

Wnioskowanie w Warunkach Niepewności - Projekt

Mateusz Pindyk

169831

Dane

Fetal Health Classification

Classify the health of a fetus as Normal, Suspect or Pathological using CTG data

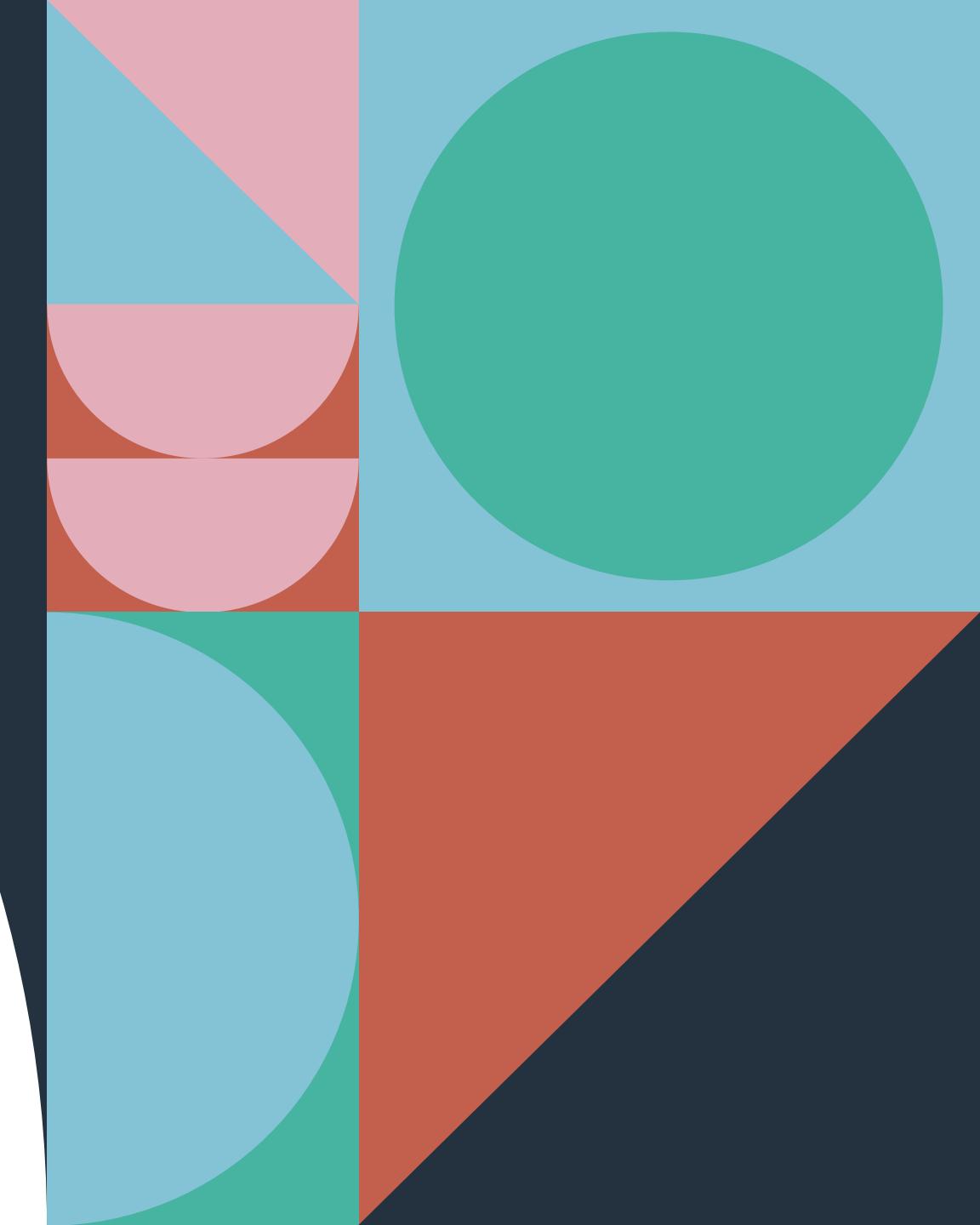


- Użyta baza danych (Fetal Health Classification) zawiera 2126 zapisów cech wyodrębnionych z badań kardiotokogramu, które następnie zostały sklasyfikowane przez położników w 3 klasy: NORMAL, SUSPECT, PATHOLOGICAL.
- Kardiotokografię (KTG) wykonuje się w czasie ciąży w celu monitorowania tętna płodu i skurczów macicy. Monitoruje dobrostan płodu i pozwala na wczesne wykrycie zagrożenia dla płodu.
- Interpretacja KTG pomaga w określeniu, czy ciąża jest wysokiego, czy niskiego ryzyka. Nieprawidłowy wynik KTG może wskazywać na potrzebę dalszych badań i potencjalnej interwencji.
- Źródło: <https://www.kaggle.com/datasets/andrewmvd/fetal-health-classification>

Dostępne dane

- 'baseline value- wartość bazowa FHR (ilość uderzeń serca płodu na minutę)
- 'accelerations' – liczba przyspieszeń na sekundę
- „fetal_movement” – liczba ruchów płodu na sekundę
- „uterine_contractions” – liczba skurczów macicy na sekundę
- „light_decelerations” – liczba lekkich spowolnień na sekundę
- „severe_decelerations” – liczba poważnych spowolnień na sekundę
- „prolongued_decelerations” – liczba przedłużonych opóźnień na sekundę
- „abnormal_short_term_variability” – procent czasu z nienormalną zmiennością krótkoterminową
- „mean_value_of_short_term_variability” – średnia wartość zmienności krótkoterminowej
- „percentage_of_time_with_abnormal_long_term_variability” - procent czasu z nienormalną zmiennością długoterminową
- „mean_value_of_long_term_variability” - średnia wartość zmienności długoterminowej
- „histogram_width” – szerokość histogramu FHR
- „histogram_min” – minimalna (niska częstotliwość) histogramu FHR
- 'histogram_max' - Maksymalna (wysoka częstotliwość) histogramu FHR
- „histogram_number_of_peaks” – liczba pików histogramu
- „histogram_number_of_zeroes” – liczba zer histogramu
- „histogram_mode” – tryb histogramu
- „histogram_mean” – średnia histogramu
- „histogram_median” – mediana histogramu
- „histogram_variance” – wariancja histogramu
- „histogram_tendency” – tendencja histogramu
- Cel

Praca w R



Załadowanie pliku i podstawowych bibliotek

```
#### MATEUSZ PINDYK ####  
#### PROJEKT ####  
#### Wnioskowanie w Warunkach Niepewności ####  
#### Fetal Health Classification ####
```

```
#setwd("")  
library(readr)  
library(dplyr)  
fetal_health <- read_csv("fetal_health.csv") #load dataset
```

Zamiana kolumny celu z danych liczbowych (1,2,3) na słowne

- Dane w kolumnie celu były sklasyfikowane za pomocą liczb 1 odpowiadającej za stan „Normalny”,
2 - „Podejrzany”,
3 - „Patologiczny”.
- Stwierdziłem, że czytelniej będzie zaprezentować te dane słownie.

```
#### declaring names for health conditions ####
fetal_health$fetal_health <- factor(
  fetal_health$fetal_health,
  levels = c(1, 2, 3),
  labels = c("NORMAL", "SUSPECT", "PATHOLOGICAL")
)
```

Zmniejszenie liczby kolumn

- Usunąłem 10 kolumn, które, według mnie miały mały wpływ na kolejne kolumny, co ułatwiło późniejszą analizę.

```
#### deleting spare columns ####
fetal_health <- fetal_health %>% select(-one_of(
  c(
    'prolongued_decelerations',
    'accelerations',
    'light_decelerations',
    'severe_decelerations',
    'histogram_tendency',
    "histogram_number_of_peaks",
    "histogram_number_of_zeroes",
    "histogram_mode",
    "histogram_median",
    "histogram_variance"
  )
))
```

Usunięcie skrajnych wartości

- By zachować spójność analizy, usunąłem skrajne wartości (minima i maksima) z kolumn, które je zawierały.

```
> ##### deleting minimal and maximal values #####
> nrow_fetal_bef <- nrow(fetal_health)
>
> columns_to_clean <- c(
+   "baseline_value",
+   "mean_value_of_short_term_variability",
+   "mean_value_of_long_term_variability",
+   "histogram_width",
+   "histogram_min",
+   "histogram_max",
+   "histogram_mean"
+ )
> # Function to remove rows with min and max values in a specific column
> clean_column <- function(data, column) {
+   min_val <- min(data[[column]], na.rm = TRUE)
+   max_val <- max(data[[column]], na.rm = TRUE)
+
+   # Remove rows with the min and max values
+   data <-
+     data[data[[column]] != min_val & data[[column]] != max_val, ]
+   return(data)
+ }
>
> # Apply the cleaning function to each column of interest
> for (column in columns_to_clean) {
+   fetal_health <- clean_column(fetal_health, column)
+ }
>
> cat(
+   "Deleted" ,
+   nrow_fetal_bef - nrow(fetal_health) ,
+   "rows containing minimal or maximal values."
+ )
Deleted 264 rows containing minimal or maximal values.
```

