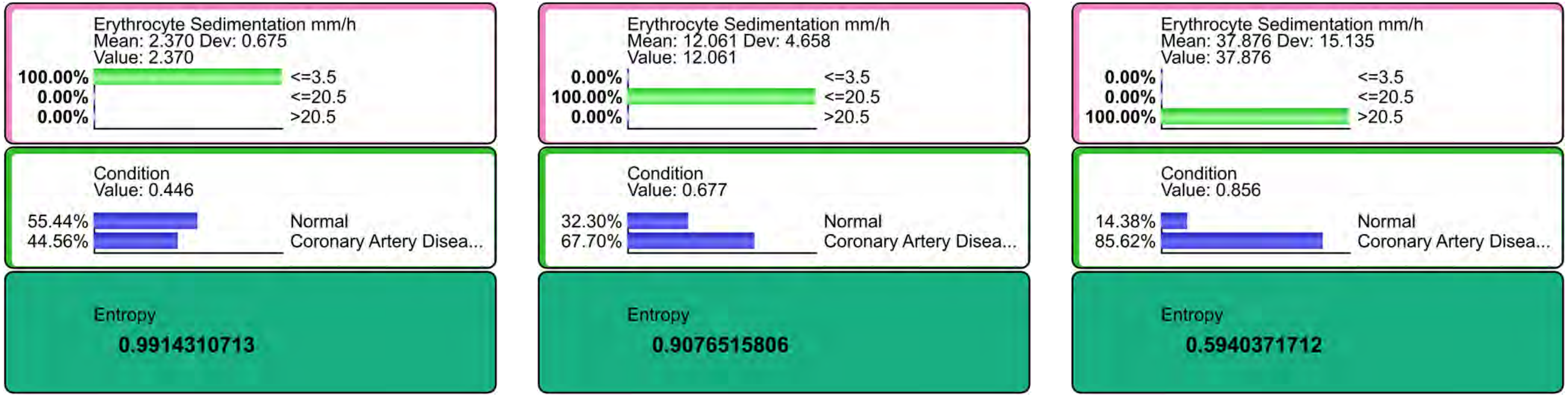
Maximum Entropy



Minimum Entropy

Information Theory

Conditional Entropy



$$H(\text{Condition} \mid \text{Erythrocyte Sedimentation}) = 0.815$$

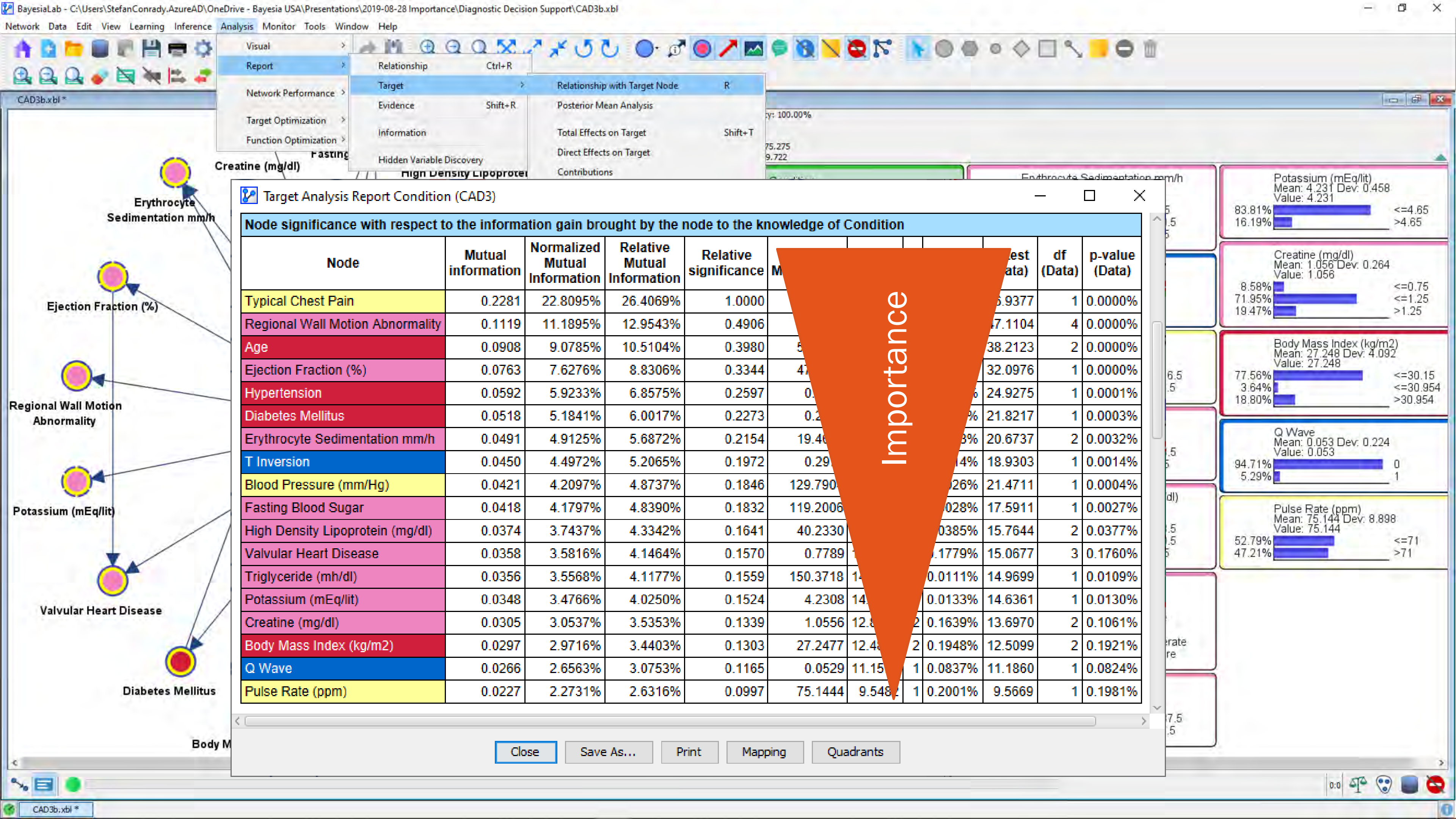
Information Theory

Mutual Information

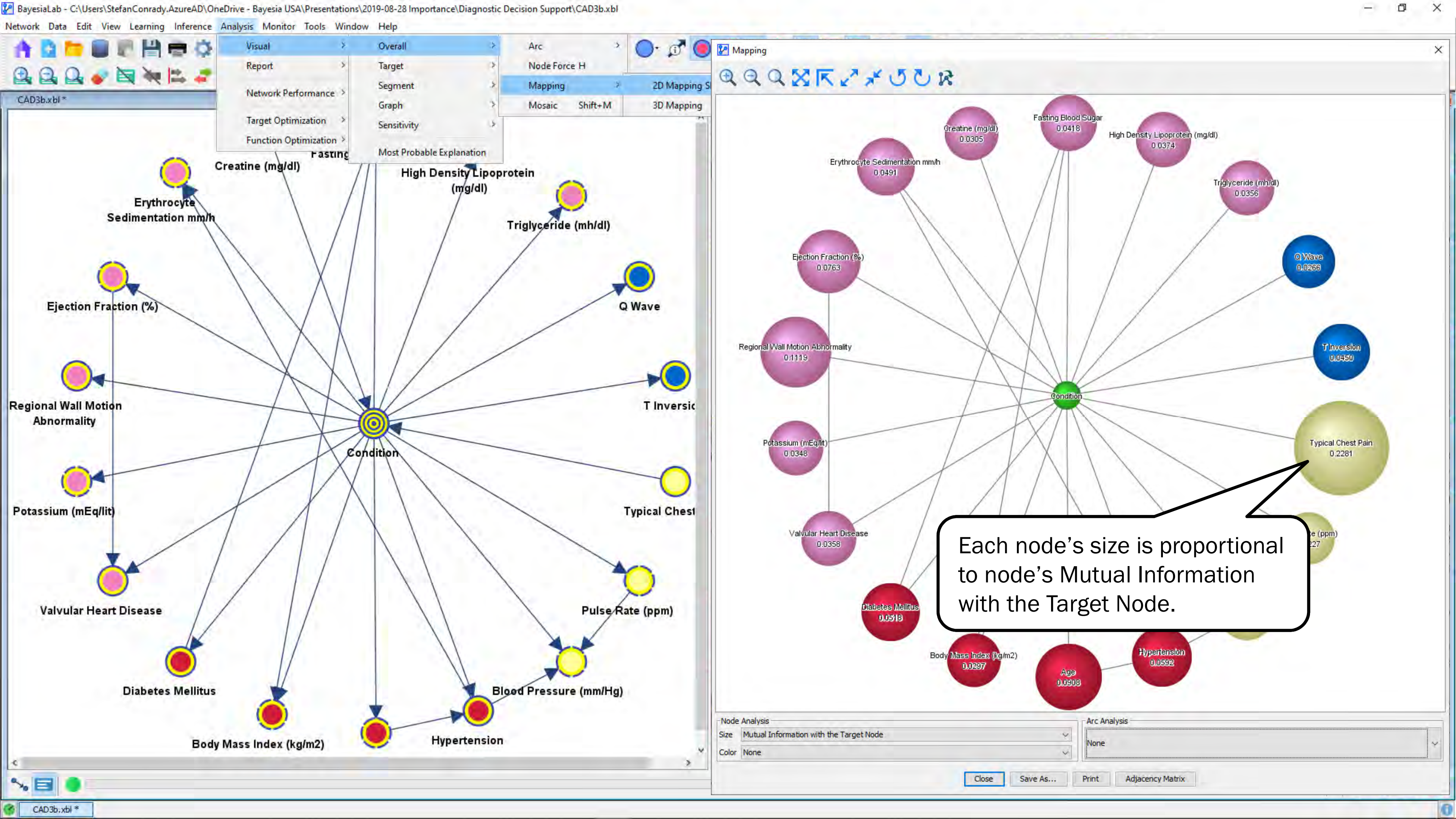
$$\underbrace{I(\text{Condition, ESR})}_{\substack{\text{Mutual Information} \\ 0.0494}} = \underbrace{H(\text{Condition})}_{\substack{\text{Marginal Entropy} \\ 0.864}} - \underbrace{H(\text{Condition}|\text{ESR})}_{\substack{\text{Conditional Entropy} \\ 0.815}}$$

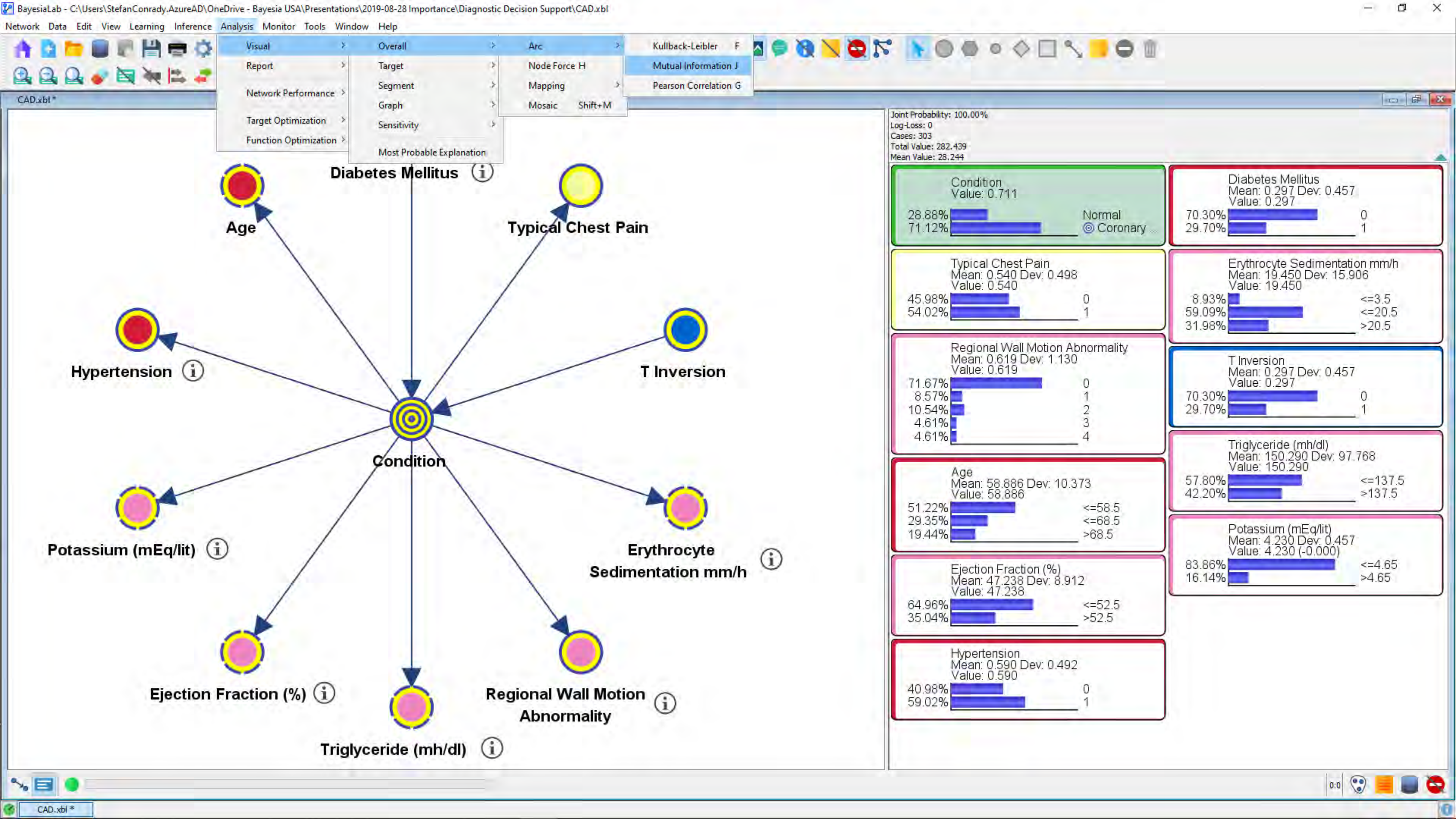
This is the amount of information Condition and ESR have in common.

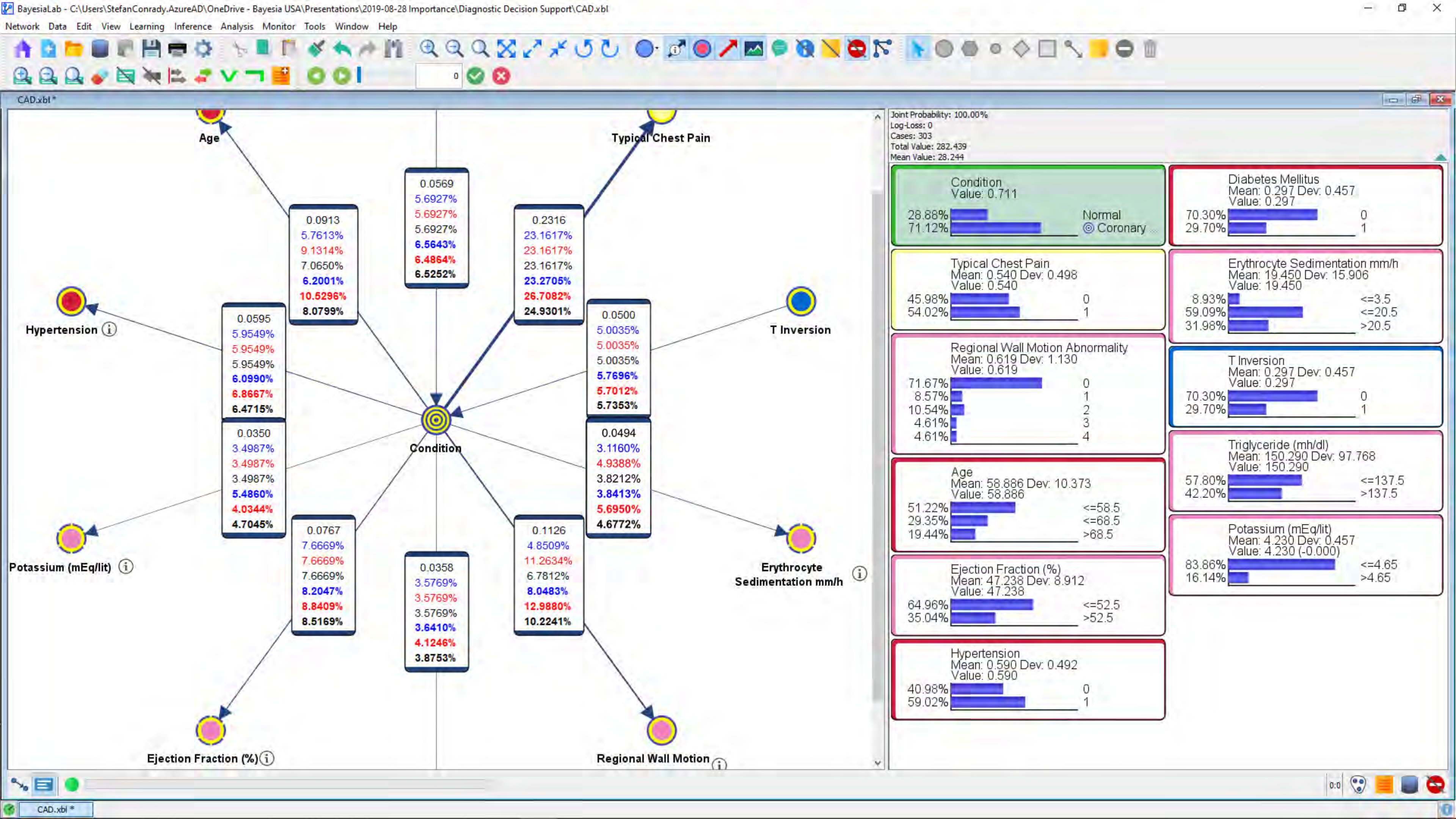
Mutual Information is a symmetrical metric.

