# **EDS-Bayes-Net**

Developing a probabilistic graphical model for the effect of the COL5A1 mutation on Chronic Fatigue and Brain Fog via the Ehlers-Danlos Syndrome pathway, using the R bnlearn package.

Casey developed an Adverse Outcome Pathway for the effect of the COL5A1 genetic mutation as it leads to Chronic Fatigue and Brain Fog via the Ehlers-Danlos pathway, using the R bnlearn package. This is described below.

# The EDS BN

# A Bayesian Network for EDS AOP

# \*\*\* Libraries

library(bnlearn)

library(graph)

library(Rgraphviz)

library(RBGL)

library(gRain)

library(ggplot2)

library(RColorBrewer)

Below creates the Model String which describes the Bayesian Network nodes and parent-child relationships among them

modelStr <- paste(

"[COL5A1\_Mutation]",

"[TypeV\_Collagen\_Function|COL5A1\_Mutation]",

"[Fibrils\_Strength|TypeV\_Collagen\_Function]",

"[Skin\_Strength|Fibrils\_Strength]",

"[Tissue\_Strength|Fibrils\_Strength]",

"[AAD|Tissue\_Strength]", #AAD = Arthritis-Adrenaline Disorder

"[Tendon\_Strength|Fibrils\_Strength]",

"[Immunity|Tendon\_Strength]",

"[Wound\_Healing|Skin\_Strength:Tendon\_Strength]",

"[Chronic\_Pain|Wound\_Healing:Immunity]",

"[Vein\_Plasticity|Fibrils\_Strength]",

"[Blood\_Flow\_to\_Brain|Vein\_Plasticity]",

"[GI\_Efficiency|Vein\_Plasticity]",

"[POTS|Vein\_Plasticity]",

"[Brain\_Fog|Blood\_Flow\_to\_Brain:POTS]",

"[Pharynx\_Structure|Vein\_Plasticity]",

"[Sleep\_Apnea|Pharynx\_Structure]",

"[Chronic\_Fatigue|AAD:Sleep\_Apnea]",

sep=""

)

Below creates the directed acyclic graph (DAG) from the string

dag = model2network(modelStr)

dag # Details

# Plot the DAG

graphviz.plot(dag)

Below highlights the observed nodes. Teal highlighted nodes are observed (Mark’s data is available, such is the case for his genomic data and symptoms) and white nodes are unobserved.

graphviz.plot(dag,

highlight = list(nodes = c("COL5A1\_Mutation","AAD",

"Immunity", "Wound\_Healing","Chronic\_Pain",

"GI\_Efficiency", "POTS", "Brain\_Fog",

"Sleep\_Apnea", "Chronic\_Fatigue"),

col = "black", fill = "#d6ffeb"))

Here is the resulting plot:

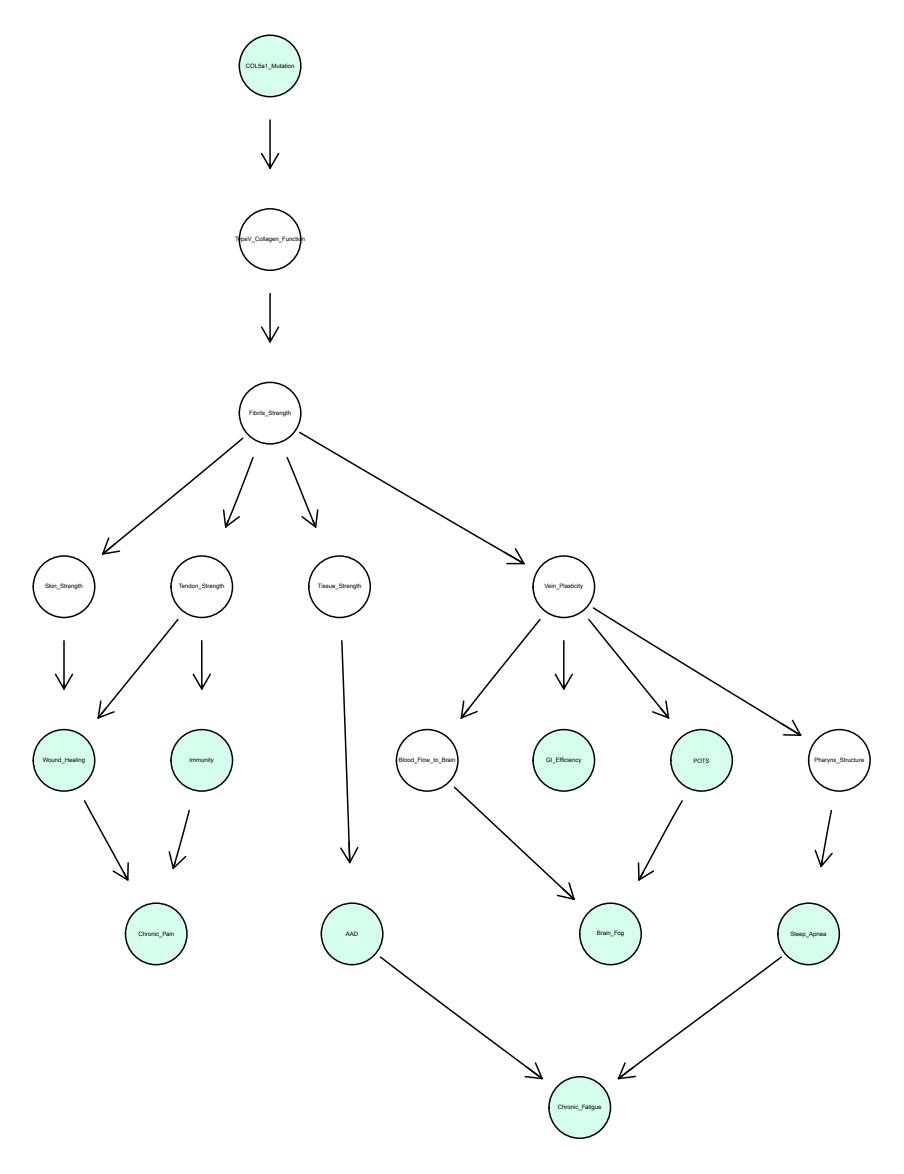


Figure 1: R bnlearn generated COL5A1 AOP

We can create a matrix for each patient we want to run through the AOP. For example, I created dummy data for Mark’s observed nodes below. We know that he has the mutation and we know what his symptoms are. We could do the same for Kira’s data.

#Mark

probsMark =

matrix(data = c(

0.0, 0.0, 1.0, # COL5A1\_Mutation ("None", "Heterozygous", "Homozygous")

0.3, 0.7, 0.0, # AAD ("None", "Mild", "Severe")

0.9, 0.1, 0.8, # Immunity ("Weak", "Ok", "Strong")

0.6, 0.4, 0.9, # Wound\_Healing ("Weak", "Ok", "Strong")

0.8, 0.2, 1.0, # Chronic\_Pain ("None", "Mild", "Severe")

0.2, 0.8, 0.6, # GI\_Efficiency ("Weak", "Ok", "Strong")

0.7, 0.3, 0.2, # POTS ("None", "Mild", "Severe")

0.0, 0.5, 0.5, # Brain\_Fog ("None", "Mild", "Severe")

0.0, 0.8, 0.2, # Sleep\_Apnea ("None", "Mild", "Severe")

0.0, 0.8, 0.2), # Chronic\_Fatigue ("None", "Mild", "Severe")

nrow = 10, ncol = 3, byrow = TRUE,

dimnames = list(c("COL5A1\_Mutation","AAD", "Immunity", "Wound\_Healing",

"Chronic\_Pain", "GI\_Efficiency", "POTS", "Brain\_Fog",

"Sleep\_Apnea", "Chronic\_Fatigue"),c("1","2","3"))

)

# Set the input probabilities for each run (each run = each patient)

probsUse <- probsMark

numEnts=1 #because we only have Mark right now

# Convert the chosen input probabilities to a data frame

probs <- as.list(data.frame(t(probsUse)))

Now we need to set the CPTs, or Conditional Probability Tables, for each node as I do below for COL5A1. We can’t do this until we determine the probability functions the edges represent. These will be determined by smaller networks (see Nikki’s work later in this document).