Grupowanie danych

- W celu stworzenia węzłów, pogrupowałem dane w pętli na 5 grup

```
##### grouping data #####
data_group <- function(data_select) {  
  #fetal_health$baseline_value_group <- data_group(fetal_health$`baseline_...`)  
  
  # List of column names to process (excluding the 12th column)
  names_of_columns <- names(fetal_health)
  names_of_columns <- names_of_columns[-12]  
  
  # Loop over the columns and create new grouped columns
  for (i in names_of_columns) {
    new_column_name <- paste(i, 'group', sep = "_")
    fetal_health[[new_column_name]] <-
      as.factor(data_group(fetal_health[[i]])) #as factors
  }  
  
  # New data frame with groups
  fetal_health_grouped <- fetal_health[, c(13:23, 12)]
  names(fetal_health_grouped)[1] <- "baseline_value_group"
```

Grupowanie danych

Przed

baseline value	fetal_movement	uterine_contractions	abnormal_short_term_variability	
120	0.000	0.000		73
132	0.000	0.006		17
133	0.000	0.008		16
134	0.000	0.008		16
132	0.000	0.008		16
122	0.000	0.000		83
122	0.000	0.002		84
122	0.000	0.003		86
151	0.000	0.001		64
150	0.000	0.001		64
131	0.072	0.008		28
131	0.222	0.006		28
130	0.222	0.004		21
130	0.380	0.004		19
130	0.441	0.005		24
131	0.383	0.003		18

Po

baseline_value_group	fetal_movement_group	uterine_contractions_group	abnormal_short_term_var
(120,130]	[-1,0.0962]	[-1,0.003]	(71.2,87]
(130,139]	[-1,0.0962]	(0.003,0.006]	[11,26.8]
(130,139]	[-1,0.0962]	(0.006,0.009]	[11,26.8]
(130,139]	[-1,0.0962]	(0.006,0.009]	[11,26.8]
(130,139]	[-1,0.0962]	(0.006,0.009]	[11,26.8]
(120,130]	[-1,0.0962]	[-1,0.003]	(71.2,87]
(120,130]	[-1,0.0962]	[-1,0.003]	(71.2,87]
(120,130]	[-1,0.0962]	[-1,0.003]	(71.2,87]
(149,160]	[-1,0.0962]	[-1,0.003]	(56.4,71.2]
(149,160]	[-1,0.0962]	[-1,0.003]	(56.4,71.2]
(130,139]	[-1,0.0962]	(0.006,0.009]	(26.8,41.6]
(130,139]	(0.192,0.289]	(0.003,0.006]	(26.8,41.6]
(130,139]	(0.385,1.48]	(0.003,0.006]	[11,26.8]
(130,139]	(0.289,0.385]	(0.003,0.006]	[11,26.8]
(130,139]	(0.385,1.48]	(0.003,0.006]	[11,26.8]
(130,139]	(0.289,0.385]	[-1,0.003]	[11,26.8]

Badanie niezależności

- Zbadałem niezależność węzłów za pomocą funkcji ci.test z biblioteki bnlearn, wpisałem je do ramki danych, oraz stworzyłem ramkę wypisującą, czy p-value jest mniejsze od 0.05.

```
#### independence tests ####

library(bnlearn)

# Create an empty data frame
independence_matrix <-
  data.frame(matrix(
    ncol = ncol(fetal_health_grouped),
    nrow = ncol(fetal_health_grouped)
  ))

# Assign the column names from fetal_health_grouped to the empty data frame
colnames(independence_matrix) <- colnames(fetal_health_grouped)
rownames(independence_matrix) <- colnames(fetal_health_grouped)

fetal_health_grouped <-
  as.data.frame(fetal_health_grouped) #conv to df due it only works with bnlearn tests

for (i in names(independence_matrix)) {
  for (j in names(independence_matrix)) {
    if (i != j) {
      test_result <-
        ci.test(i, j, test = "x2", data = fetal_health_grouped)
      p_value <- test_result$p.value
      independence_matrix[i, j] <- p_value
    }
  }
}

independence_matrix_less <- independence_matrix < 0.05
```

Macierz niezależności

	baseline_value_group	fetal_movement_group	uterine_contractions_group	abnormal_short_term_variability_group	mean_value_of_short_term_variability_group	percentage_of_time_with_abnormal_long_term_variability_group	mean_value_of_long_term_variability_group	histogram_width_group	histogram_min_group	histogram_max_group	histogram_mean_group	fetal_health
baseline_value_group	NA	3.345372e-03	4.652211e-15	1.300484e-80	6.605652e-32	5.953408e-33	4.709090e-14	1.182457e-15	2.416651e-109	5.720400e-81	0.000000e+00	1.504267e-52
fetal_movement_group	3.345372e-03	NA	2.559166e-02	3.729193e-06	4.576707e-08	9.621242e-01	5.263686e-01	4.371323e-09	1.512357e-06	1.034925e-02	3.665136e-08	3.962044e-30
uterine_contractions_group	4.652211e-15	2.559166e-02	NA	1.754081e-18	1.521824e-11	2.551527e-09	2.912849e-01	1.687793e-09	1.069190e-10	9.430463e-05	1.869777e-06	1.216046e-29
abnormal_short_term_variability_group	1.300484e-80	3.729193e-06	1.754081e-18	NA	7.957782e-94	2.887073e-76	9.679175e-35	5.077835e-41	8.403389e-54	7.828743e-18	4.310307e-71	1.093344e-106
mean_value_of_short_term_variability_group	6.605652e-32	4.576707e-08	1.521824e-11	7.957782e-94	NA	1.462914e-35	1.360846e-37	2.478902e-148	1.128413e-103	2.489593e-53	1.277597e-53	2.985231e-23
percentage_of_time_with_abnormal_long_term_variability_group	5.953408e-33	9.621242e-01	2.551527e-09	2.887073e-76	1.462914e-35	NA	2.983746e-23	1.177173e-62	1.555343e-55	1.161523e-18	7.468627e-12	8.695684e-149
mean_value_of_long_term_variability_group	4.709090e-14	5.263686e-01	2.912849e-01	9.679175e-35	1.360846e-37	2.983746e-23	NA	5.389369e-28	3.286182e-32	4.758397e-08	1.081246e-02	1.251103e-05
histogram_width_group	1.182457e-15	4.371323e-09	1.687793e-09	5.077835e-41	2.478902e-148	1.177173e-62	5.389369e-28	NA	0.000000e+00	6.743849e-252	3.176054e-14	1.368078e-29
histogram_min_group	2.416651e-109	1.512357e-06	1.069190e-10	8.403389e-54	1.128413e-103	1.555343e-55	3.286182e-32	0.000000e+00	NA	2.984495e-43	8.404605e-112	2.311632e-51
histogram_max_group	5.720400e-81	1.034925e-02	9.430463e-05	7.828743e-18	2.489593e-53	1.161523e-18	4.758397e-08	6.743849e-252	2.984495e-43	NA	3.728988e-113	4.044406e-04
histogram_mean_group	0.000000e+00	3.665136e-08	1.869777e-06	4.310307e-71	1.277597e-53	7.468627e-12	1.081246e-02	3.176054e-14	8.404605e-112	3.728988e-113	NA	6.816116e-87
fetal_health	1.504267e-52	3.962044e-30	1.216046e-29	1.093344e-106	2.985231e-23	8.695684e-149	1.251103e-05	1.368078e-29	2.311632e-51	4.044406e-04	6.816116e-87	NA