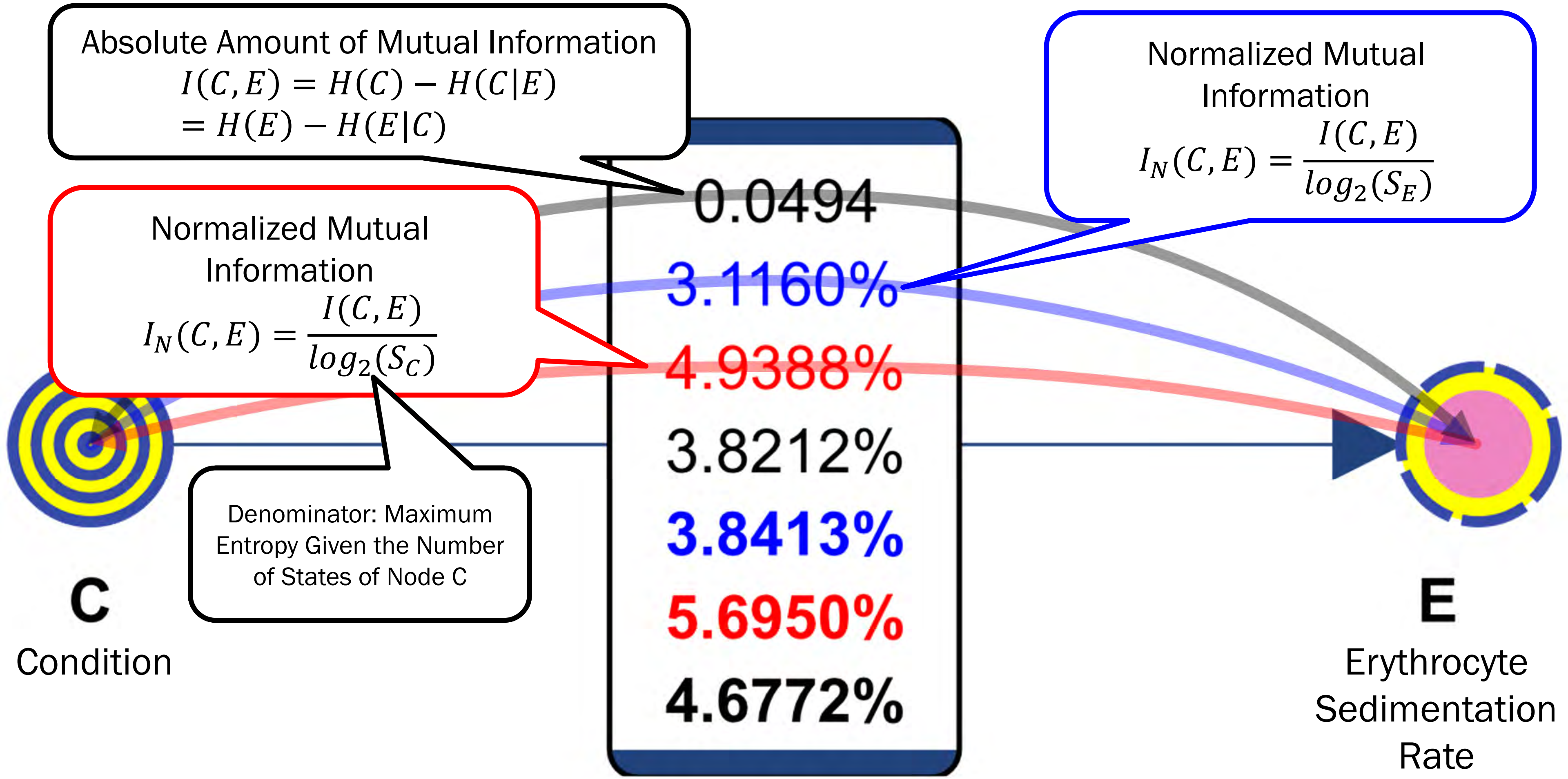


Node	Mutual information	Normalized Mutual Information	Relative Mutual Information	Relative significance	G-test (Data)	df (Data)	p-value (Data)
Typical Chest Pain	0.2281	22.8095%	26.4069%	1.0000	5.9377	1	0.0000%
Regional Wall Motion Abnormality	0.1119	11.1895%	12.9543%	0.4906	7.1104	4	0.0000%
Age	0.0908	9.0785%	10.5104%	0.3980	38.2123	2	0.0000%
Ejection Fraction (%)	0.0763	7.6276%	8.8306%	0.3344	32.0976	1	0.0000%
Hypertension	0.0592	5.9233%	6.8575%	0.2597	24.9275	1	0.0001%
Diabetes Mellitus	0.0518	5.1841%	6.0017%	0.2273	21.8217	1	0.0003%
Erythrocyte Sedimentation mm/h	0.0491	4.9125%	5.6872%	0.2154	20.6737	2	0.0032%
T Inversion	0.0450	4.4972%	5.2065%	0.1972	18.9303	1	0.0014%
Blood Pressure (mm/Hg)	0.0421	4.2097%	4.8737%	0.1846	21.4711	1	0.0004%
Fasting Blood Sugar	0.0418	4.1797%	4.8390%	0.1832	17.5911	1	0.0027%
High Density Lipoprotein (mg/dl)	0.0374	3.7437%	4.3342%	0.1641	15.7644	2	0.0377%
Valvular Heart Disease	0.0358	3.5816%	4.1464%	0.1570	15.0677	3	0.1760%
Triglyceride (mh/dl)	0.0356	3.5568%	4.1177%	0.1559	14.9699	1	0.0109%
Potassium (mEq/lit)	0.0348	3.4766%	4.0250%	0.1524	14.6361	1	0.0130%
Creatine (mg/dl)	0.0305	3.0537%	3.5353%	0.1339	13.6970	2	0.1061%
Body Mass Index (kg/m2)	0.0297	2.9716%	3.4403%	0.1303	12.5099	2	0.1921%
Q Wave	0.0266	2.6563%	3.0753%	0.1165	11.1860	1	0.0824%
Pulse Rate (ppm)	0.0227	2.2731%	2.6316%	0.0997	9.5669	1	0.1981%


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD3b.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Overall
Target
Segment
Graph
Sensitivity
Most Probable Explanation
Arc
Node Force H
Mapping
Mosaic Shift+M
2D Mapping
3D Mapping
CAD3b.xbl *
Creatine (mg/dl)
Fasting
High Density Lipoprotein (mg/dl)
Erythrocyte Sedimentation mm/h
Triglyceride (mh/dl)
Ejection Fraction (%)
Q Wave
Regional Wall Motion Abnormality
T Inversic
Condition
Potassium (mEq/lit)
Typical Chest
Valvular Heart Disease
Pulse Rate (ppm)
Diabetes Mellitus
Blood Pressure (mm/Hg)
Body Mass Index (kg/m2)
Hypertension
Mapping
Creatine (mg/dl) 0.0305
Fasting Blood Sugar 0.0418
High Density Lipoprotein (mg/dl) 0.0374
Erythrocyte Sedimentation mm/h 0.0491
Triglyceride (mh/dl) 0.0356
Ejection Fraction (%) 0.0763
Q Wave 0.0266
Regional Wall Motion Abnormality 0.1119
T Inversion 0.0450
Condition
Potassium (mEq/lit) 0.0348
Typical Chest Pain 0.2281
Valvular Heart Disease 0.0358
Diabetes Mellitus 0.0518
Body Mass Index (kg/m2) 0.0297
Age 0.0908
Hypertension 0.0592
Each node's size is proportional to node's Mutual Information with the Target Node.
Node Analysis
Size Mutual Information with the Target Node
Color None
Arc Analysis
None
Close
Save As...
Print
Adjacency Matrix
CAD3b.xbl *

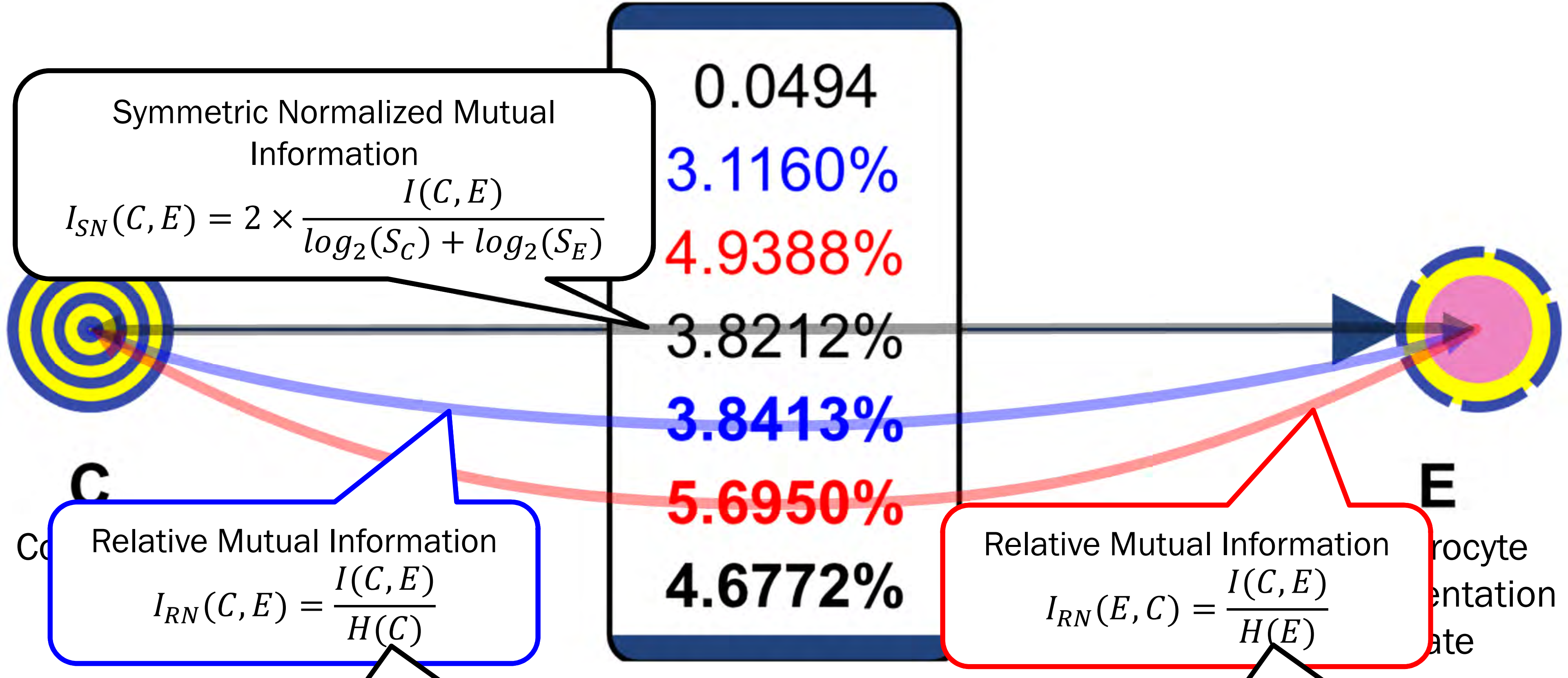

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD.xbl
Network
Data
Edit
View
Learning
Inference
Analysis
Monitor
Tools
Window
Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Overall
Target
Segment
Graph
Sensitivity
Most Probable Explanation
Arc
Node Force H
Mapping
Mosaic Shift+M
Kullback-Leibler F
Mutual Information J
Pearson Correlation G
CAD.xbl *
Diabetes Mellitus
Age
Typical Chest Pain
Hypertension
T Inversion
Condition
Potassium (mEq/lit)
Erythrocyte Sedimentation mm/h
Ejection Fraction (%)
Regional Wall Motion Abnormality
Triglyceride (mh/dl)
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 282.439
Mean Value: 28.244
Condition
Value: 0.711
28.88% Normal
71.12% Coronary
Diabetes Mellitus
Mean: 0.297 Dev: 0.457
Value: 0.297
70.30% 0
29.70% 1
Typical Chest Pain
Mean: 0.540 Dev: 0.498
Value: 0.540
45.98% 0
54.02% 1
Erythrocyte Sedimentation mm/h
Mean: 19.450 Dev: 15.906
Value: 19.450
8.93% <=3.5
59.09% <=20.5
31.98% >20.5
Regional Wall Motion Abnormality
Mean: 0.619 Dev: 1.130
Value: 0.619
71.67% 0
8.57% 1
10.54% 2
4.61% 3
4.61% 4
T Inversion
Mean: 0.297 Dev: 0.457
Value: 0.297
70.30% 0
29.70% 1
Triglyceride (mh/dl)
Mean: 150.290 Dev: 97.768
Value: 150.290
57.80% <=137.5
42.20% >137.5
Age
Mean: 58.886 Dev: 10.373
Value: 58.886
51.22% <=58.5
29.35% <=68.5
19.44% >68.5
Potassium (mEq/lit)
Mean: 4.230 Dev: 0.457
Value: 4.230 (-0.000)
83.86% <=4.65
16.14% >4.65
Ejection Fraction (%)
Mean: 47.238 Dev: 8.912
Value: 47.238
64.96% <=52.5
35.04% >52.5
Hypertension
Mean: 0.590 Dev: 0.492
Value: 0.590
40.98% 0
59.02% 1
CAD.xbl *


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Diagnostic Decision Support\CAD.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
CAD.xbl *
Age
Typical Chest Pain
Hypertension
T Inversion
Condition
Potassium (mEq/lit)
Erythrocyte Sedimentation mm/h
Ejection Fraction (%)
Regional Wall Motion
0.0913 5.7613% 9.1314% 7.0650% 6.2001% 10.5296% 8.0799%
0.0569 5.6927% 5.6927% 5.6927% 6.5643% 6.4864% 6.5252%
0.2316 23.1617% 23.1617% 23.1617% 23.2705% 26.7082% 24.9301%
0.0595 5.9549% 5.9549% 5.9549% 6.0990% 6.8667% 6.4715%
0.0500 5.0035% 5.0035% 5.0035% 5.7696% 5.7012% 5.7353%
0.0350 3.4987% 3.4987% 3.4987% 5.4860% 4.0344% 4.7045%
0.0494 3.1160% 4.9388% 3.8212% 3.8413% 5.6950% 4.6772%
0.0767 7.6669% 7.6669% 7.6669% 8.2047% 8.8409% 8.5169%
0.0358 3.5769% 3.5769% 3.5769% 3.6410% 4.1246% 3.8753%
0.1126 4.8509% 11.2634% 6.7812% 8.0483% 12.9880% 10.2241%
Joint Probability: 100.00%
Log-Loss: 0
Cases: 303
Total Value: 282.439
Mean Value: 28.244
Condition
Value: 0.711
28.88% Normal
71.12% Coronary
Diabetes Mellitus
Mean: 0.297 Dev: 0.457
Value: 0.297
70.30% 0
29.70% 1
Typical Chest Pain
Mean: 0.540 Dev: 0.498
Value: 0.540
45.98% 0
54.02% 1
Erythrocyte Sedimentation mm/h
Mean: 19.450 Dev: 15.906
Value: 19.450
8.93% <=3.5
59.09% <=20.5
31.98% >20.5
Regional Wall Motion Abnormality
Mean: 0.619 Dev: 1.130
Value: 0.619
71.67% 0
8.57% 1
10.54% 2
4.61% 3
4.61% 4
T Inversion
Mean: 0.297 Dev: 0.457
Value: 0.297
70.30% 0
29.70% 1
Triglyceride (mh/dl)
Mean: 150.290 Dev: 97.768
Value: 150.290
57.80% <=137.5
42.20% >137.5
Age
Mean: 58.886 Dev: 10.373
Value: 58.886
51.22% <=58.5
29.35% <=68.5
19.44% >68.5
Potassium (mEq/lit)
Mean: 4.230 Dev: 0.457
Value: 4.230 (-0.000)
83.86% <=4.65
16.14% >4.65
Ejection Fraction (%)
Mean: 47.238 Dev: 8.912
Value: 47.238
64.96% <=52.5
35.04% >52.5
Hypertension
Mean: 0.590 Dev: 0.492
Value: 0.590
40.98% 0
59.02% 1
CAD.xbl *



Absolute Amount of Mutual Information
$I(C,E) = H(C) - H(C|E)$
$= H(E) - H(E|C)$
Normalized Mutual Information
$I_N(C,E) = \frac{I(C,E)}{log_2(S_E)}$
Normalized Mutual Information
$I_N(C,E) = \frac{I(C,E)}{log_2(S_C)}$
Denominator: Maximum Entropy Given the Number of States of Node C
0.0494
3.1160%
4.9388%
3.8212%
3.8413%
5.6950%
4.6772%
C
Condition
E
Erythrocyte Sedimentation Rate

Mutual Information



Symmetric Normalized Mutual Information

$$I_{SN}(C,E) = 2 \times \frac{I(C,E)}{log_2(S_C) + log_2(S_E)}$$

Relative Mutual Information

$$I_{RN}(C,E) = \frac{I(C,E)}{H(C)}$$

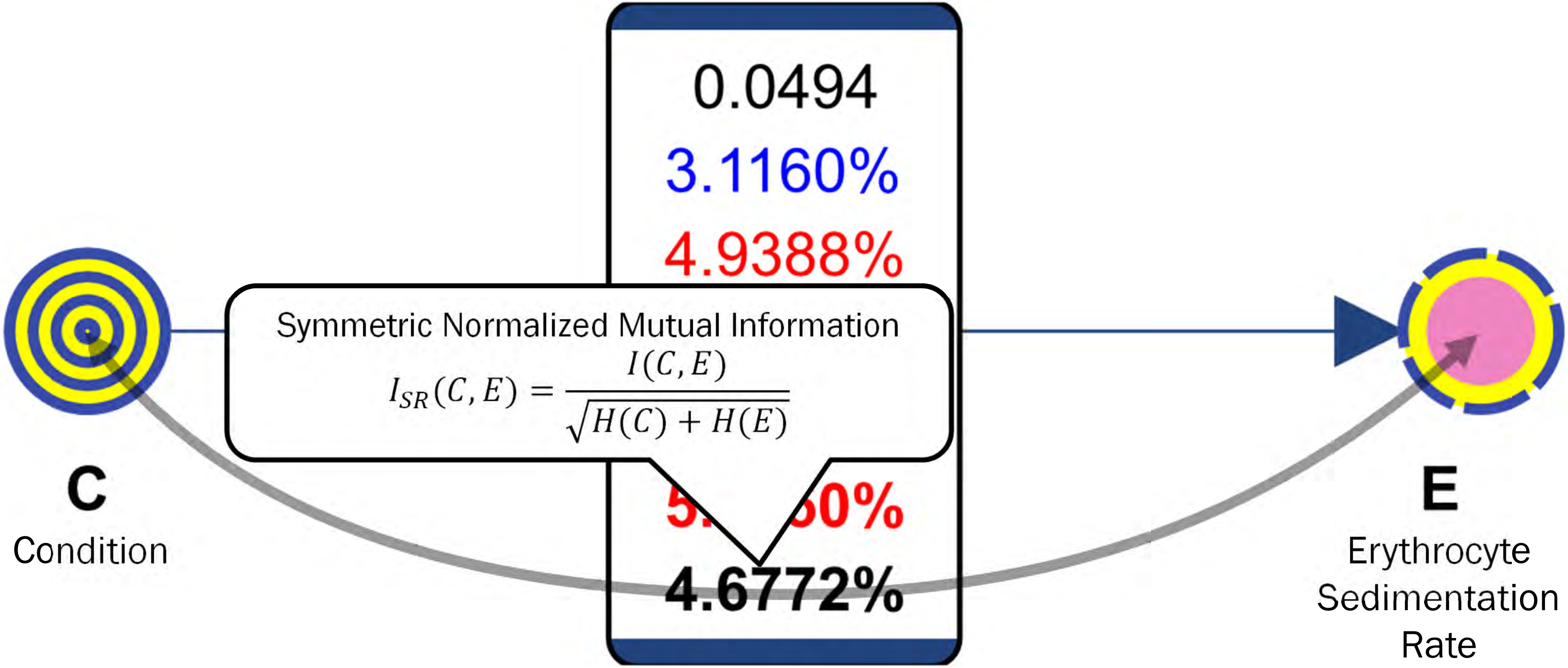
Denominator: Entropy of C

Relative Mutual Information

$$I_{RN}(E,C) = \frac{I(C,E)}{H(E)}$$

Denominator: Entropy of E

Mutual Information



Predictive Modeling

Importance of Information

- With Mutual Information, we have captured the average amount of information shared between two variables.
- However, the case-specific relevance very much depends on the state of the actual observation.
- The Bayes Factor can quantify how observations are consistent with a hypothesis or other pieces of evidence observed so far.