COL5A1\_Mutation.vals <- c("None", "Heterozygous", "Homozygous")

COL5A1\_Mutation.cpt <-

array(probs$COL5A1\_Mutation,

dim=3,

dimnames = list(COL5A1\_Mutation=COL5A1\_Mutation.vals))

#need to create CPTs for each node

Below "fits" the dag and cpt together into a complete network model. This step means we have connected our probability functions to our network and can now run Mark and Kira’s respective observed variables through the node.

bn = custom.fit(dag, cpt)

# or if any of the nodes are ordinal discrete, we can add them in a list

# bn = custom.fit(dag, cpt, ordinal=c(" Chronic\_Fatigue ","Brain\_Fog"))

Below plots nodes and chart probabilities to produce the probability of Brain Fog or CFS for each patient.

graphviz.chart(bn, type = "barprob", grid = TRUE, bar.col = "darkgreen", strip.bg = "lightskyblue")

# Turn data into a data frame

dfResults <- data.frame(Company=results[2,], Probability=as.numeric(results[1,]))

dfResults

# \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

# Plot the Results

# Set up dynamic coloring for the bar chart

topVal <- max(dfResults$Probability,0.60) # Set top value for coloring, with 0.60 minimum

palette <- colorRampPalette(brewer.pal(9, 'RdYlGn'))

ncols <- topVal\*1000

allColors <- palette(ncols)#apply the function to get 100 colours

colors <- array(dim=numEnts)

for(i in 1:numEnts) {

colors[i] <- allColors[as.numeric(dfResults$Probability[i])\*1000]

}

p<-ggplot(data=dfResults,

aes(x=Company, y=Probability)) + # Order alphabetically

geom\_bar(stat="identity", color="black", fill=colors) +

geom\_text(aes(label=format(round(Probability, 3))), vjust=1.6, color="darkblue", size=3.5) +

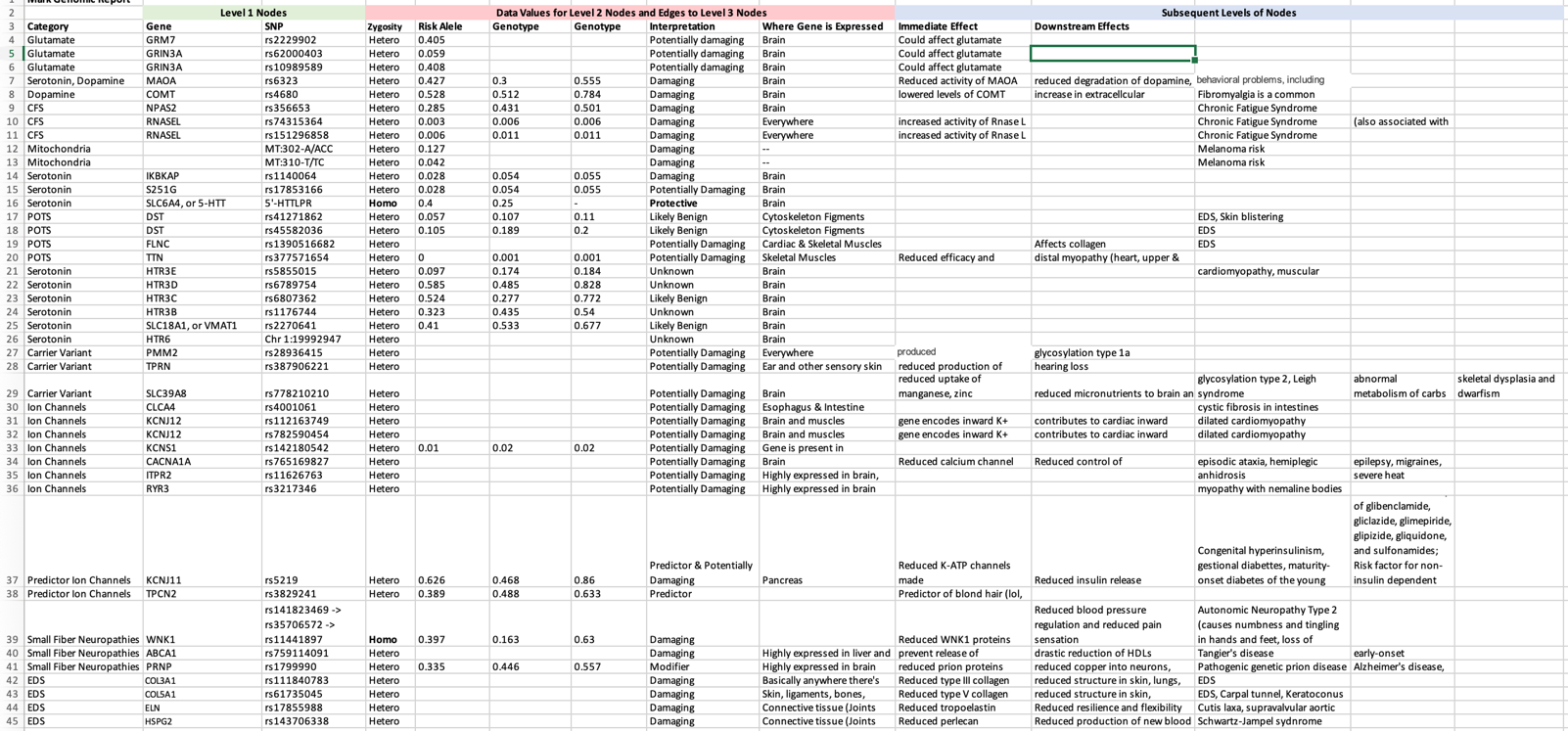
ggtitle(expression(atop("Investment Probabilities", atop(italic("Bayesian Network Analysis")))))