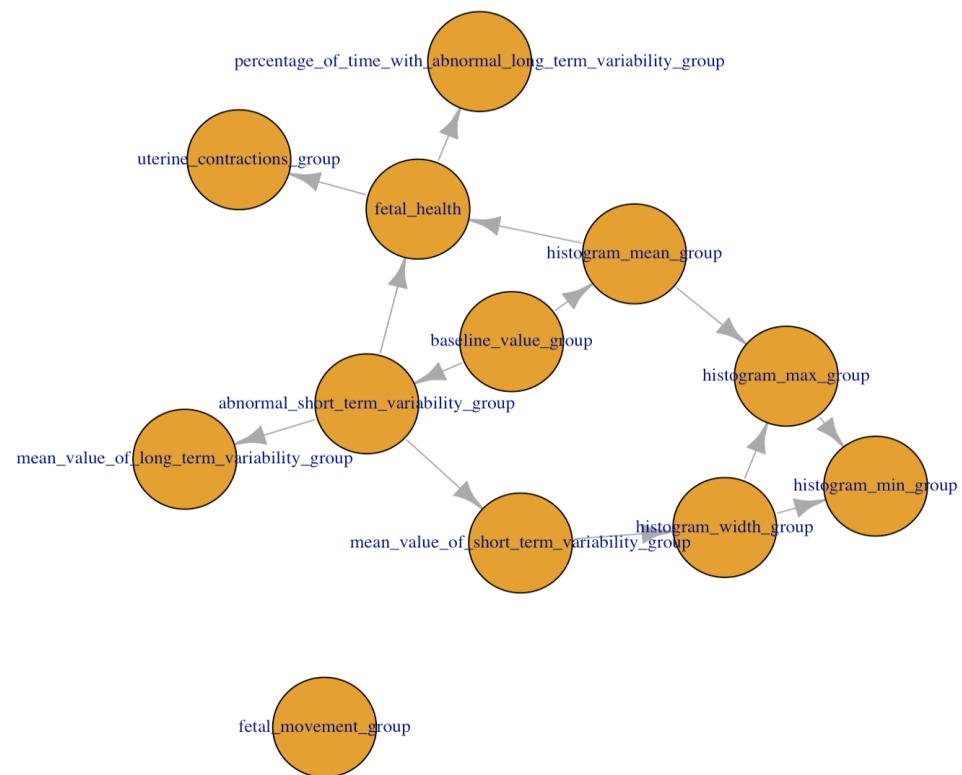
Macierz p-value < 0.05

Sieć HC

```
#### HC Network ####

library(igraph)
hc_network <- hc(fetal_health_grouped)
# Convert the Bayesian network to an igraph object
bn_igraph <- as.igraph(hc_network)

# Plot the igraph object
plot(
  bn_igraph,
  vertex.size = 30,
  vertex.label.cex = 0.8,
  edge.arrow.size = 0.5
)
```



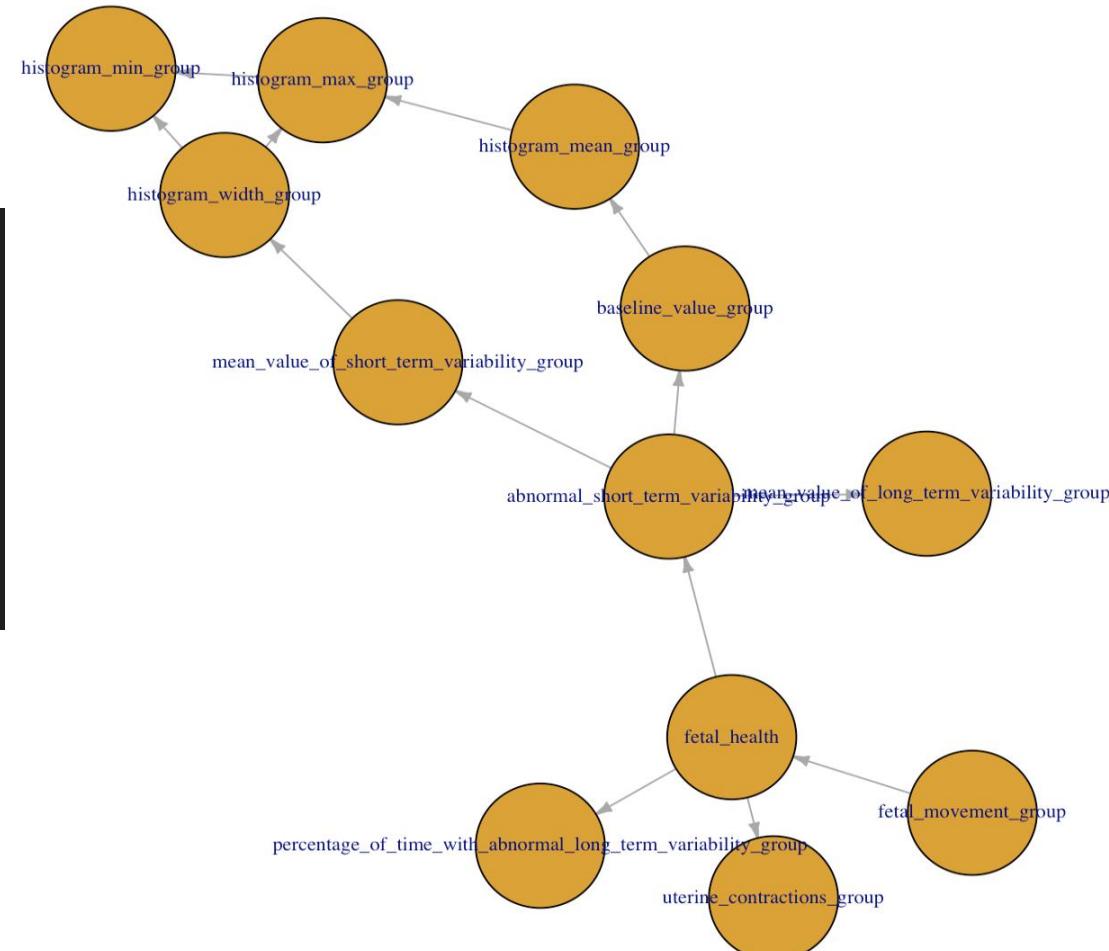
Sieć HC

Niestety w grafie brakowało jednego z węzłów. Dodałem go uwzględniające dane dotyczące niezależności.

```
# connect the missing node
hc_network <-
  hc(fetal_health_grouped, whitelist = matrix(c("fetal_movement_group", "fetal_health"), ncol = 2))

bn_igraph <- as.igraph(hc_network)

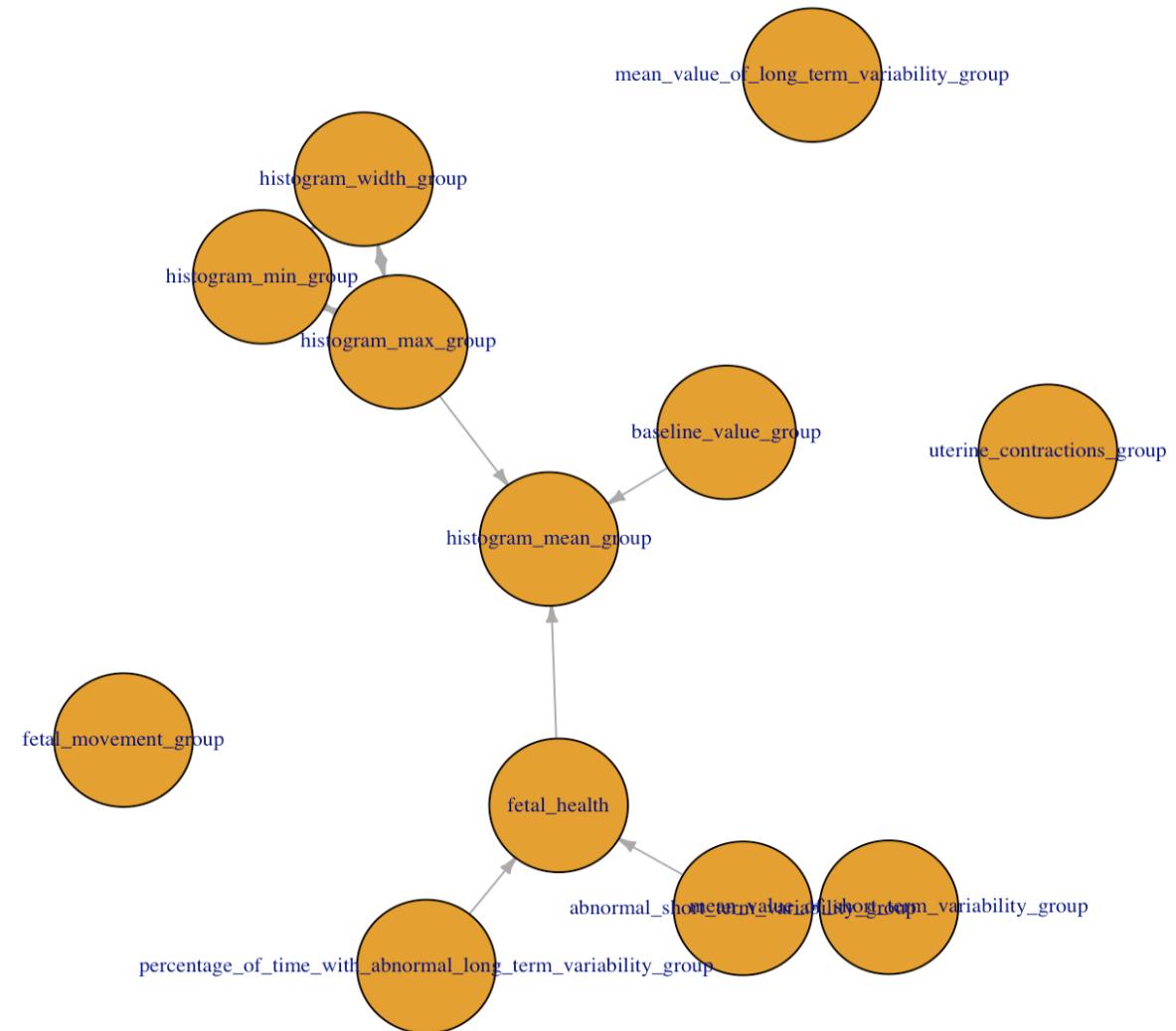
plot(
  bn_igraph,
  vertex.size = 30,
  vertex.label.cex = 0.7,
  edge.arrow.size = 0.2
)
```