Target Analysis Report Condition (CAD3) [8]

Node significance with respect to the information gain brought by the node to the knowledge of the target state

Condition = Coronary Artery Disease (71.3794%)

Node	Binary Mutual Information	Relative Binary Mutual Information	Binary Relative Significance	Posterior Mean Value	Posterior Modal Value		*Prior Modal Value*		Bayes Factor				
Typical Chest Pain	0.2281	26.4069%	1.0000	0.7116	1	71.1649%	*1*	*54.1254%*	0.3949	[illegible]	17.0395%	0	17.0395%
Regional Wall Motion Abnormality	0.1119	12.9543%	0.4906	0.8472	0	62.0370%	*0*	*71.5864%*	-0.2066	2	3.7786%	0	9.5493%
Age	0.0908	10.5104%	0.3980	60.8882	<=58.5 (1/3)	40.7407%	*<=58.5 (1/3)*	*51.1217%*	-0.3275	>68.[illegible]	5.9718%	<=58.5 (1/3)	10.3809%
Ejection Fraction (%)	0.0763	8.8306%	0.3344	45.9874	<=52.5 (1/2)	75.0000%	*<=52.5 (1/2)*	*65.0486%*	0.2054	[illegible]	[illegible]	[illegible]	[illegible]
Hypertension	0.0592	6.8575%	0.2597	0.6806	1	68.0556%	*1*	*59.1048%*	0.2034	[illegible]	[illegible]	[illegible]	[illegible]
Diabetes Mellitus	0.0518	6.0017%	0.2273	0.3704	0	62.9630%	*0*	*70.2735%*	-0.1585	[illegible]	[illegible]	[illegible]	[illegible]
Erythrocyte Sedimentation mm/h	0.0491	5.6872%	0.2154	21.4426	<=20.5 (2/3)	56.0185%	*<=20.5 (2/3)*	*59.0661%*	-0.0764	[illegible]	[illegible]	[illegible]	[illegible]
T Inversion	0.0450	5.2065%	0.1972	0.3657	0	63.4259%	*0*	*70.2750%*	-0.1479	1	6.8490%	0	6.8490%
Blood Pressure (mm/Hg)	0.0421	4.8737%	0.1846	131.8138	>116.5 (2/2)	84.2515%	*>116.5 (2/2)*	*77.6655%*	0.1174	>116.5 (2/2)	6.5859%	<=116.5 (1/2)	6.5859%
Fasting Blood Sugar	0.0418	4.8390%	0.1832	124.1053	>98.5 (2/2)	57.4074%	*<=98.5 (1/2)*	*50.1407%*	0.2034	>98.5 (2/2)	7.5481%	<=98.5 (1/2)	7.5481%
High Density Lipoprotein (mg/dl)	0.0374	4.3342%	0.1641	39.9114	>39.5 (3/3)	46.2963%	*>39.5 (3/3)*	*49.4946%*	-0.0964	<=39.5 (2/3)	2.7826%	>39.5 (3/3)	3.1983%
Valvular Heart Disease	0.0358	4.1464%	0.1570	0.7778	Mild	53.2407%	*Mild*	*49.1880%*	0.1142	Mild	4.0528%	None	3.0887%
Triglyceride (mh/dl)	0.0356	4.1177%	0.1559	159.2280	<=137.5 (1/2)	50.9259%	*<=137.5 (1/2)*	*57.7338%*	-0.1810	>137.5 (2/2)	6.8079%	<=137.5 (1/2)	6.8079%
Potassium (mEq/lit)	0.0348	4.0250%	0.1524	4.2710	<=4.65 (1/2)	79.1667%	*<=4.65 (1/2)*	*83.8134%*	-0.0823	>4.65 (2/2)	4.6467%	<=4.65 (1/2)	4.6467%
Creatine (mg/dl)	0.0305	3.5353%	0.1339	1.0659	<=1.25 (2/3)	66.3745%	*<=1.25 (2/3)*	*71.9472%*	-0.1163	>1.25 (3/3)	3.6361%	<=1.25 (2/3)	5.5727%
Body Mass Index (kg/m2)	0.0297	3.4403%	0.1303	27.0373	<=30.15 (1/3)	79.6296%	*<=30.15 (1/3)*	*77.5644%*	0.0379	<=30.15 (1/3)	2.0652%	>30.954 (3/3)	3.5228%
Q Wave	0.0266	3.0753%	0.1165	0.0741	0	92.5926%	*0*	*94.7126%*	-0.0327	1	2.1200%	0	2.1200%
Pulse Rate (ppm)	0.0227	2.6316%	0.0997	75.9062	>71 (2/2)	52.7778%	*<=71 (1/2)*	*52.7873%*	0.1608	>71 (2/2)	5.5651%	<=71 (1/2)	5.5651%

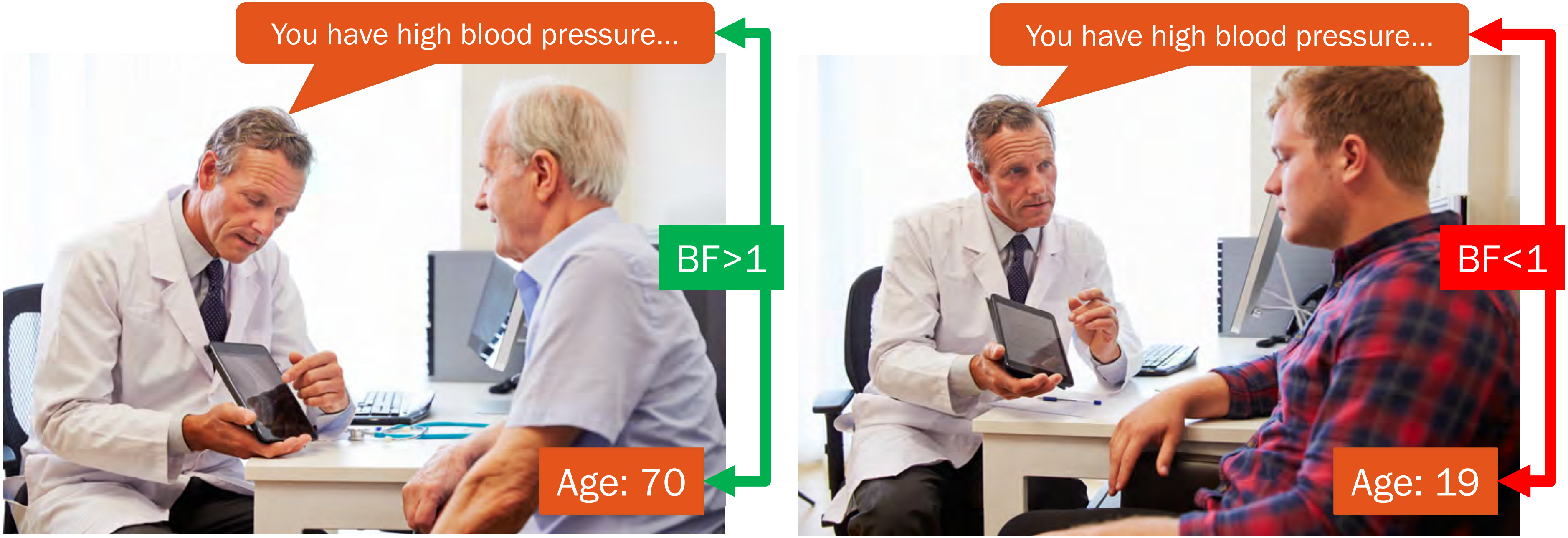
Close | Save As... | Print | Mapping | Quadrants

Bayes Factor>1:
Supporting Evidence

Bayes Factor<1:
Contradicting Evidence

Predictive Modeling

Bayes Factor: Measuring the Agreement of Pieces of Evidence



Predictive Modeling

- New **Target Analysis Report** to be revealed as part of the BayesiaLab 9 launch on October 10 at the 7th Annual BayesiaLab Conference.

Predictive Modeling

- So far, all our measure everything were mostly about “importance” with respect to one variable.
- However, predictive models are not limited to merely predicting a single target variable.
- Through BayesiaLab’s **Unsupervised Learning**, we can learn models that simultaneously predict all variables in a domain.
- The learning process itself is out of scope for today. However, we want to understand how to evaluate the importance of variables and relationships in such a network.

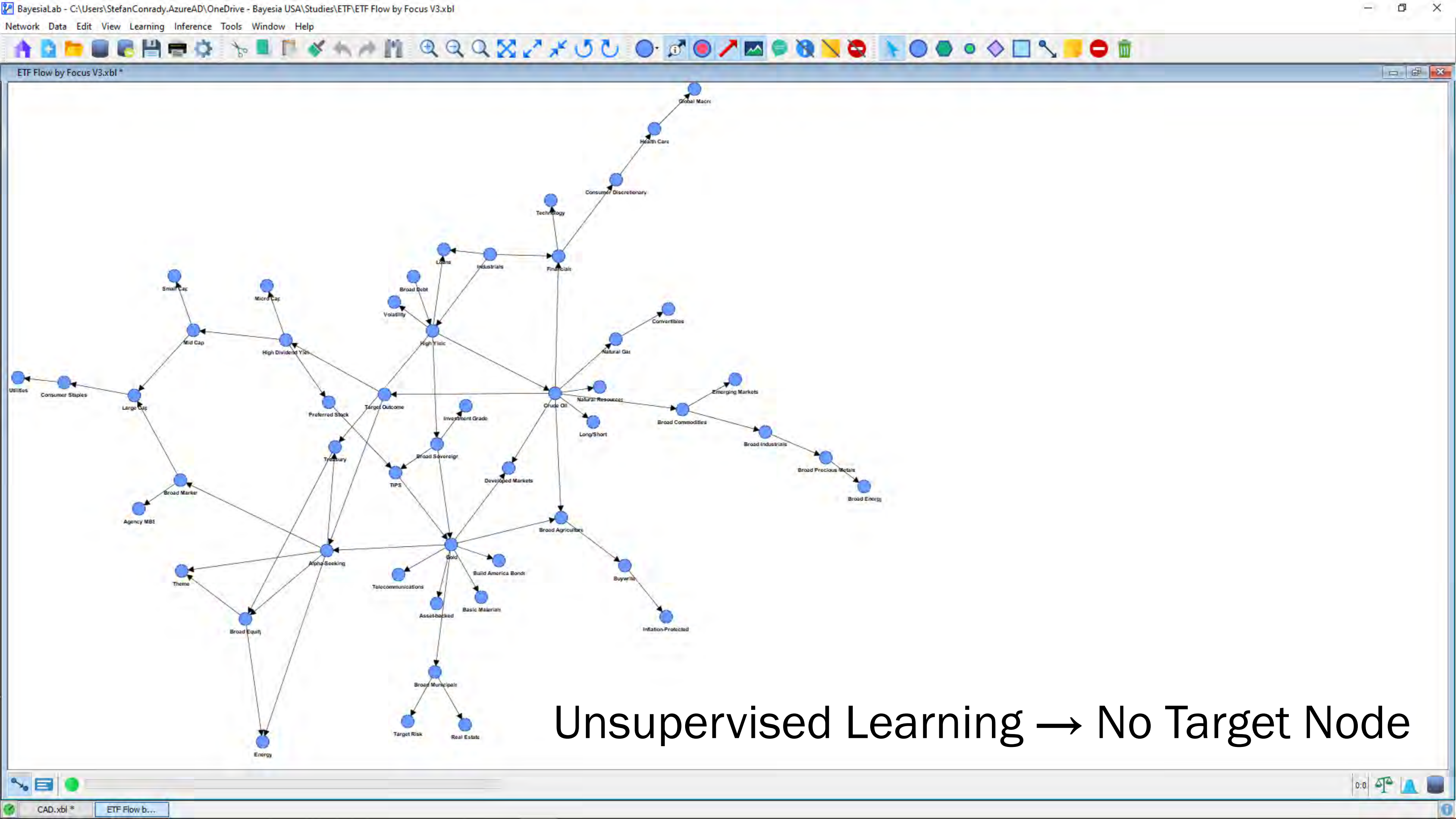
Predictive Models

Unsupervised Learning

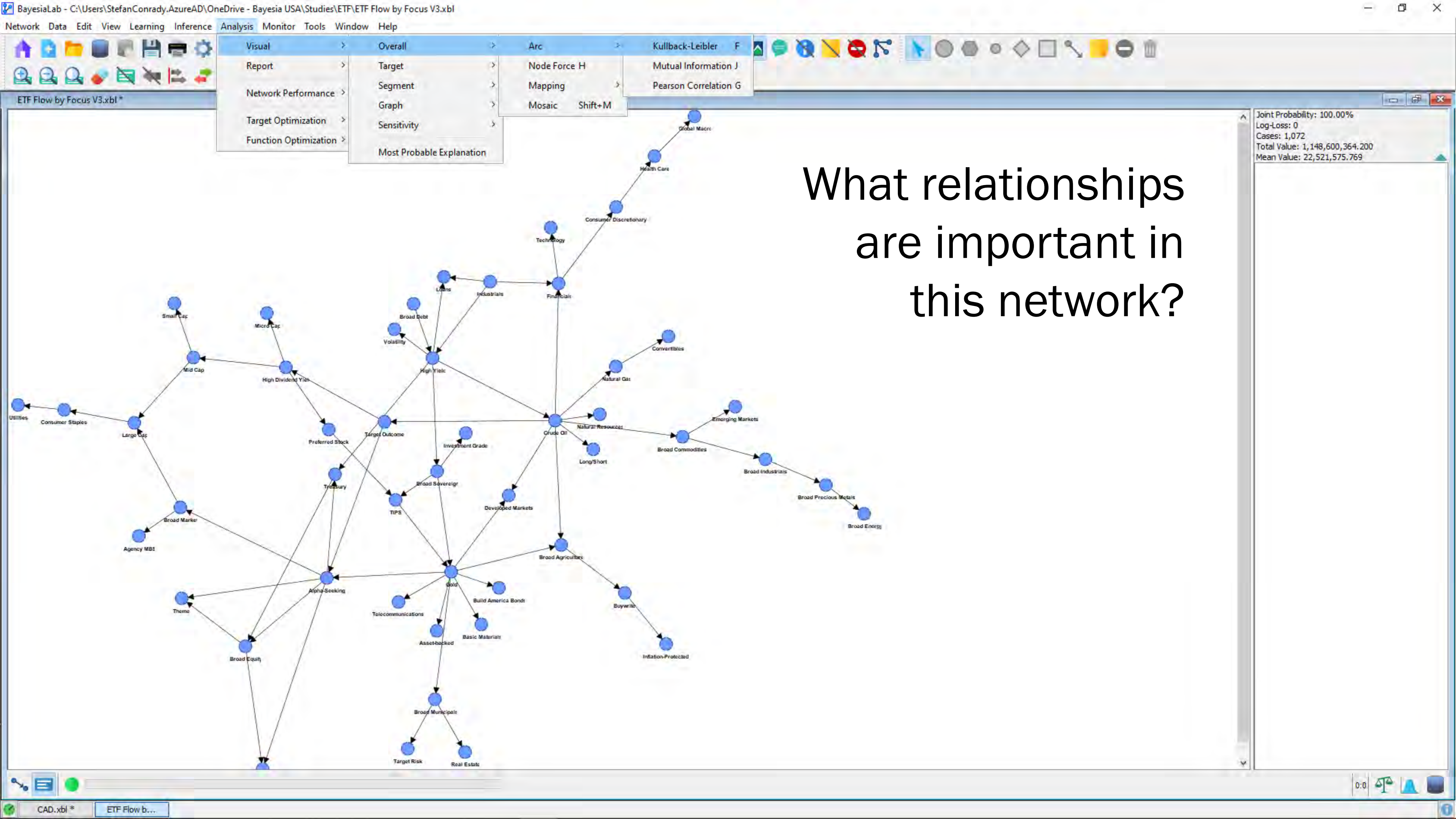
- 1,147 Exchange-Traded Funds
- Timeframe: 2014–2018
- Daily Flow grouped by 50 investment themes
- 1,000 daily observations