p

The next step will be to add treatment nodes as parents to the intermediary nodes. We can then run Mark and Kira’s data through the AOP again, and see how their symptoms (primarily Brain Fog and Chronic Fatigue in this model) are affected. We can also propose experiments to transfer more of the unobserved nodes to observed nodes.

### Target Mutations AOP:

Casey compiled the data from Mark’s genomics report and researched potential downstream effects of each genetic mutation, using the Online Mendelian Database for Man ([omim.org](https://www.omim.org/)) database to determine most likely targets for modeling. This 'compilation’ spreadsheet is included in the Dropbox, but a screenshot is pasted below for reference.

Table 1: Mark Genomic Data with Downstream Effects

Casey identified potentially damaging suspect genes a created an AOP in PowerPoint (PPT) to determine parent-child relationships. The genomic variants of greatest interest based on her research are highlighted in red.

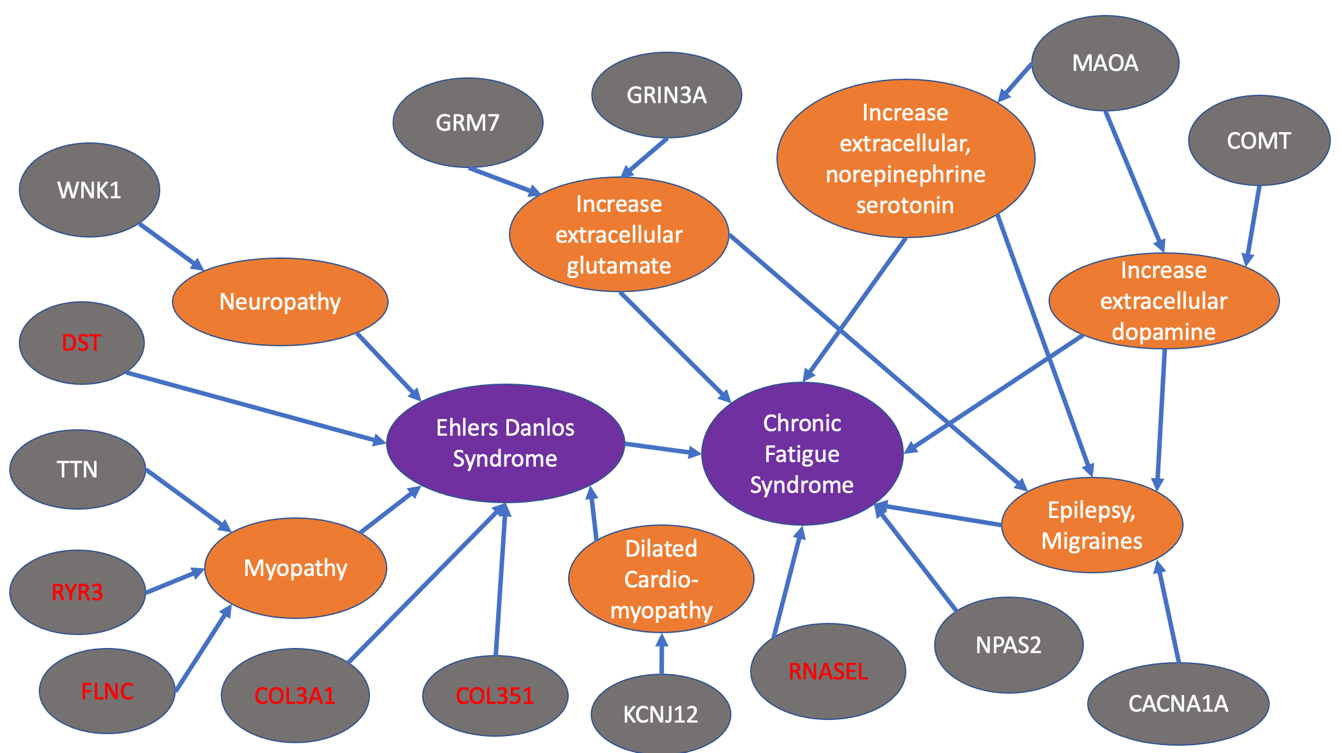


Figure 2: PPT generated Mark Genomic Targets of Interest AOP

She then created this AOP in the R bnlearn package using the same code as described above, in which teal highlighted nodes are observed (Mark’s data is available) and white nodes are unobserved. This was the resulting plot:

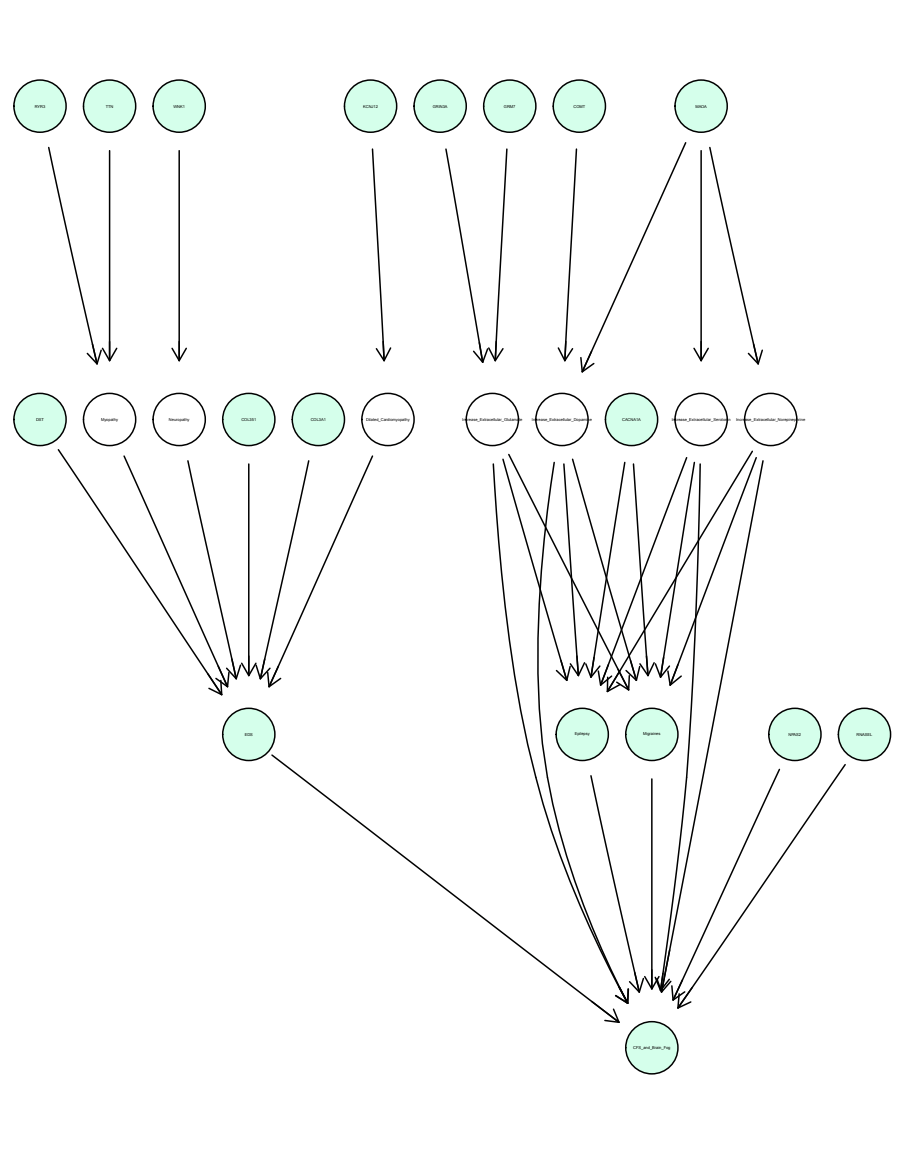


Figure 3: R bnlearn generated Mark Genomic Targets of Interest AOP

These AOPs can only be computationally useful once the probability functions (edges) are determined and justified. We plan to do so with biochemical modeling. Further, any node currently marked as ‘unobserved’ should be examined to determine if it *can* be observed via experimentation. This will strengthen the model.