Sieć IAMB

```
#### IAMB Network ####
iamb_network <- iamb(fetal_health_grouped)
bn_igraph <- as.igraph(iamb_network)

plot(
  bn_igraph,
  vertex.size = 30,
  vertex.label.cex = 0.7,
  edge.arrow.size = 0.2
)
```



Sieć IAMB

Ta sieć okazała się jednak bardziej wymagająca do dopracowania.

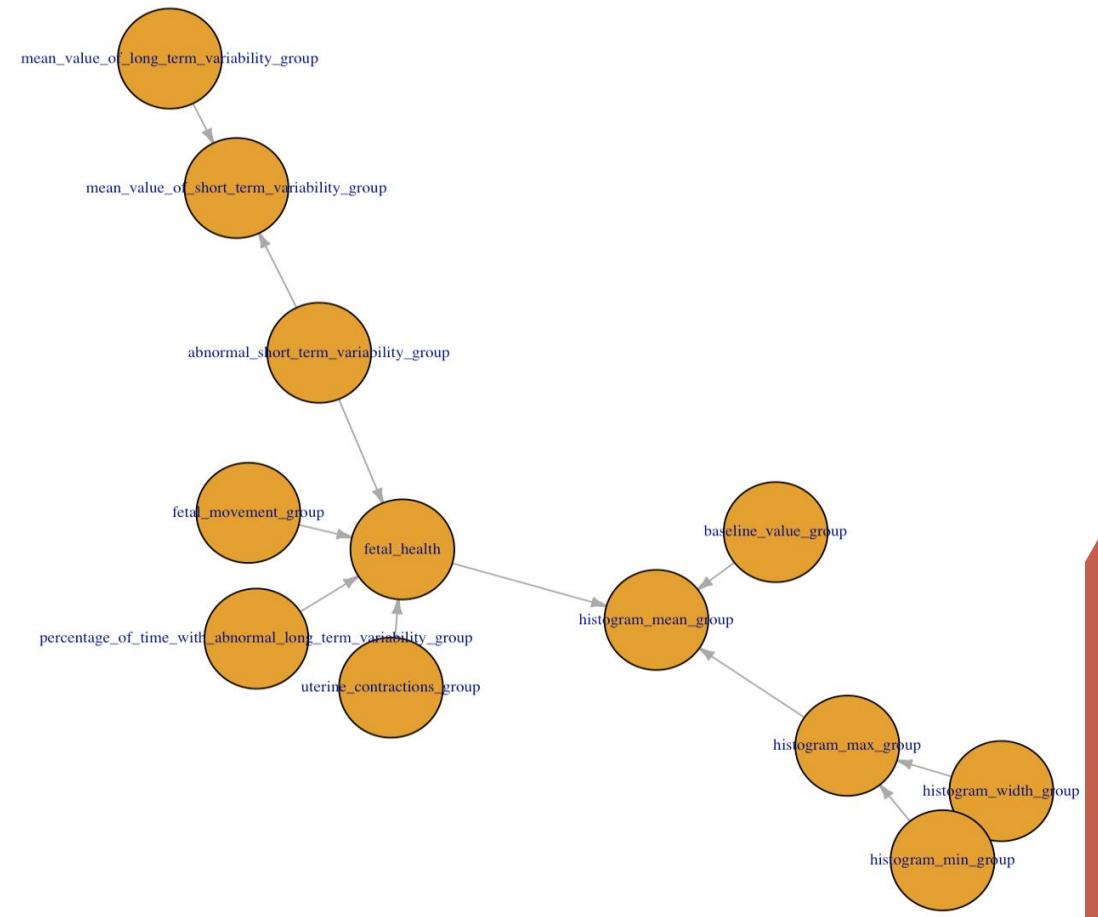
```
# connect the missing nodes

iamb_network <-
  iamb(fetal_health_grouped, whitelist = matrix(
    c(
      "fetal_movement_group",
      "fetal_health",
      "uterine_contractions_group" ,
      "fetal_health",
      "percentage_of_time_with_abnormal_long_term_variability_group",
      "fetal_health",
      "mean_value_of_long_term_variability_group",
      "mean_value_of_short_term_variability_group",
      "abnormal_short_term_variability_group",
      "fetal_health",
      "fetal_health",
      "histogram_mean_group"
    ),
    ncol = 2,
    byrow = T
  ))

iamb_network <- set.arc(iamb_network,      "histogram_width_group",
                         "histogram_max_group")
iamb_network <- set.arc(iamb_network,      "histogram_min_group",
                         "histogram_max_group")
iamb_network <- set.arc(iamb_network,      "histogram_min_group",
                         "histogram_width_group")

bn_igraph <- as.igraph(iamb_network)

plot(
  bn_igraph,
  vertex.size = 25,
  vertex.label.cex = 0.6,
  edge.arrow.size = 0.2|
```