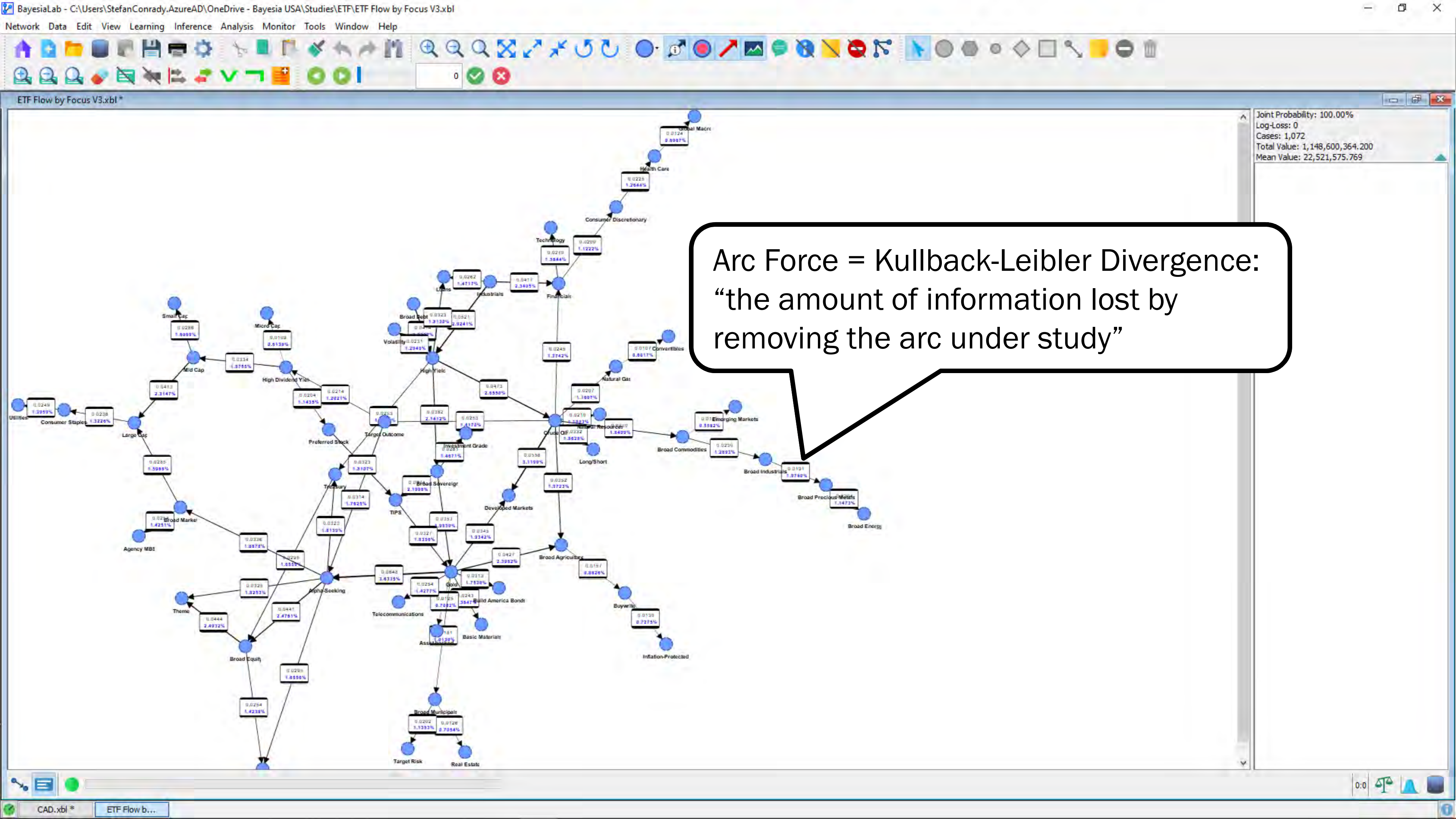
- Alpha-Seeking
- Basic Materials
- Broad Equity
- Consumer Discretionary
- Energy
- Financials
- High Dividend Yield
- Industrials
- Mid Cap
- Natural Resources
- Preferred Stock
- Technology
- Agency MBS
- Asset-backed
- Broad Agriculture
- Broad Commodities
- Broad Debt
- Broad Energy
- Broad Industrials
- Broad Market
- Broad Municipals
- Broad Sovereign
- Build America Bonds
- Buywrite
- Consumer Staples
- Crude Oil
- Developed Markets
- Emerging Markets
- Global Macro
- Gold
- Health Care
- High Yield
- Inflation-Protected
- Investment Grade
- Large Cap
- Loans
- Long/Short
- Micro Cap
- Natural Gas
- Real Estate
- Small Cap
- TIPS
- Target Outcome
- Target Risk
- Telecommunications
- Theme
- Treasury
- Utilities
- Volatility
- Broad Precious Metals

See Webinar on Analyzing Capital Flows of Exchange-Traded Funds:
https://www.bayesia.com/2018-04-13-analyzing-capital-flows

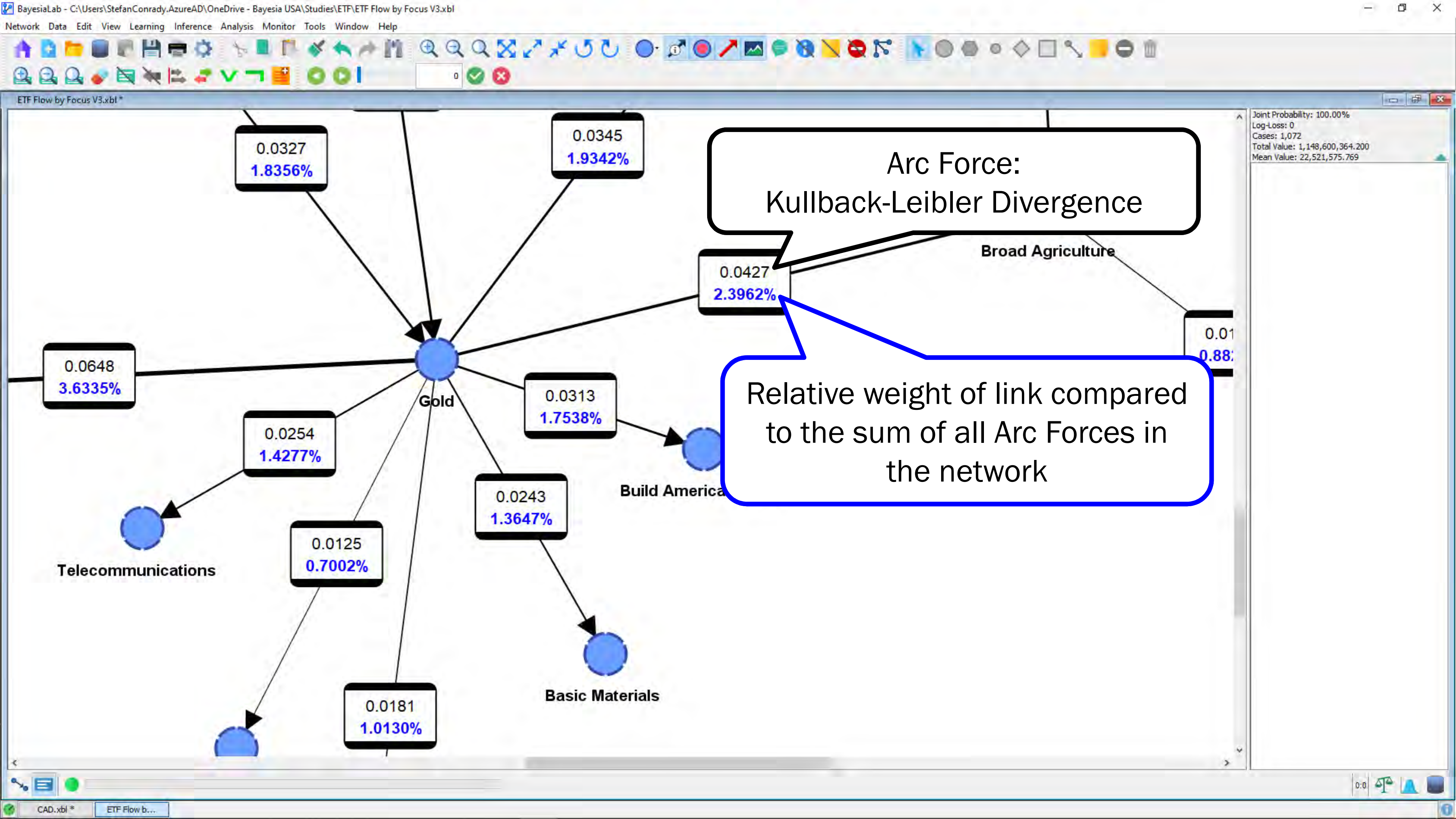

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Studies\ETF\ETF Flow by Focus V3.xbl
Network Data Edit View Learning Inference Tools Window Help
ETF Flow by Focus V3.xbl *
Health Care
Consumer Discretionary
Industrials
Broad Debt
Volatility
Mid Cap
Utilities
Consumer Staples
Preferred Stock
Target Outcome
Convertibles
Emerging Markets
Investment Grade
Long/Short
Broad Commodities
Broad Industrials
Broad Precious Metals
TIPS
Developed Markets
Broad Market
Agency MBS
Alpha-Seeking
Theme
Telecommunications
Build America Bonds
Basic Materials
Inflation-Protected
Target Risk
Energy
CAD.xbl *
ETF Flow b...
Unsupervised Learning → No Target Node

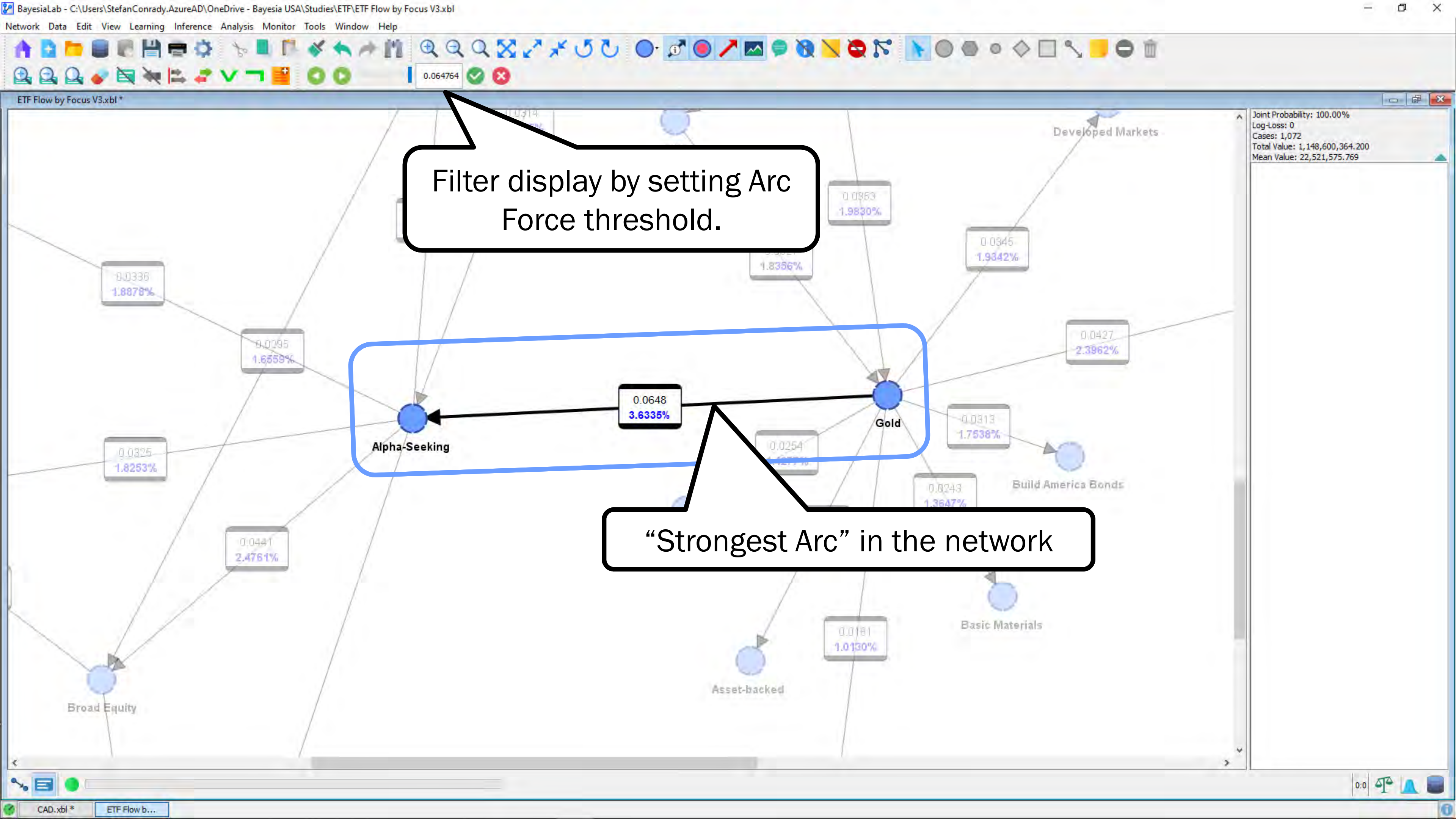
What relationships are important in this network?

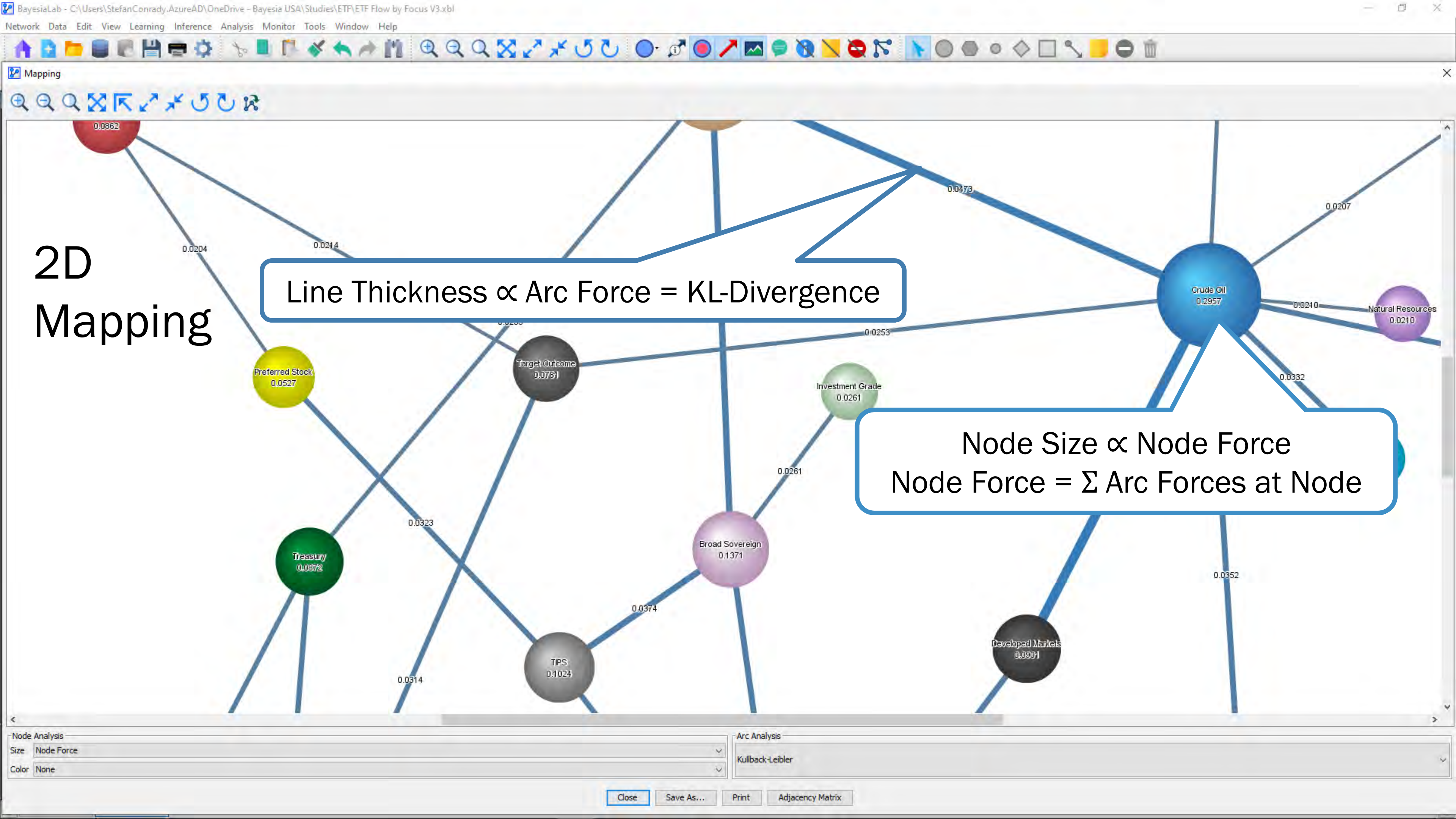




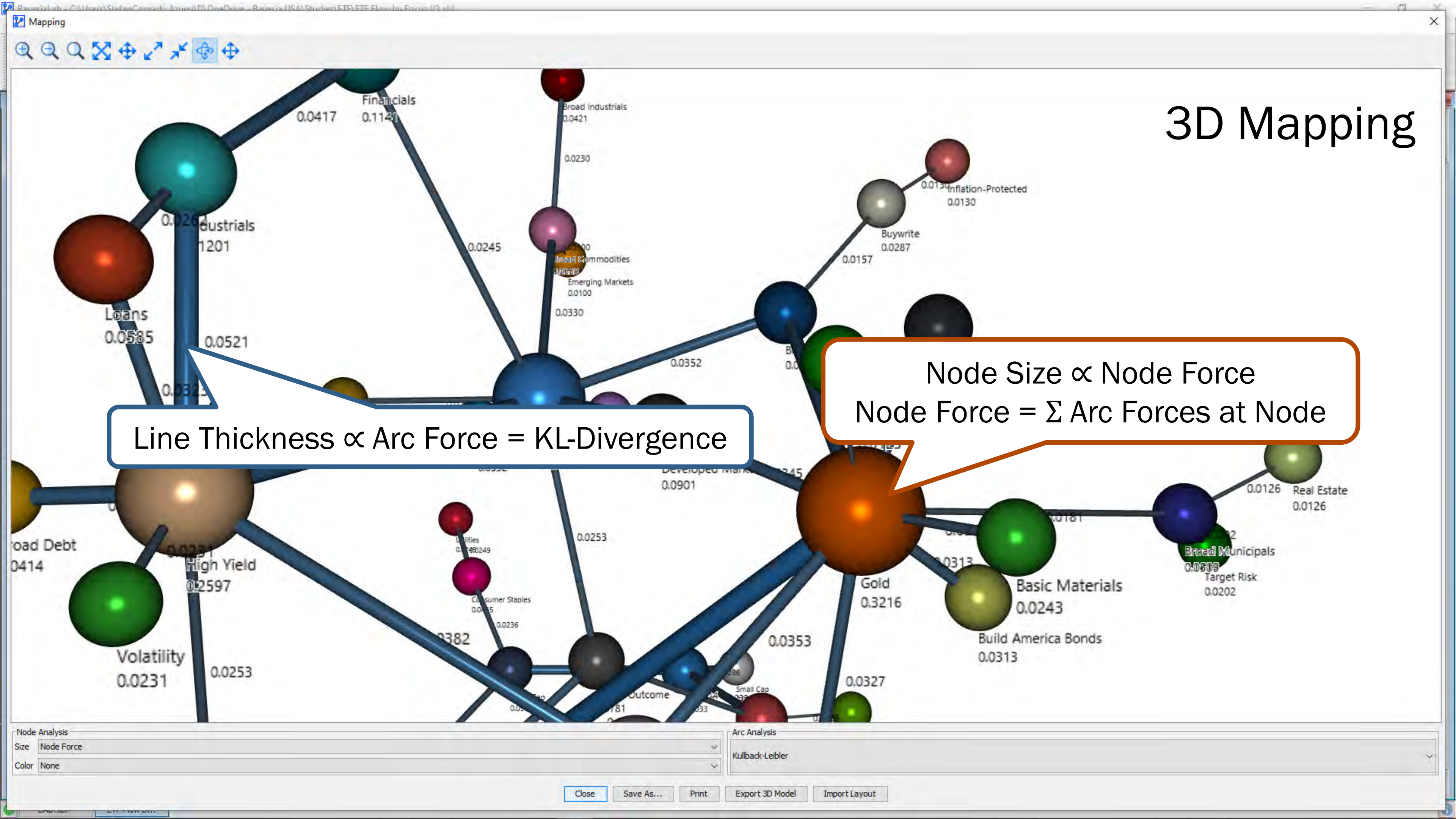
Arc Force = Kullback-Leibler Divergence: “the amount of information lost by removing the arc under study”


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Studies\ETF\ETF Flow by Focus V3.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
ETF Flow by Focus V3.xbl *
Joint Probability: 100.00%
Log-Loss: 0
Cases: 1,072
Total Value: 1,148,600,364.200
Mean Value: 22,521,575.769
Arc Force:
Kullback-Leibler Divergence
Relative weight of link compared to the sum of all Arc Forces in the network
0.0327
1.8356%
0.0345
1.9342%
0.0427
2.3962%
0.0648
3.6335%
0.0313
1.7538%
0.0254
1.4277%
0.0243
1.3647%
0.0125
0.7002%
0.0181
1.0130%
Gold
Broad Agriculture
Build America
Telecommunications
Basic Materials
CAD.xbl *
ETF Flow b...