Literature

## AOP Articles

Foran, C. M., Rycroft, T., Keisler, J., Perkins, E. J., Linkov, I., & Garcia-Reyero, N. (2019). A modular approach for assembly of quantitative adverse outcome pathways. ALTEX-Alternatives to animal experimentation, 36(3), 353-362.

Halappanavar, S., van den Brule, S., Nymark, P. et al. Adverse outcome pathways as a tool for the design of testing strategies to support the safety assessment of emerging advanced materials at the nanoscale. Part Fibre Toxicol 17, 16 (2020). <https://doi.org/10.1186/s12989-020-00344-4>

Knapen, D., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta‐Casaluci, L., ... & Villeneuve, D. L. (2018). Adverse outcome pathway networks I: development and applications. Environmental toxicology and chemistry, 37(6), 1723-1733.

Masison, J., Beezley, J., Mei, Y., Ribeiro, H. A. L., Knapp, A. C., Sordo Vieira, L., ... & Laubenbacher, R. (2021). A modular computational framework for medical digital twins. Proceedings of the National Academy of Sciences, 118(20), e2024287118.

Villeneuve, D. L., Angrish, M. M., Fortin, M. C., Katsiadaki, I., Leonard, M., Margiotta‐Casaluci, L., ... & Knapen, D. (2018). Adverse outcome pathway networks II: network analytics. Environmental toxicology and chemistry, 37(6), 1734-1748.

## Clinical Articles

### EDS

Johnston, J. M., Connizzo, B. K., Shetye, S. S., Robinson, K. A., Huegel, J., Rodriguez, A. B., ... & Soslowsky, L. J. (2017). Collagen V haploinsufficiency in a murine model of classic Ehlers–Danlos syndrome is associated with deficient structural and mechanical healing in tendons. Journal of Orthopaedic Research, 35(12), 2707-2715.

Laguette, M. J., Abrahams, Y., Prince, S., & Collins, M. (2011). Sequence variants within the 3′-UTR of the COL5A1 gene alters mRNA stability: implications for musculoskeletal soft tissue injuries. Matrix Biology, 30(5-6), 338-345.

Malfait, F., Wenstrup, R. J., & De Paepe, A. (2010). Clinical and genetic aspects of Ehlers-Danlos syndrome, classic type. Genetics in medicine, 12(10), 597-605.

Peltonen, L., Halila, R., & Ryhänen, L. (1985). Enzymes converting procollagens to collagens. Journal of cellular biochemistry, 28(1), 15-21.

Royer, S. P., & Han, S. J. (2022). Mechanobiology in the comorbidities of Ehlers Danlos syndrome. Frontiers in Cell and Developmental Biology, 710.

Teti, A. (1992). Regulation of cellular functions by extracellular matrix. Journal of the American Society of Nephrology, 2(10), S83.

Wilson, G. N., Tonk, S. S., Tonk, V. S., & Lampe, R. (2020). Complement Gene Mutation and Ehlers-Danlos Syndrome. Journal of Biosciences and Medicines, 8(06), 28.

## Graph Theory Articles

Bertsimas, D., Dunn, J., Gibson, E. *et al.* Optimal survival trees. *Mach Learn* **111**, 2951–3023 (2022). <https://doi.org/10.1007/s10994-021-06117-0>

Dimitris Bertsimas, Jack Dunn, Nishanth Mundru (2019) Optimal Prescriptive Trees. INFORMS Journal on Optimization 1(2):164-183. <https://doi.org/10.1287/ijoo.2018.0005>

Behrouz, A., Lécuyer, M., Rudin, C., & Seltzer, M. (2022). Fast optimization of weighted sparse decision trees for use in optimal treatment regimes and optimal policy design. arXiv preprint arXiv:2210.06825.

Liu, J., Zhong, C., Li, B., Seltzer, M., & Rudin, C. (2022). FasterRisk: Fast and Accurate Interpretable Risk Scores. arXiv preprint arXiv:2210.05846.