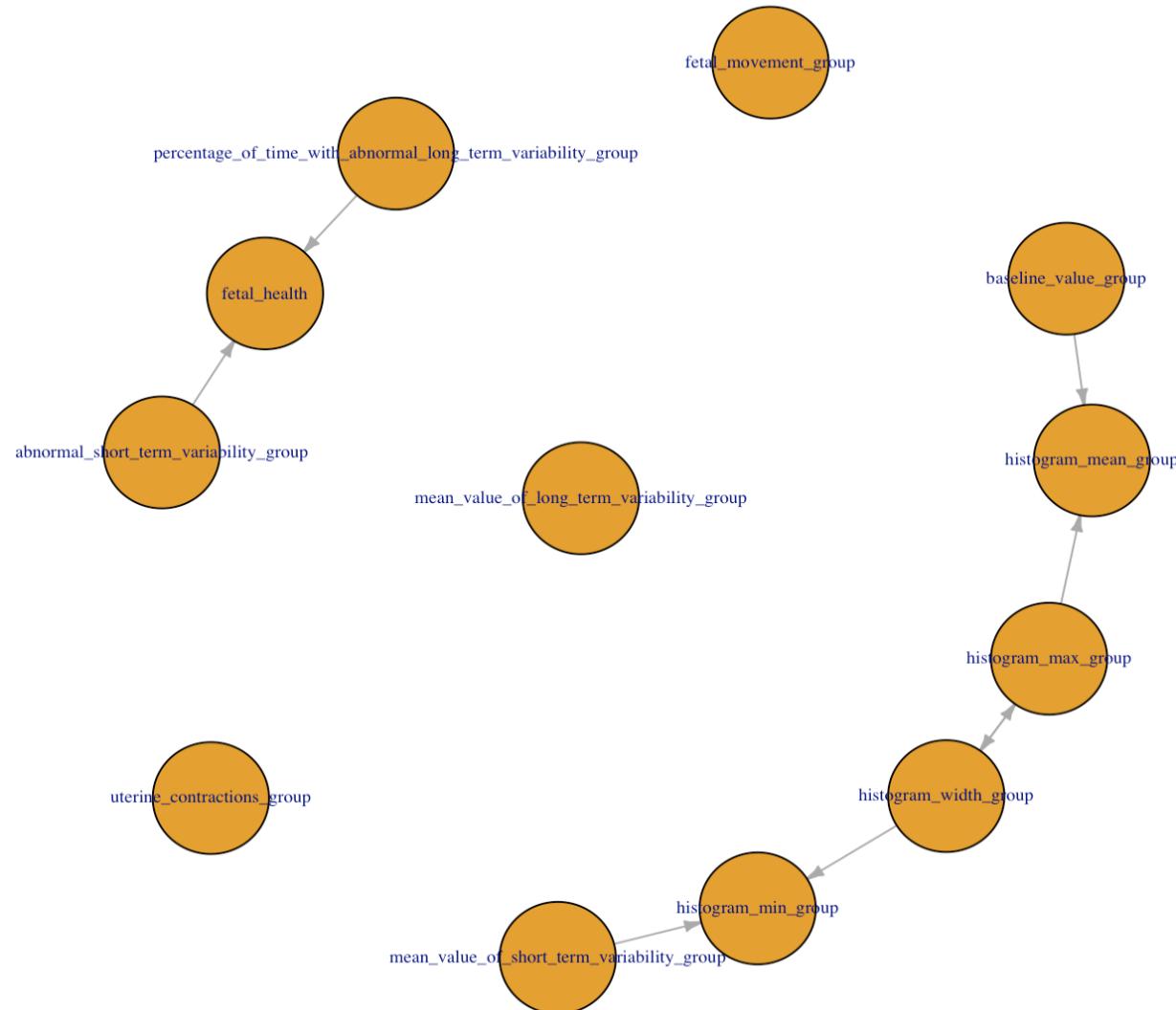


Sieć Fast.IAMB

```
#### Fast.IAMB Network ####
fast_iamb_network <- fast.iamb(fetal_health_grouped)

bn_igraph <- as.igraph(fast_iamb_network)

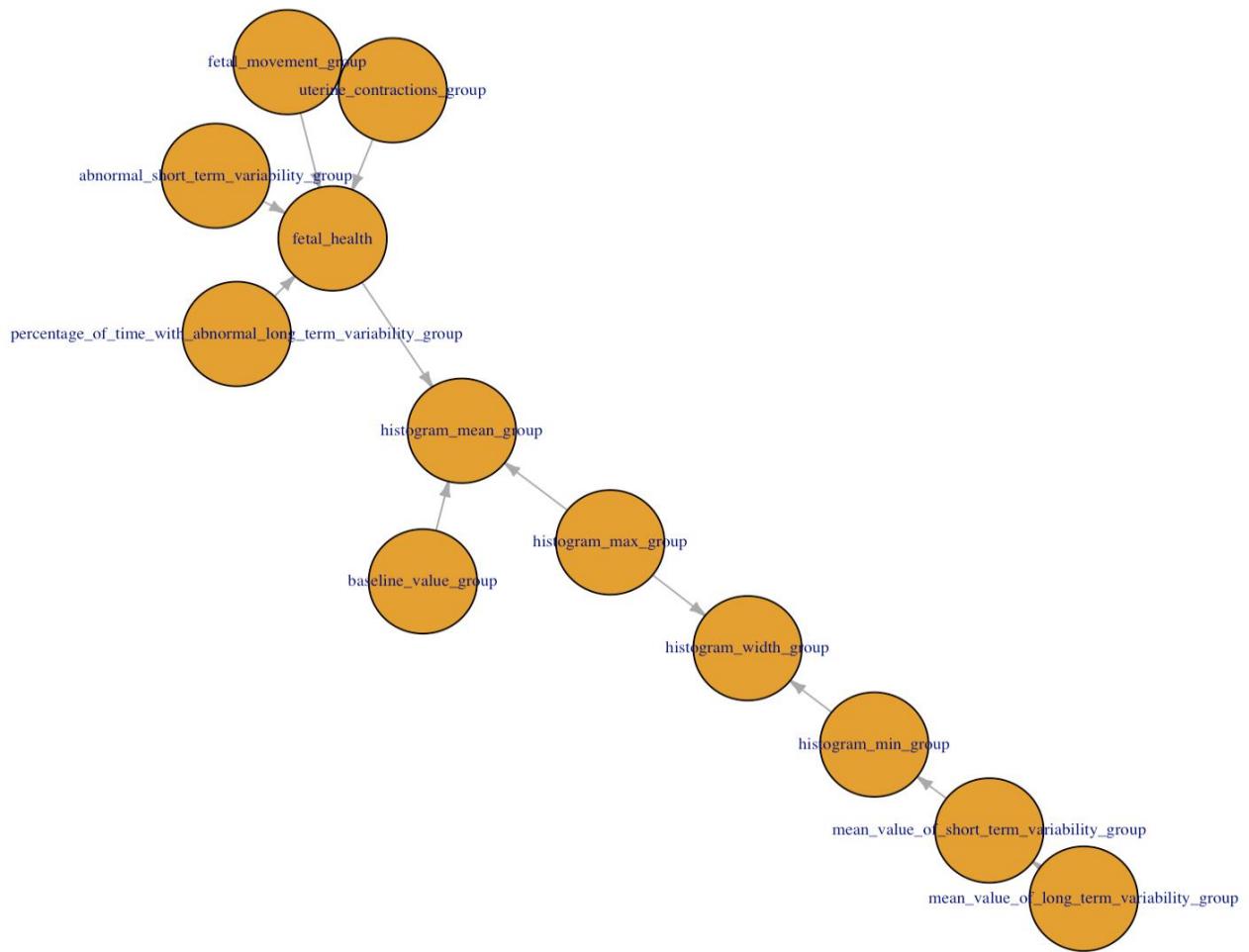
plot(
  bn_igraph,
  vertex.size = 25,
  vertex.label.cex = 0.6,
  edge.arrow.size = 0.2
)
```



Sieć Fast.IAMB

Kolejne własnoręczne połączenia

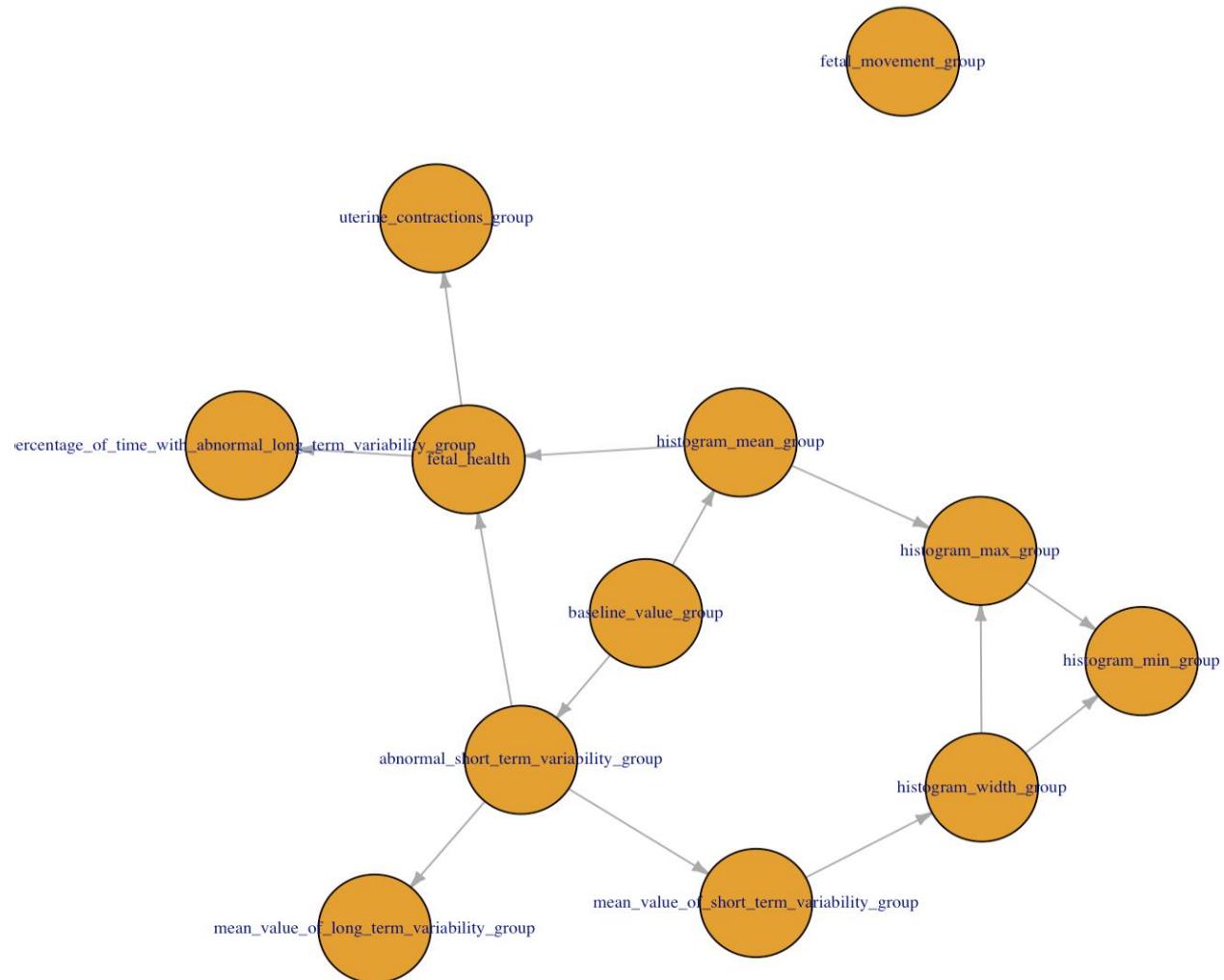
```
fast_iamb_network <-  
  fast.iamb(fetal_health_grouped, whitelist = matrix(  
    c(  
      "mean_value_of_long_term_variability_group" ,  
      "mean_value_of_short_term_variability_group",  
      "mean_value_of_short_term_variability_group",  
      "histogram_min_group",  
      "histogram_max_group",  
      "histogram_mean_group",  
      "baseline_value_group",  
      "histogram_mean_group",  
      "uterine_contractions_group" ,  
      "fetal_health",  
      "percentage_of_time_with_abnormal_long_term_variability_group",  
      "fetal_health",  
      "abnormal_short_term_variability_group",  
      "fetal_health",  
      "fetal_movement_group",  
      "fetal_health",  
      "fetal_health",  
      "histogram_mean_group"  
    ),  
    ncol = 2,  
    byrow = T  
  ))  
  
bn_igraph <- as.igraph(fast_iamb_network)  
plot(  
  bn_igraph,  
  vertex.size = 25,  
  vertex.label.cex = 0.6,  
  edge.arrow.size = 0.2  
)
```