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Studies\ETF\ETF Flow by Focus V3.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
0.064764
ETF Flow by Focus V3.xbl *
Filter display by setting Arc Force threshold.
"Strongest Arc" in the network
0.0648
3.6335%
Alpha-Seeking
Gold
Developed Markets
Build America Bonds
Basic Materials
Asset-backed
Broad Equity
Joint Probability: 100.00%
Log-Loss: 0
Cases: 1,072
Total Value: 1,148,600,364.200
Mean Value: 22,521,575.769
CAD.xbl *
ETF Flow b...



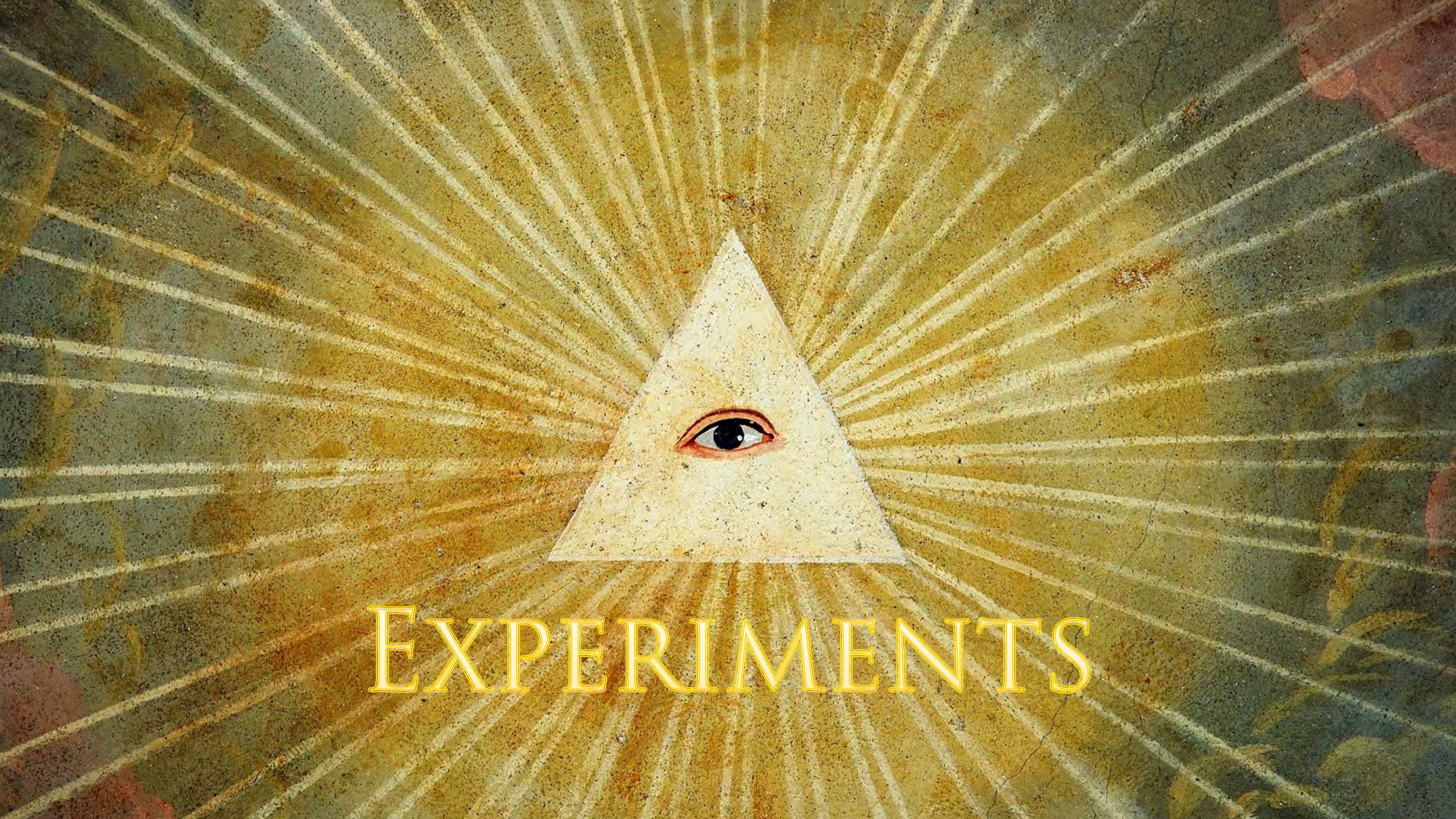
BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Studies\ETF\ETF Flow by Focus V3.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Mapping
2D Mapping
Line Thickness ∝ Arc Force = KL-Divergence
Node Size ∝ Node Force
Node Force = Σ Arc Forces at Node
0.0862
0.0204
0.0214
0.0473
0.0207
Crude Oil
0.2957
0.0210
Natural Resources
0.0210
0.0253
Preferred Stock
0.0527
Target Outcome
0.0781
Investment Grade
0.0261
0.0332
0.0261
0.0323
Treasury
0.0872
Broad Sovereign
0.1371
0.0352
0.0374
Developed Markets
0.0901
TIPS
0.1024
0.0314
Node Analysis
Size Node Force
Color None
Arc Analysis
Kullback-Leibler
Close
Save As...
Print
Adjacency Matrix


3D Mapping
Line Thickness ∝ Arc Force = KL-Divergence
Node Size ∝ Node Force
Node Force = Σ Arc Forces at Node
Mapping
Financials
0.0417
Broad Industrials
0.0421
0.0230
Inflation-Protected
0.0130
Buywrite
0.0287
0.0157
0.0245
Emerging Markets
0.0100
0.0330
Loans
0.0585
0.0521
0.0352
0.0901
Real Estate
0.0126
0.0126
Broad Municipals
Target Risk
0.0202
0.0253
High Yield
0.2597
Gold
0.3216
Basic Materials
0.0243
Build America Bonds
0.0313
Volatility
0.0231
0.0253
0.0236
0.0353
0.0327
Small Cap
Outcome
Node Analysis
Size
Node Force
Color
None
Arc Analysis
Kullback-Leibler
Close
Save As...
Print
Export 3D Model
Import Layout

Prediction

Causation


Data
Theory
Deterministic
Probabilistic
Causation
given that I do



Experiments

Experiments

$$\text{Spring Stiffness} = \frac{\text{Modulus of Spring Steel} \times \text{Wire Diameter}^4}{8 \times \text{Number of Active Coils} \times \text{Mean Coil Diameter}^3}$$


0%
25
50
75
CONVERSION
RATE
A
CONVERSION
RATE
B
START / STOP
TEST

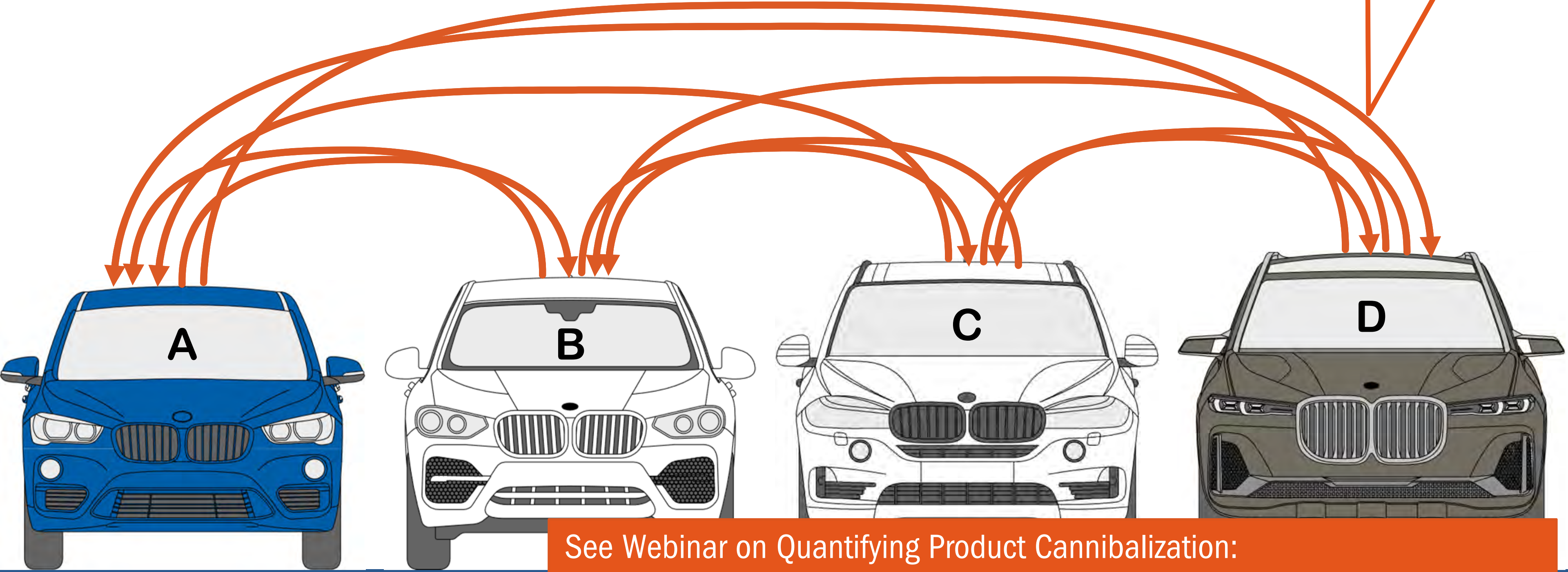

NO EXPERIMENTS


CAUSAL
INFERENCE
LABORATORY

Causation

Theoretical assumption: potential substitution between all models in SUV portfolio.

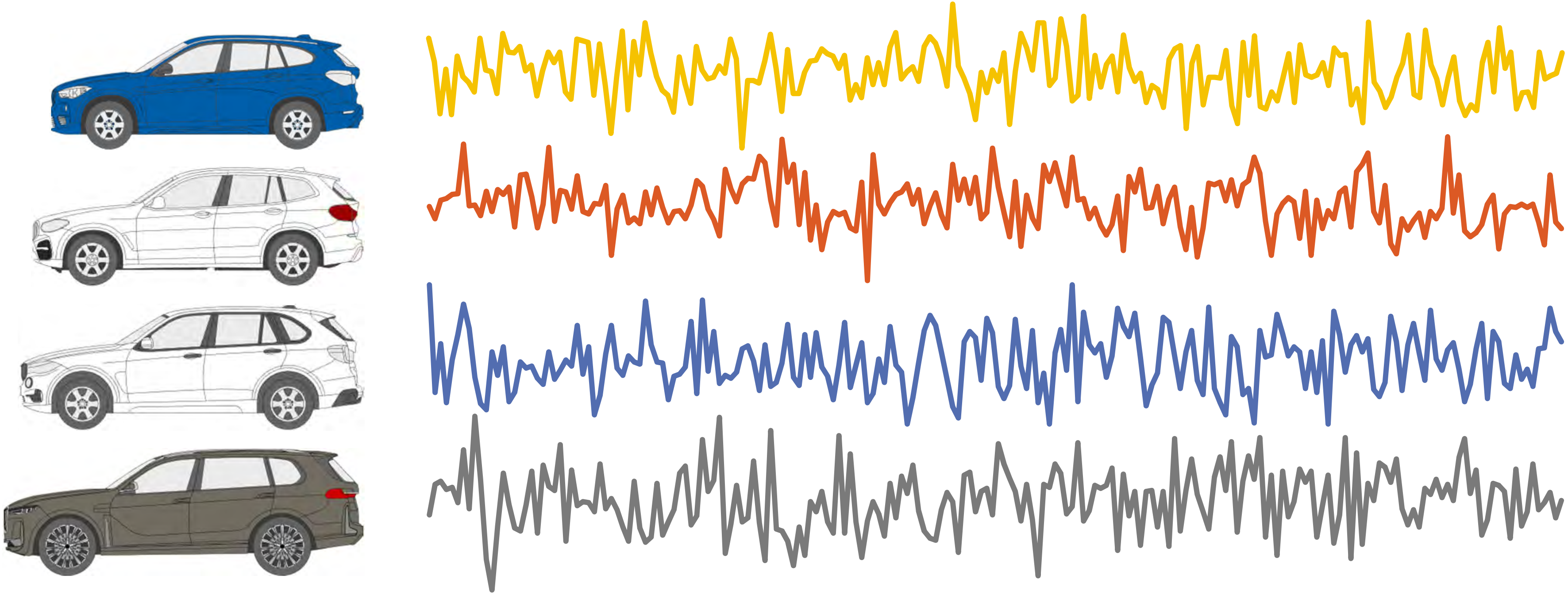
Example: Cannibalization/Substitution Between Vehicles in Portfolio

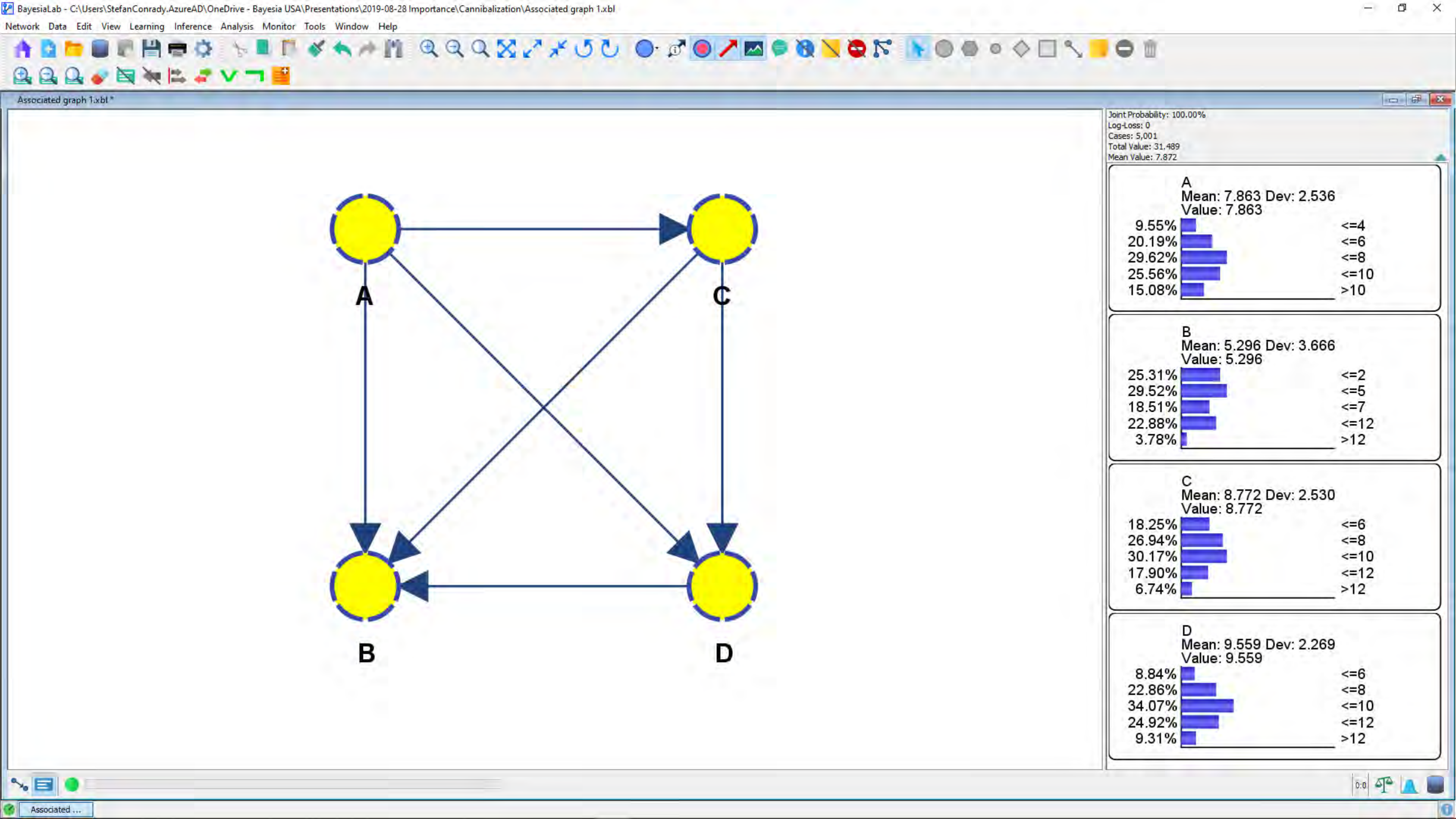


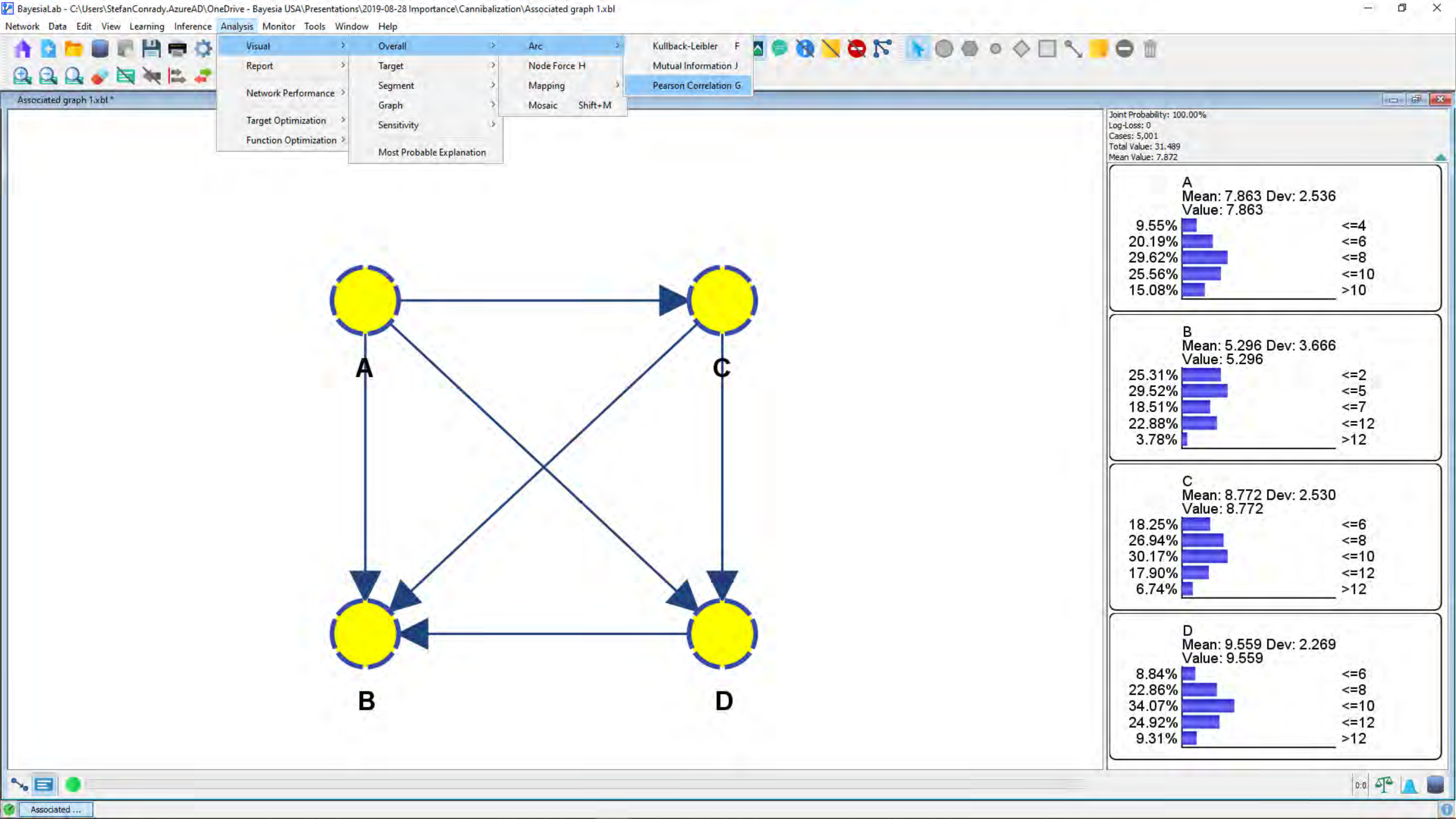
See Webinar on Quantifying Product Cannibalization:
https://www.bayesia.com/2018-03-23-quantifying-product-cannibalization