Sieć TABU

```
tabu_network <- tabu(fetal_health_grouped)

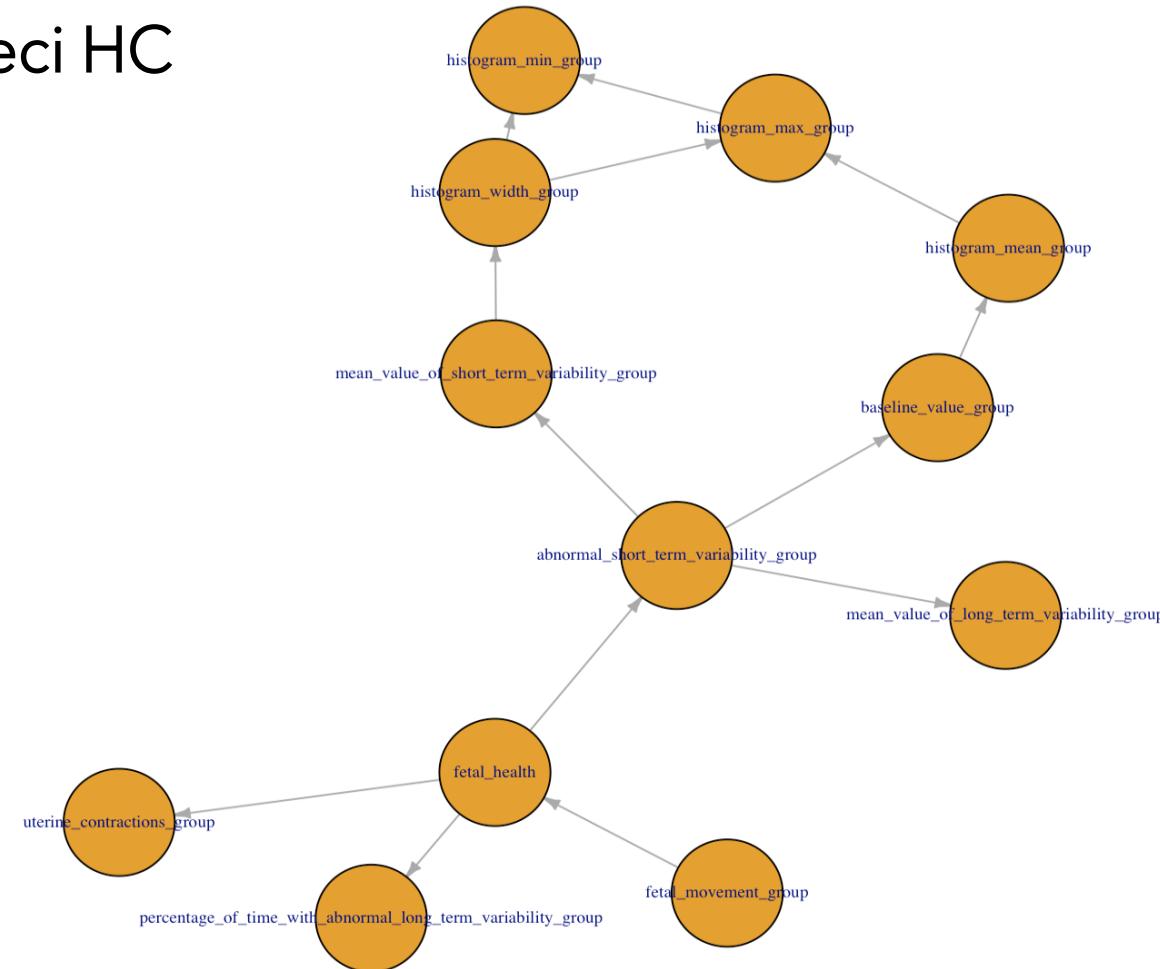
bn_igraph <- as.igraph(tabu_network)
plot(
  bn_igraph,
  vertex.size = 25,
  vertex.label.cex = 0.6,
  edge.arrow.size = 0.2
)
```



Sieć TABU

Kolejne połączenie – takie samo jak w sieci HC

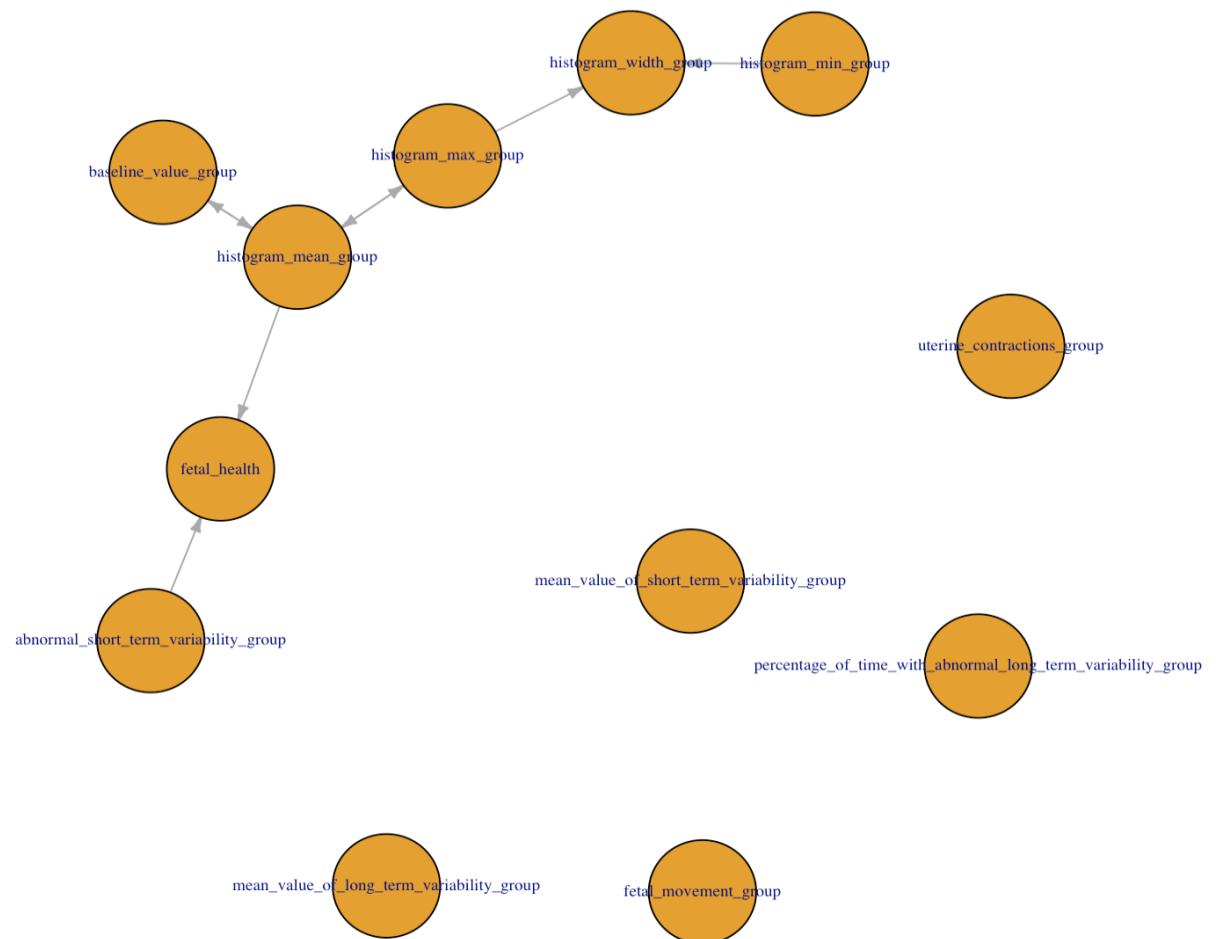
```
tabu_network <-
  tabu(fetal_health_grouped, whitelist = matrix(c("fetal_movement_group", "fetal_health"), ncol = 2))
bn_igraph <- as.igraph(tabu_network)
plot(
  bn_igraph,
  vertex.size = 25,
  vertex.label.cex = 0.6,
  edge.arrow.size = 0.2
)
```