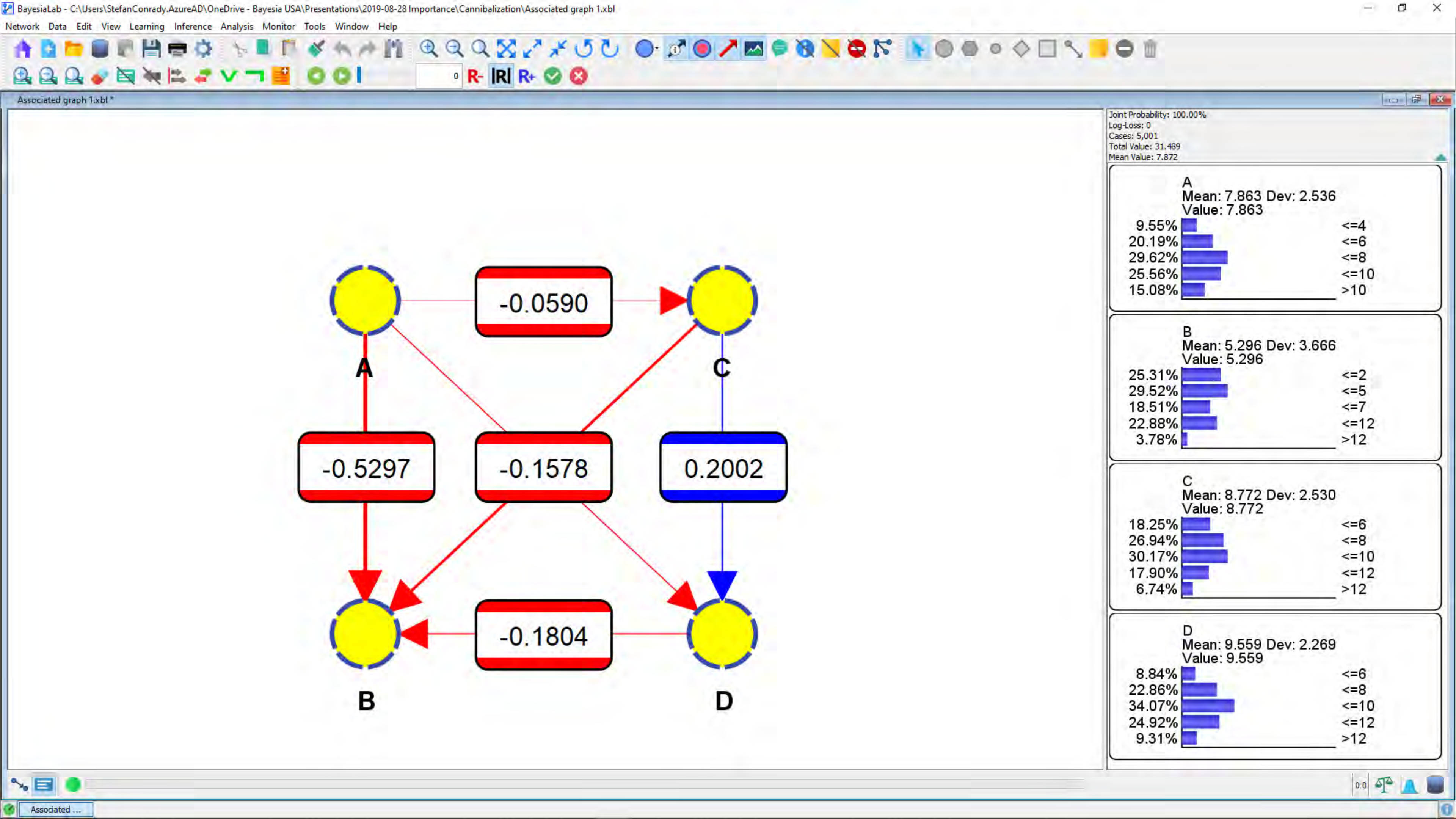
Causation

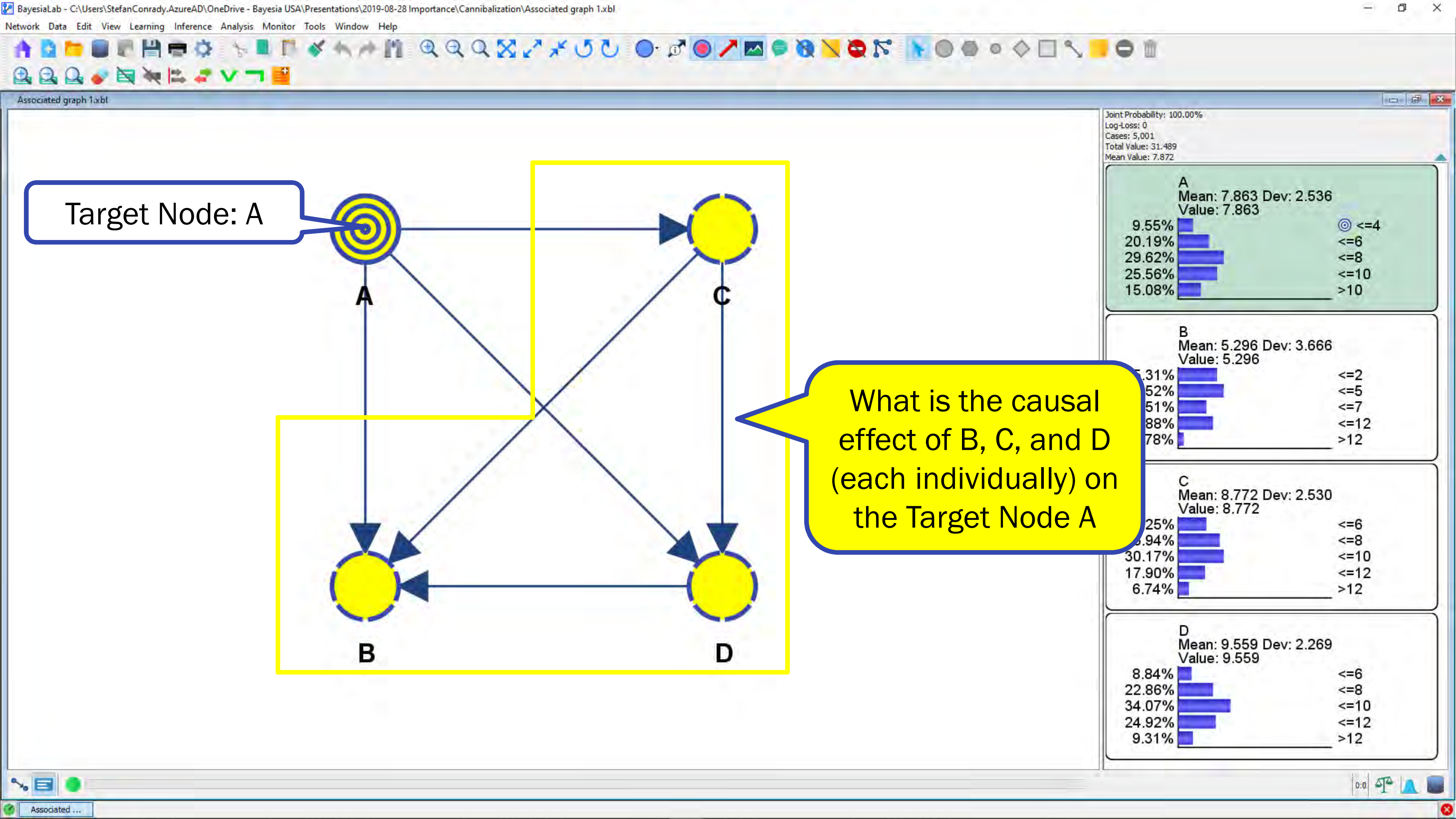
Data for Estimation of Bayesian Network Model




BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl *
A
C
B
D
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
25.31% <=2
29.52% <=5
18.51% <=7
22.88% <=12
3.78% >12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
18.25% <=6
26.94% <=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Overall
Target
Segment
Graph
Sensitivity
Most Probable Explanation
Arc
Node Force H
Mapping
Mosaic Shift+M
Kullback-Leibler F
Mutual Information J
Pearson Correlation G
Associated graph 1.xbl *
A
B
C
D
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
25.31% <=2
29.52% <=5
18.51% <=7
22.88% <=12
3.78% >12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
18.25% <=6
26.94% <=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...

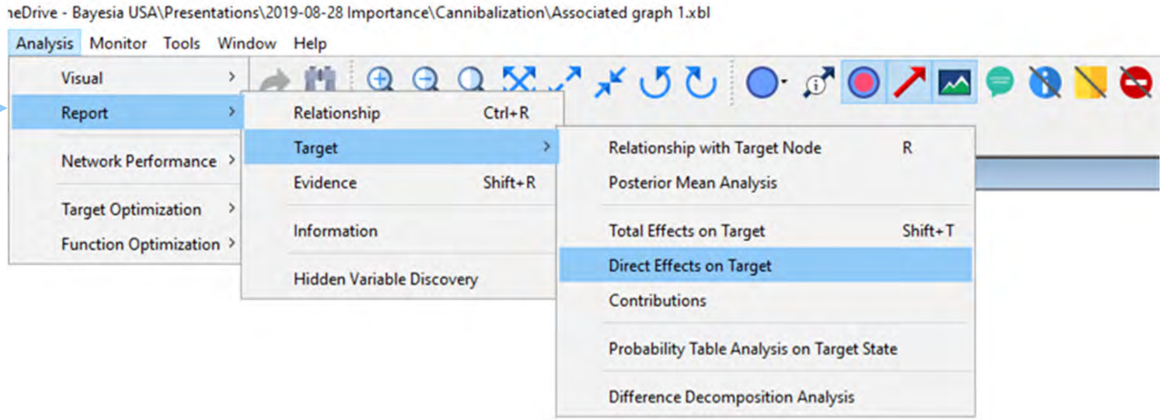

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl *
-0.0590
-0.5297
-0.1578
0.2002
-0.1804
A
B
C
D
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
25.31% <=2
29.52% <=5
18.51% <=7
22.88% <=12
3.78% >12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
18.25% <=6
26.94% <=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...

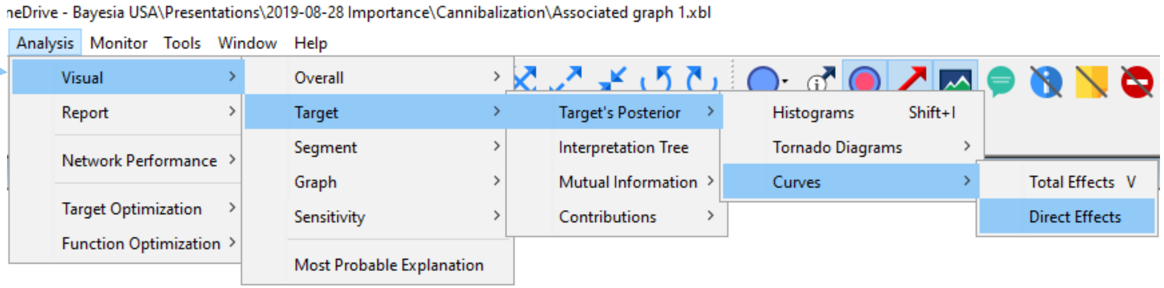

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl
Target Node: A
A
C
B
D
What is the causal effect of B, C, and D (each individually) on the Target Node A
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
<=2
<=5
<=7
<=12
>12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
<=6
<=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...

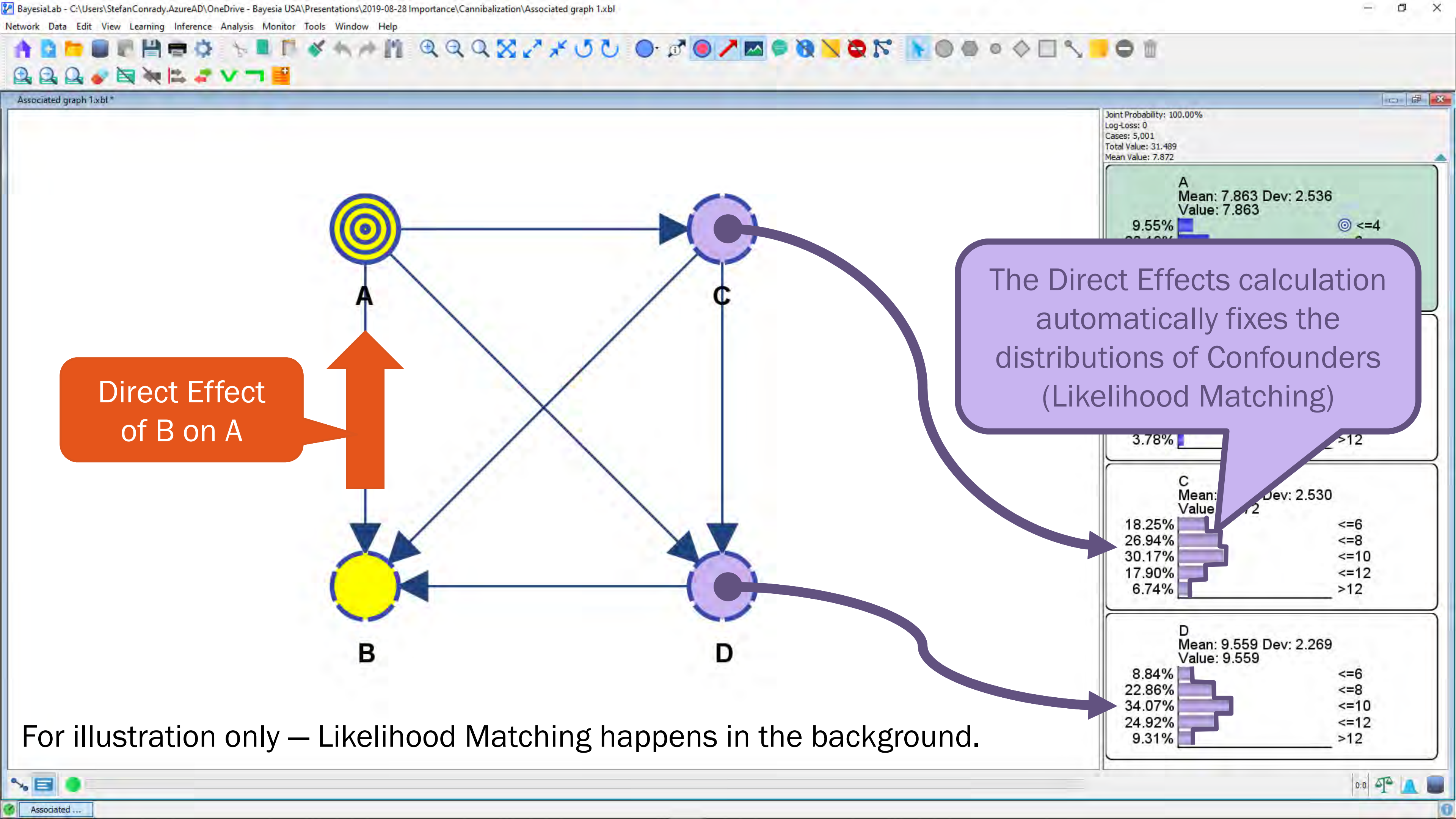
Causation

Direct Effects Analysis

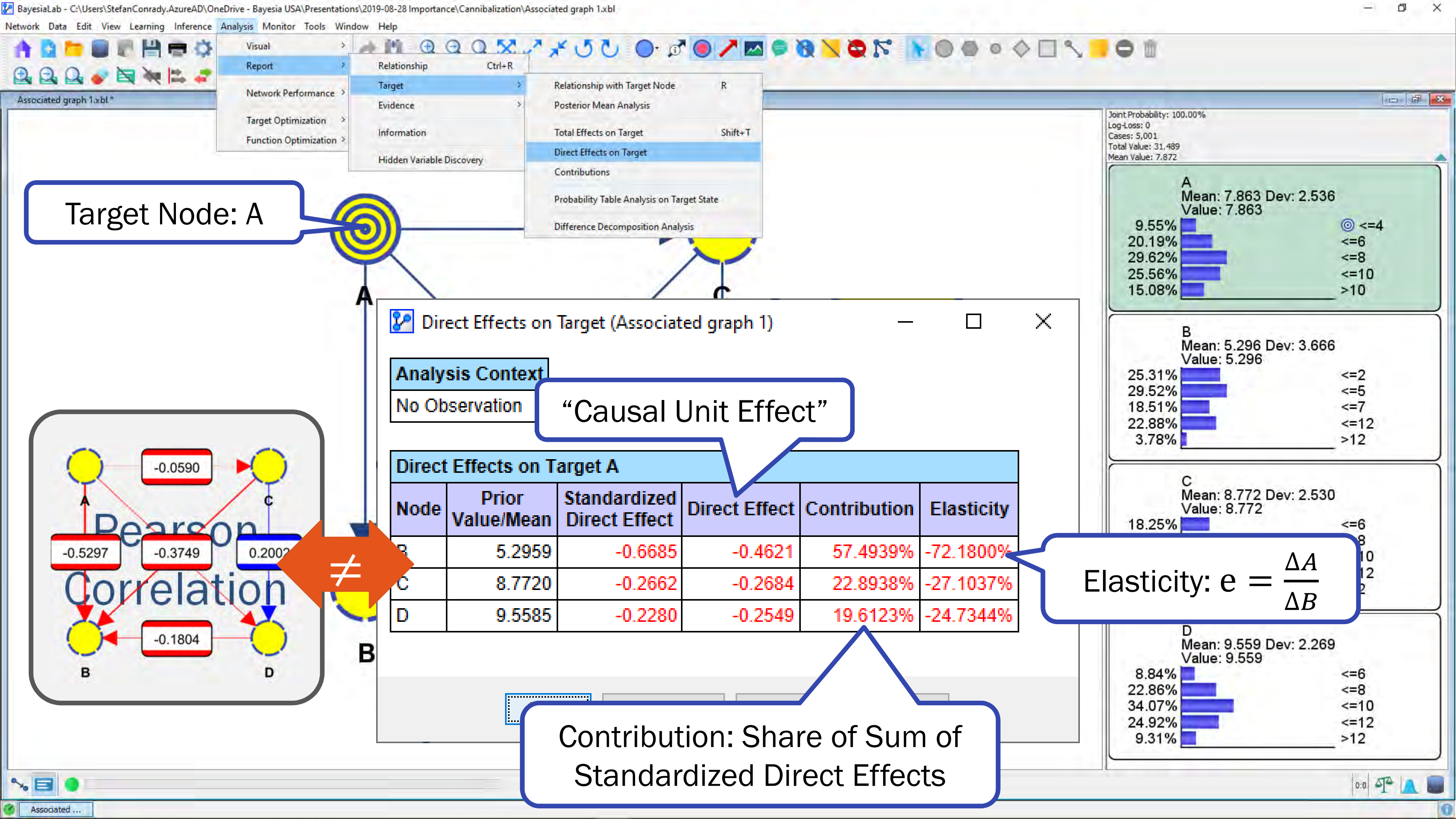
- Report
- Visual





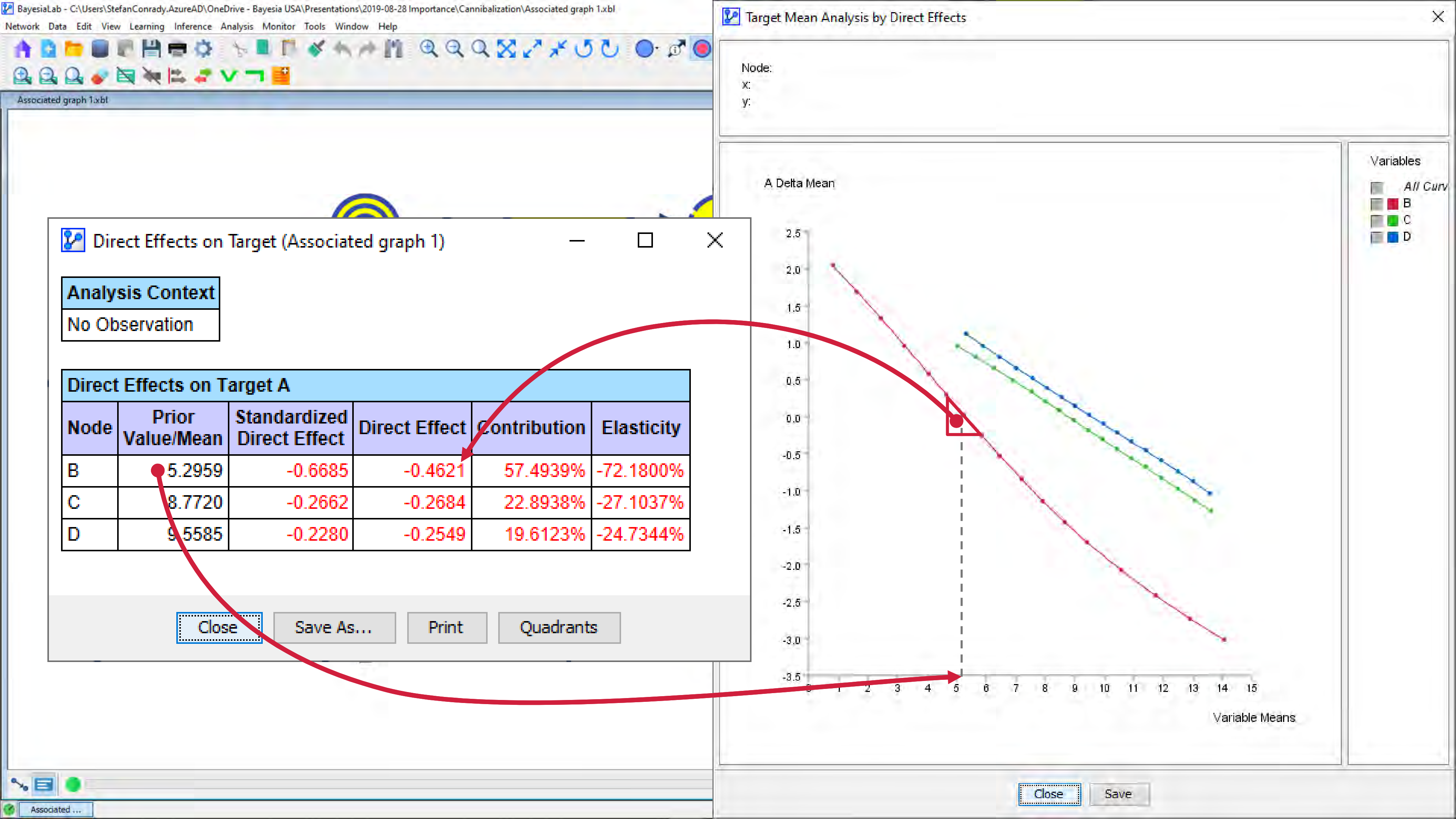


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl *
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55%
<=4
3.78%
>12
C
Mean:
Value
Dev: 2.530
18.25%
26.94%
30.17%
17.90%
6.74%
<=6
<=8
<=10
<=12
>12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84%
22.86%
34.07%
24.92%
9.31%
<=6
<=8
<=10
<=12
>12
A
C
B
D
Direct Effect of B on A
The Direct Effects calculation automatically fixes the distributions of Confounders (Likelihood Matching)
For illustration only — Likelihood Matching happens in the background.
Associated ...



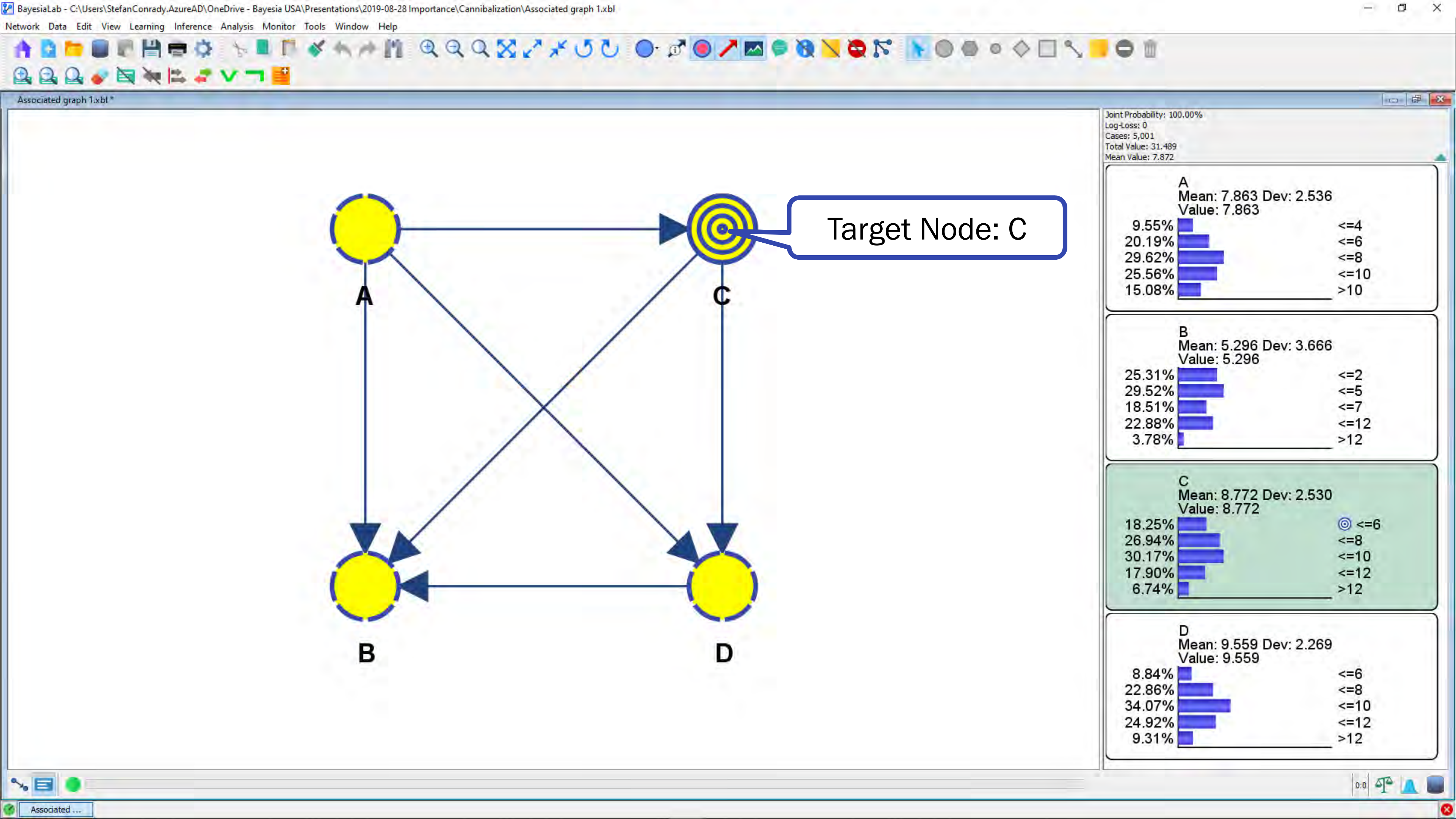
Direct Effects on Target A

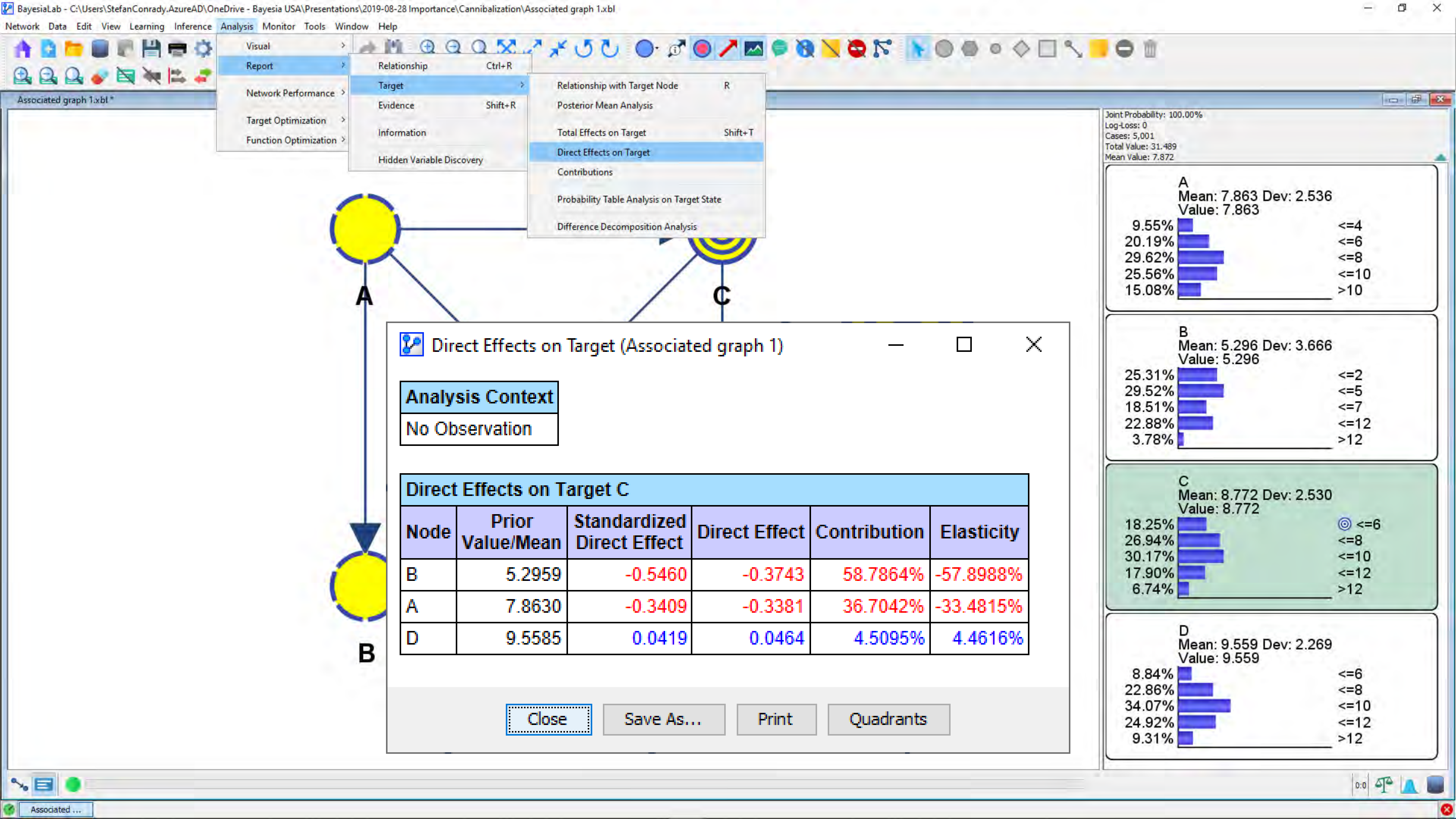
Node	Prior Value/Mean	Standardized Direct Effect	Direct Effect	Contribution	Elasticity
B	5.2959	-0.6685	-0.4621	57.4939%	-72.1800%
C	8.7720	-0.2662	-0.2684	22.8938%	-27.1037%
D	9.5585	-0.2280	-0.2549	19.6123%	-24.7344%

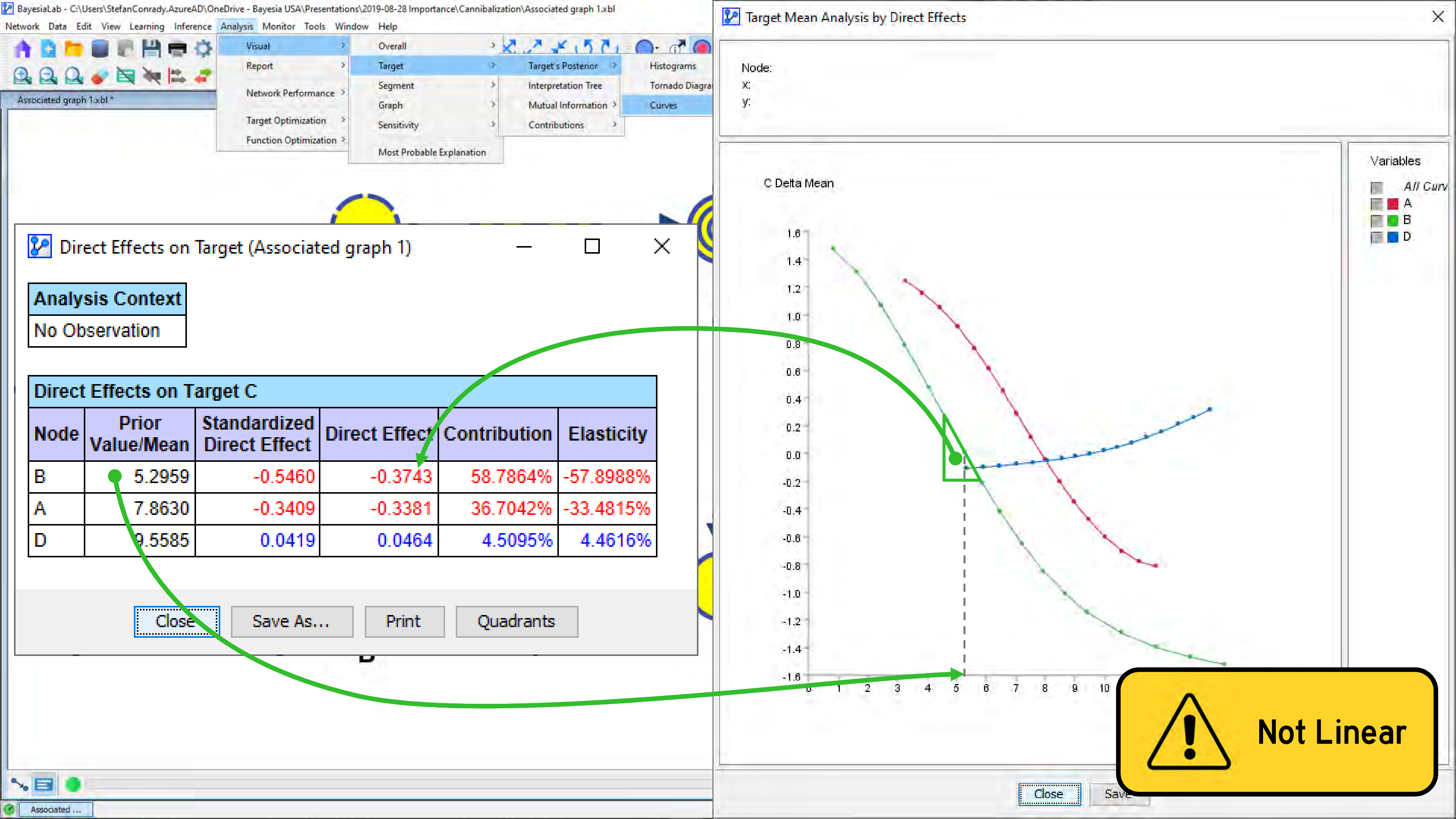

BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl
Direct Effects on Target (Associated graph 1)
Analysis Context
No Observation
Direct Effects on Target A

Node	Prior Value/Mean	Standardized Direct Effect	Direct Effect	Contribution	Elasticity
B	5.2959	-0.6685	-0.4621	57.4939%	-72.1800%
C	8.7720	-0.2662	-0.2684	22.8938%	-27.1037%
D	9.5585	-0.2280	-0.2549	19.6123%	-24.7344%

Close
Save As...
Print
Quadrants
Target Mean Analysis by Direct Effects
Node:
x:
y:
A Delta Mean
Variables
All Curv
B
C
D
Variable Means
Close
Save
Associated ...