Sieć PC.Stable

```
#### PC.STABLE Network ####
pc_network <- pc.stable(fetal_health_grouped)

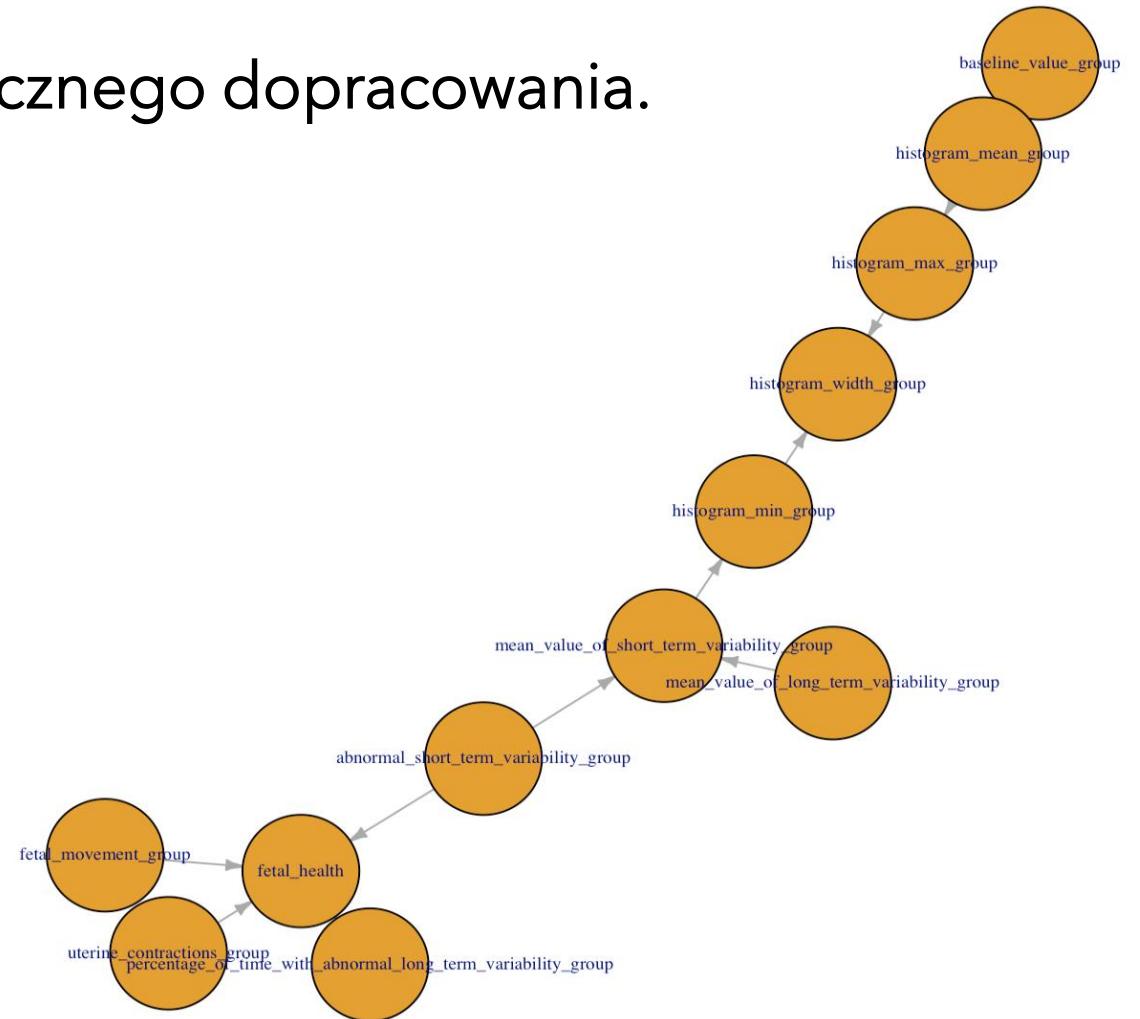
bn_igraph <- as.igraph(pc_network)
plot(
  bn_igraph,
  vertex.size = 25,
  vertex.label.cex = 0.6,
  edge.arrow.size = 0.2
)
```



Sieć PC.Stable

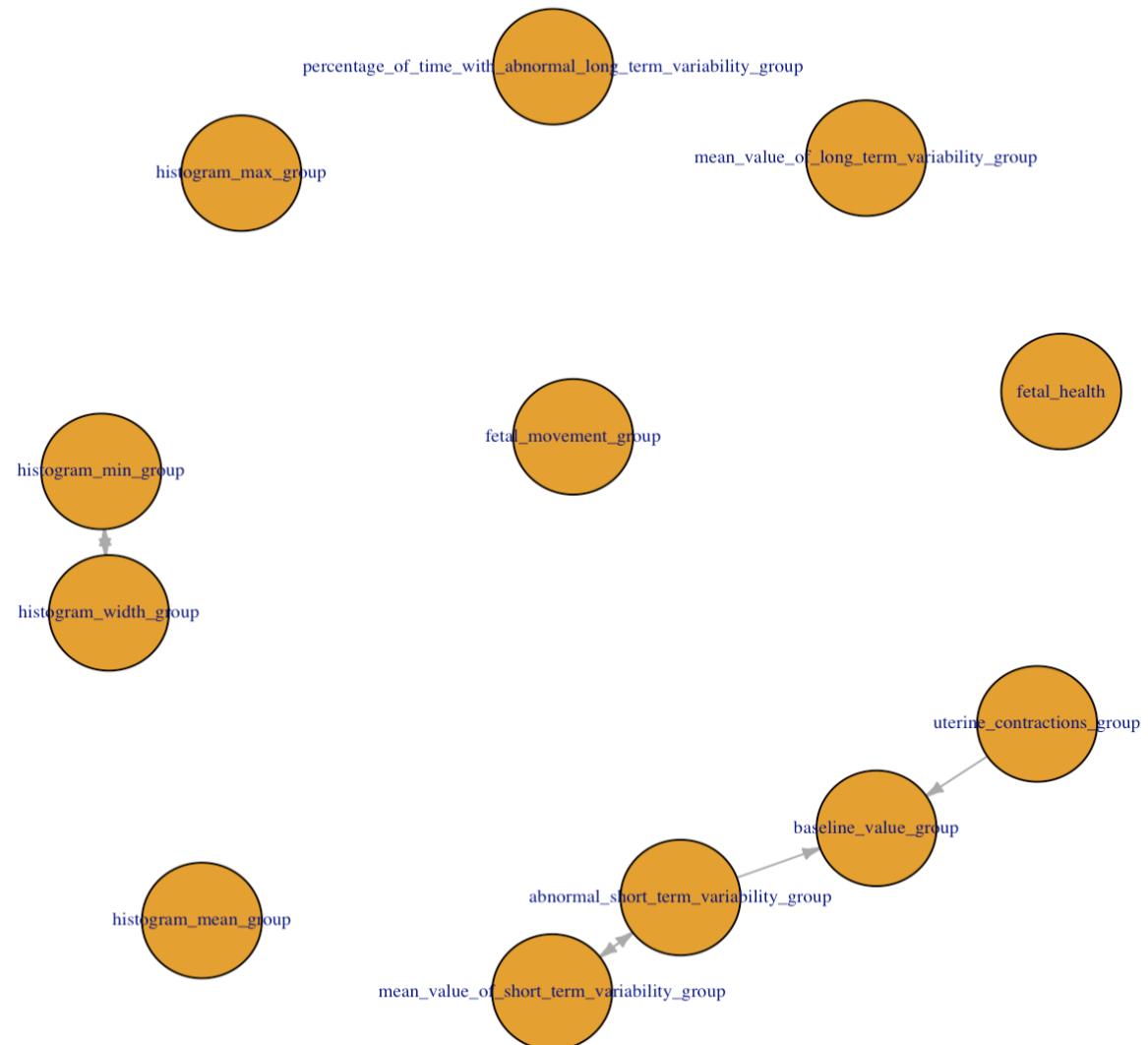
Sieć PC.Stable również wymagała ręcznego dopracowania.

```
pc_network <-  
  pc.stable(fetal_health_grouped, whitelist = matrix(  
    c(  
      "fetal_movement_group",  
      "fetal_health",  
      "mean_value_of_long_term_variability_group" ,  
      "mean_value_of_short_term_variability_group",  
      "mean_value_of_short_term_variability_group",  
      "histogram_min_group",  
      "uterine_contractions_group" ,  
      "fetal_health",  
      "percentage_of_time_with_abnormal_long_term_variability_group",  
      "fetal_health",  
      "abnormal_short_term_variability_group",  
      "fetal_health",  
      "abnormal_short_term_variability_group",  
      "mean_value_of_short_term_variability_group"  
    ),  
    ncol = 2,  
    byrow = T  
  ))  
pc_network <-  
  set.arc(pc_network, "histogram_mean_group", "histogram_max_group")  
pc_network <-  
  set.arc(pc_network, "baseline_value_group", "histogram_mean_group")  
bn_igraph <- as.igraph(pc_network)  
plot(  
  bn_igraph,  
  vertex.size = 25,  
  vertex.label.cex = 0.6,  
  edge.arrow.size = 0.2  
)
```