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl *
Target Node: C
A
B
C
D
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
25.31% <=2
29.52% <=5
18.51% <=7
22.88% <=12
3.78% >12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
18.25% <=6
26.94% <=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-08-28 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Visual
Report
Network Performance
Target Optimization
Function Optimization
Relationship Ctrl+R
Target
Evidence Shift+R
Information
Hidden Variable Discovery
Relationship with Target Node R
Posterior Mean Analysis
Total Effects on Target Shift+T
Direct Effects on Target
Contributions
Probability Table Analysis on Target State
Difference Decomposition Analysis
Associated graph 1.xbl *
A
B
C
Direct Effects on Target (Associated graph 1)
Analysis Context
No Observation
Direct Effects on Target C
Node | Prior Value/Mean | Standardized Direct Effect | Direct Effect | Contribution | Elasticity
B | 5.2959 | -0.5460 | -0.3743 | 58.7864% | -57.8988%
A | 7.8630 | -0.3409 | -0.3381 | 36.7042% | -33.4815%
D | 9.5585 | 0.0419 | 0.0464 | 4.5095% | 4.4616%
Close
Save As...
Print
Quadrants
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
25.31% <=2
29.52% <=5
18.51% <=7
22.88% <=12
3.78% >12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
18.25% <=6
26.94% <=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...



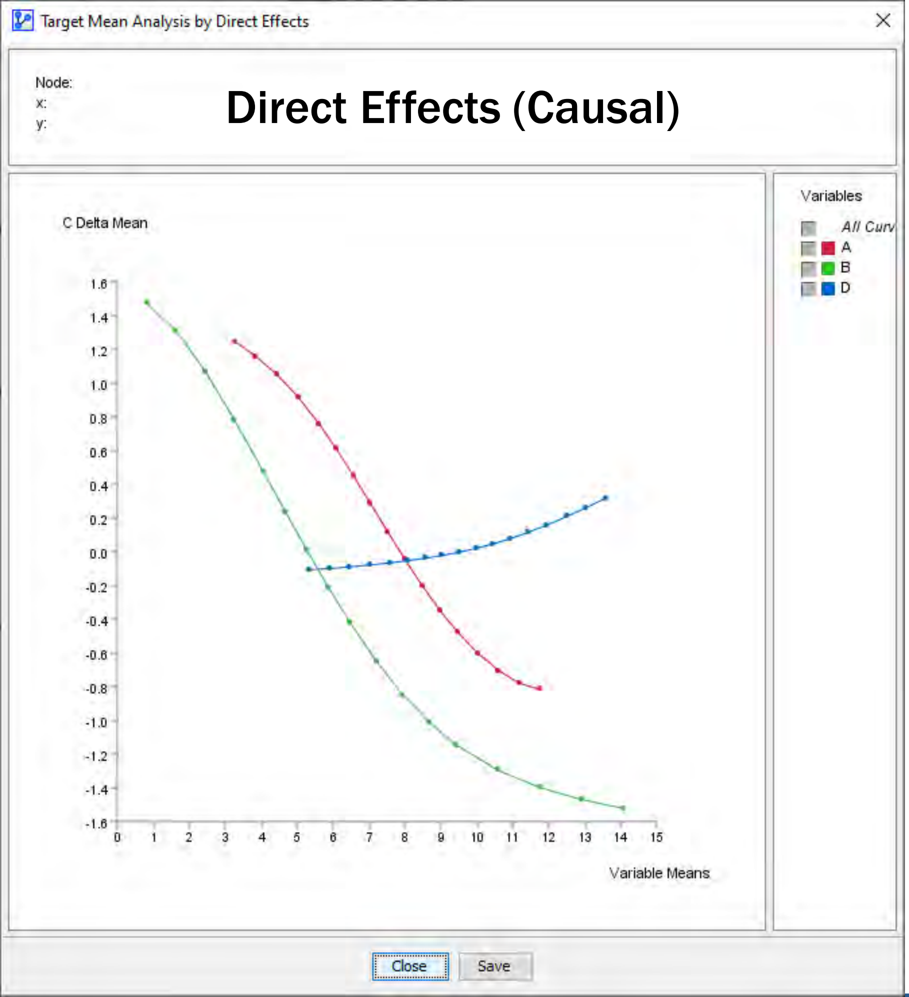
Direct Effects on Target C					
Node	Prior Value/Mean	Standardized Direct Effect	Direct Effect	Contribution	Elasticity
B	5.2959	-0.5460	-0.3743	58.7864%	-57.8988%
A	7.8630	-0.3409	-0.3381	36.7042%	-33.4815%
D	9.5585	0.0419	0.0464	4.5095%	4.4616%

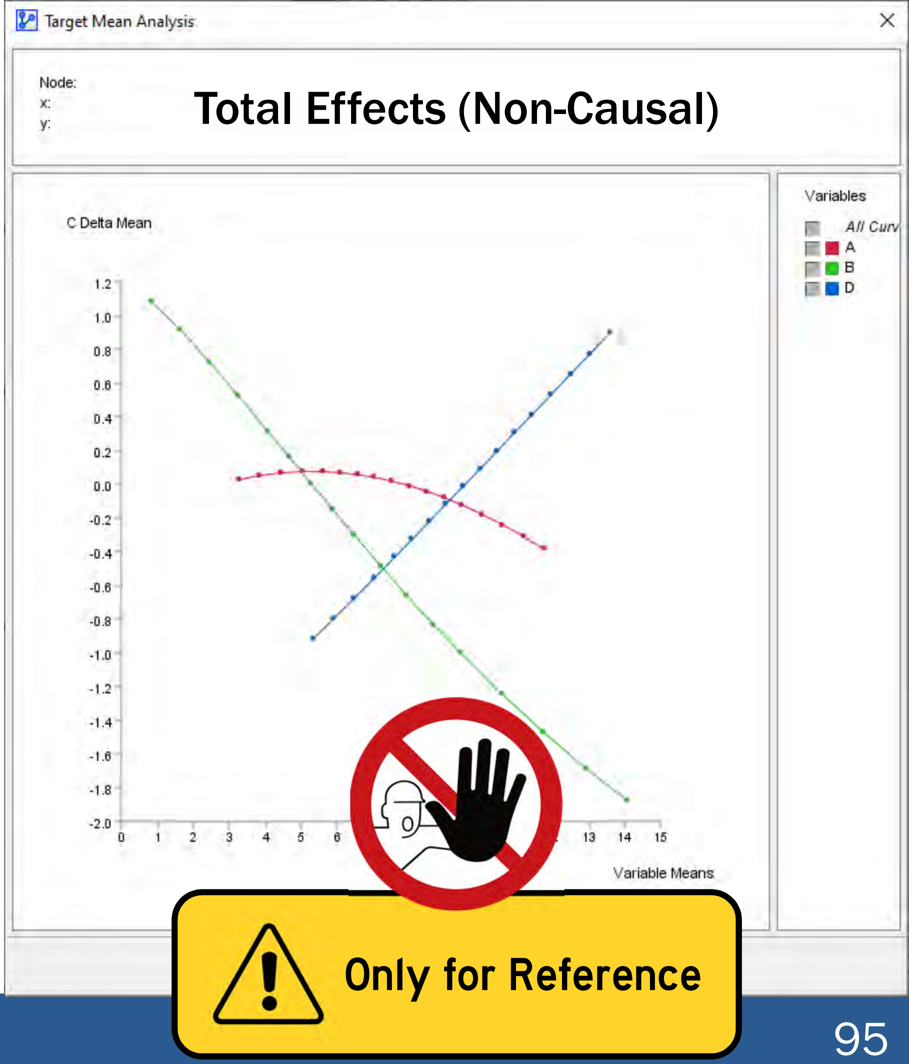
Total Effects

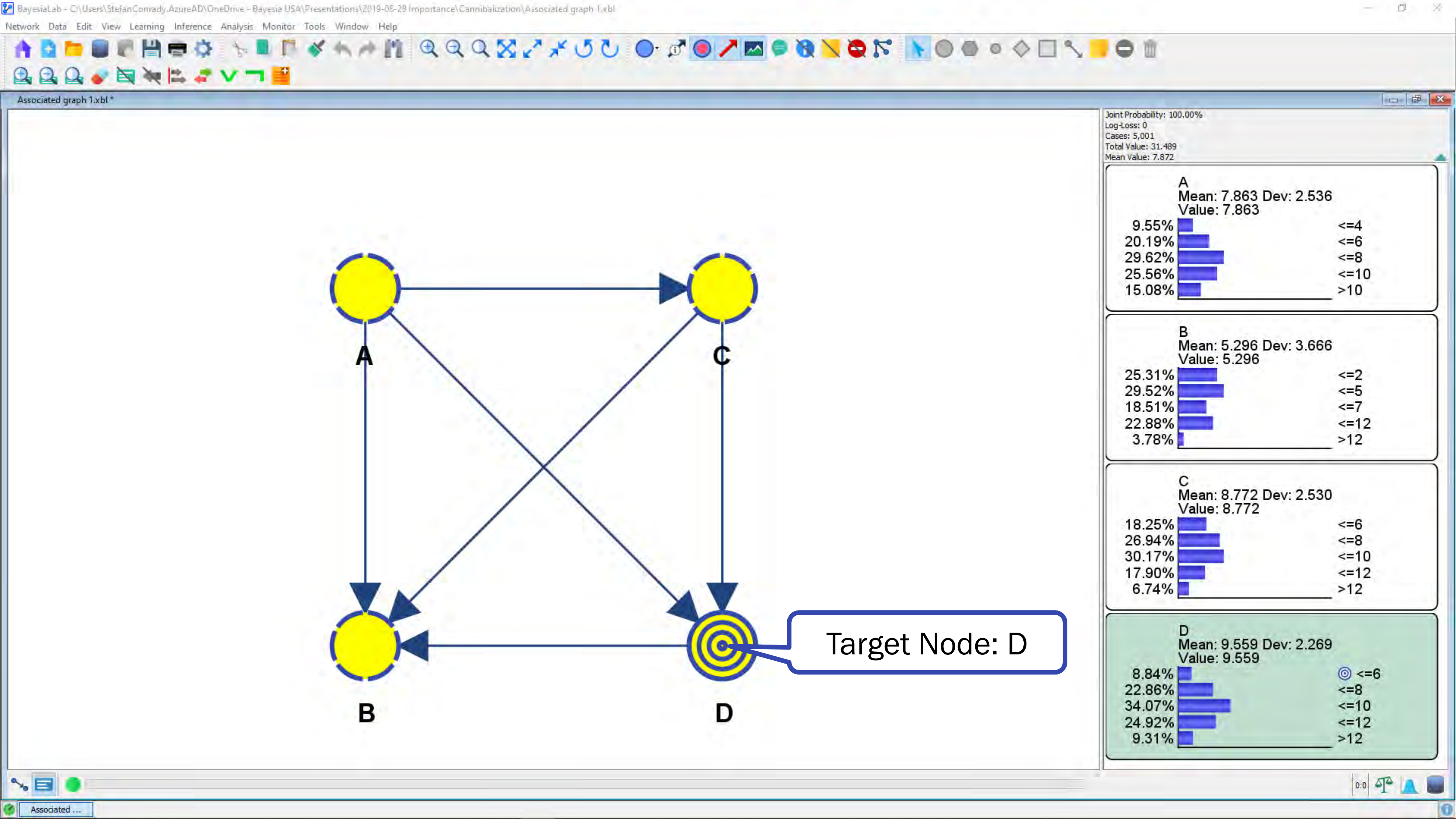
Why *"Direct"* Effects?

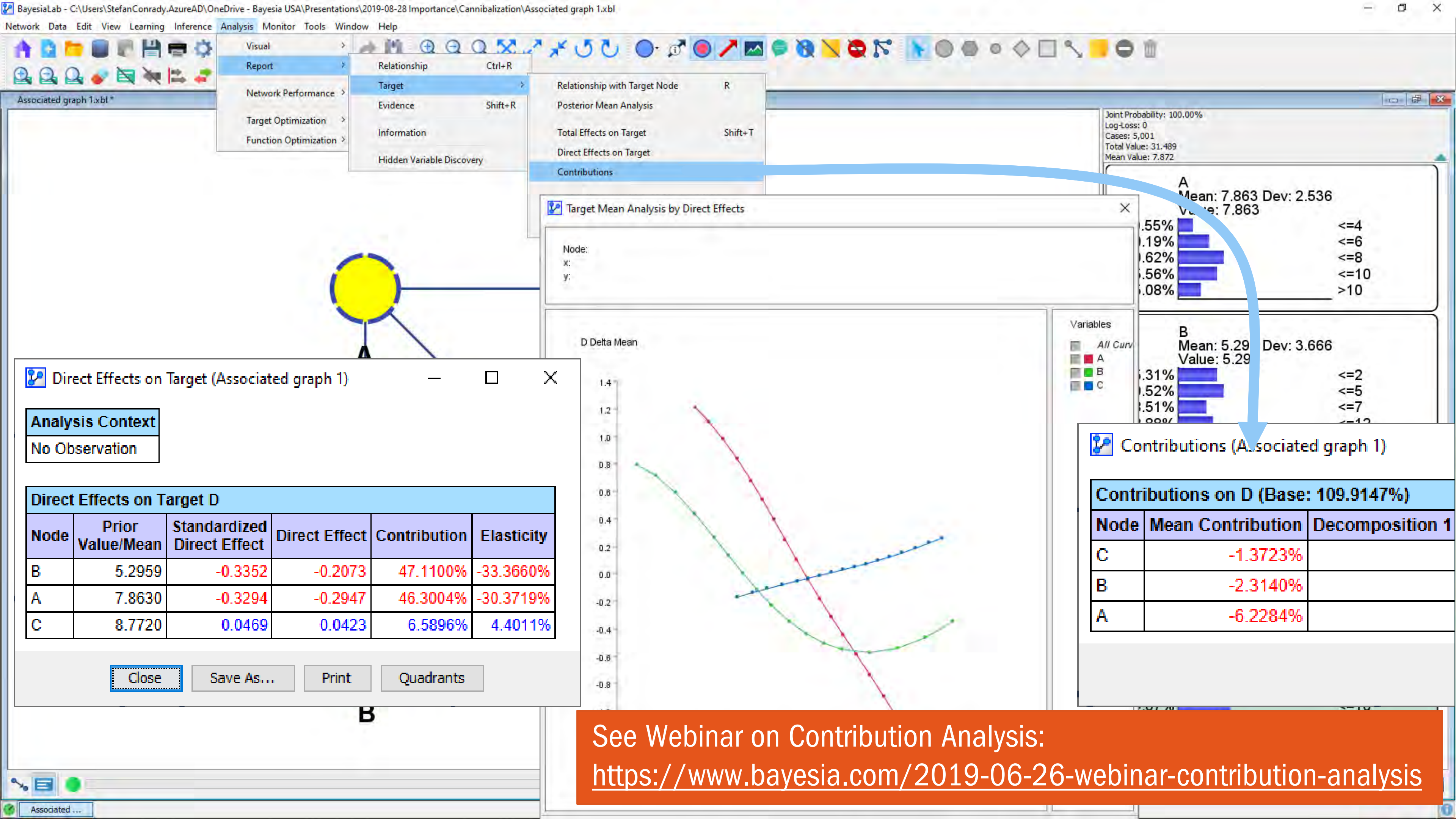
- Wouldn't Total Effects be a reasonable proxy for Direct Effects in most cases?

Total vs. Direct Effects


Target Mean Analysis by Direct Effects
Node:
x:
y:
Direct Effects (Causal)
C Delta Mean
Variables
All Curv
A
B
D
Variable Means
Close
Save


Target Mean Analysis
Node:
x:
y:
Total Effects (Non-Causal)
C Delta Mean
Variables
All Curv
A
B
D
Variable Means
Only for Reference


BayesiaLab - C:\Users\StefanConrady.AzureAD\OneDrive - Bayesia USA\Presentations\2019-06-29 Importance\Cannibalization\Associated graph 1.xbl
Network Data Edit View Learning Inference Analysis Monitor Tools Window Help
Associated graph 1.xbl *
A
C
B
D
Target Node: D
Joint Probability: 100.00%
Log-Loss: 0
Cases: 5,001
Total Value: 31.489
Mean Value: 7.872
A
Mean: 7.863 Dev: 2.536
Value: 7.863
9.55% <=4
20.19% <=6
29.62% <=8
25.56% <=10
15.08% >10
B
Mean: 5.296 Dev: 3.666
Value: 5.296
25.31% <=2
29.52% <=5
18.51% <=7
22.88% <=12
3.78% >12
C
Mean: 8.772 Dev: 2.530
Value: 8.772
18.25% <=6
26.94% <=8
30.17% <=10
17.90% <=12
6.74% >12
D
Mean: 9.559 Dev: 2.269
Value: 9.559
8.84% <=6
22.86% <=8
34.07% <=10
24.92% <=12
9.31% >12
Associated ...



Analysis Context
No Observation

Direct Effects on Target D					
Node	Prior Value/Mean	Standardized Direct Effect	Direct Effect	Contribution	Elasticity
B	5.2959	-0.3352	-0.2073	47.1100%	-33.3660%
A	7.8630	-0.3294	-0.2947	46.3004%	-30.3719%
C	8.7720	0.0469	0.0423	6.5896%	4.4011%

Contributions on D (Base: 109.9147%)		
Node	Mean Contribution	Decomposition 1
C	-1.3723%	
B	-2.3140%	
A	-6.2284%	

See Webinar on Contribution Analysis:
https://www.bayesia.com/2019-06-26-webinar-contribution-analysis

Summary

Importance in Predictive Modeling

- Total Effects
- Information Theory
  - Entropy & Mutual Information
  - Arc Force, Node Force
- Bayes Factor
- Tornado Chart

Importance in Causal Modeling

- Direct Effects
- Contributions & Synergy
- Elasticity

What part of "it's important" don't you understand?

Thank You!

stefan.conrady@bayesia.us

BayesianNetwork

linkedin.com/in/stefanconrady

facebook.com/bayesia