Sieć GS

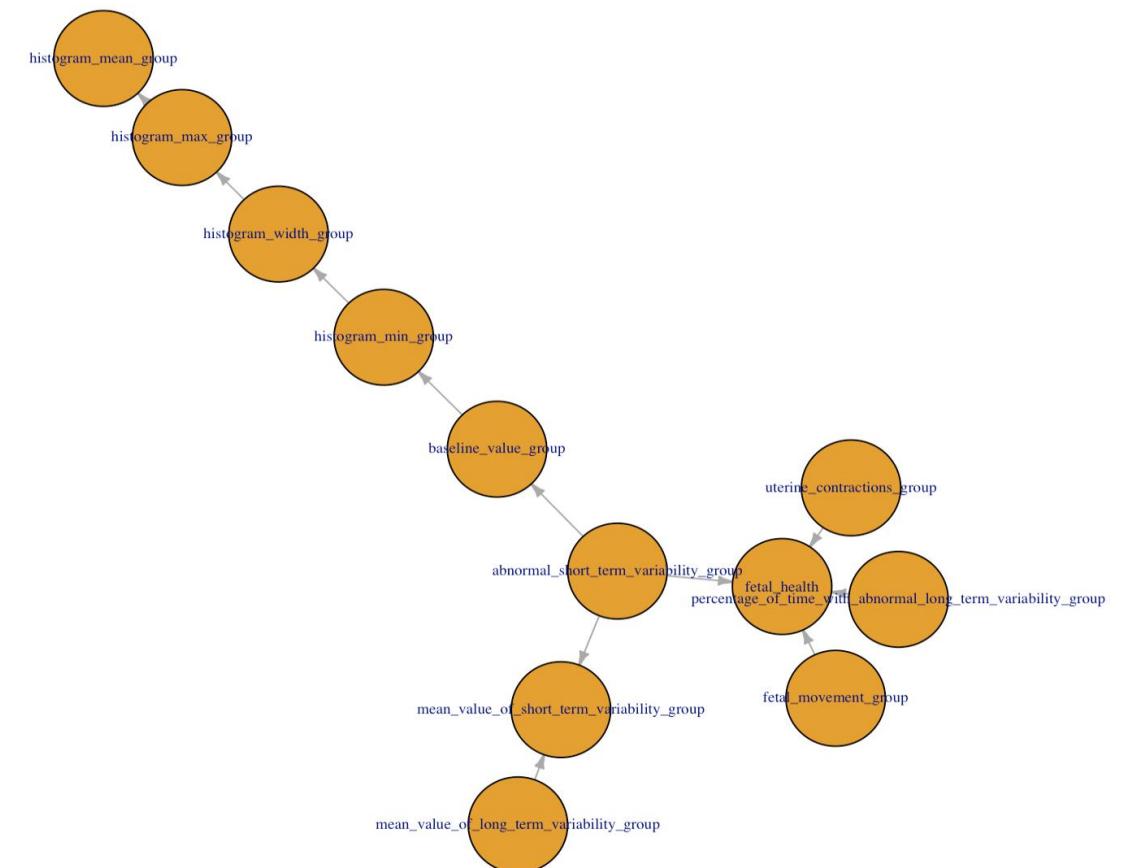
```
#### GS Network ####  
gs_network <- gs(fetal_health_grouped)  
  
bn_igraph <- as.igraph(gs_network)  
plot(  
  bn_igraph,  
  vertex.size = 25,  
  vertex.label.cex = 0.6,  
  edge.arrow.size = 0.2  
)
```



Sieć GS

Ta sieć również nie była w pełni poprawna i wymagała ręcznych połączeń.

```
gs_network <-  
  gs(fetal_health_grouped, whitelist = matrix(  
    c(  
      "histogram_width_group",  
      "histogram_max_group",  
      "baseline_value_group",  
      "histogram_min_group",  
      "abnormal_short_term_variability_group",  
      "fetal_health",  
      "percentage_of_time_with_abnormal_long_term_variability_group",  
      "fetal_health",  
      "histogram_max_group",  
      "histogram_mean_group",  
      "fetal_movement_group",  
      "fetal_health",  
      "uterine_contractions_group",  
      "fetal_health",  
      "mean_value_of_long_term_variability_group",  
      "mean_value_of_short_term_variability_group",  
      "abnormal_short_term_variability_group",  
      "mean_value_of_short_term_variability_group"  
    ),  
    ncol = 2,  
    byrow = T  
  ))  
  
gs_network <-  
  set.arc(gs_network,  
    "abnormal_short_term_variability_group",  
    "baseline_value_group")  
  
bn_igraph <- as.igraph(gs_network)  
plot(  
  bn_igraph,  
  vertex.size = 25,  
  vertex.label.cex = 0.6,  
  edge.arrow.size = 0.2  
)
```



Wybór sieci

Według wyników, grafy, które najlepiej obrazują zależności opierają się na sieci HC oraz TABU. Do dalszej pracy wybrałem sieć HC.

```
# list of network names
networks <- c(
  "hc_network",
  "iamb_network",
  "fast_iamb_network",
  "tabu_network",
  "pc_network",
  "gs_network"
)

# a data frame to store the scores
network_scores <- as.data.frame(matrix(nrow = length(networks), ncol = 2))
colnames(network_scores) <- c("Network", "Score")

# a loop to compute the scores for each network
for (i in 1:length(networks)) {
  network_name <- networks[i]
  network_object <- get(network_name) # Get the actual network object
  network_scores[i, 1] <- network_name
  network_scores[i, 2] <- score(network_object, data = fetal_health_grouped, type = "bic")
}

cat("Sieci z największym wynikiem to :\n")
eci z największym wynikiem to :
network_scores[network_scores$Score == max(network_scores$Score),]
  Network      Score
  hc_network -20285.98
  tabu_network -20285.98
```

	Network	Score
1	hc_network	-20285.98
2	iamb_network	-26495.67
3	fast_iamb_network	-26287.41
4	tabu_network	-20285.98
5	pc_network	-25400.75
6	gs_network	-26189.10

Estymacja parametrów

```
> estimation <- bn.fit(hc_network, fetal_health_grouped)
> estimation

  Bayesian network parameters

  Parameters of node baseline_value_group (multinomial distribution)

  Conditional probability table:

            abnormal_short_term_variability_group
baseline_value_group [11,26.8] (26.8,41.6] (41.6,56.4] (56.4,71.2] (71.2,87]
[109,120] 0.210884354 0.028056112 0.011061947 0.025291829 0.000000000
(120,130] 0.520408163 0.378757515 0.223451327 0.192607004 0.155339806
(130,139] 0.238095238 0.368737475 0.422566372 0.387159533 0.349514563
(139,149] 0.030612245 0.218436874 0.283185841 0.322957198 0.388349515
(149,160] 0.000000000 0.006012024 0.059734513 0.071984436 0.106796117

  Parameters of node fetal_movement_group (multinomial distribution)

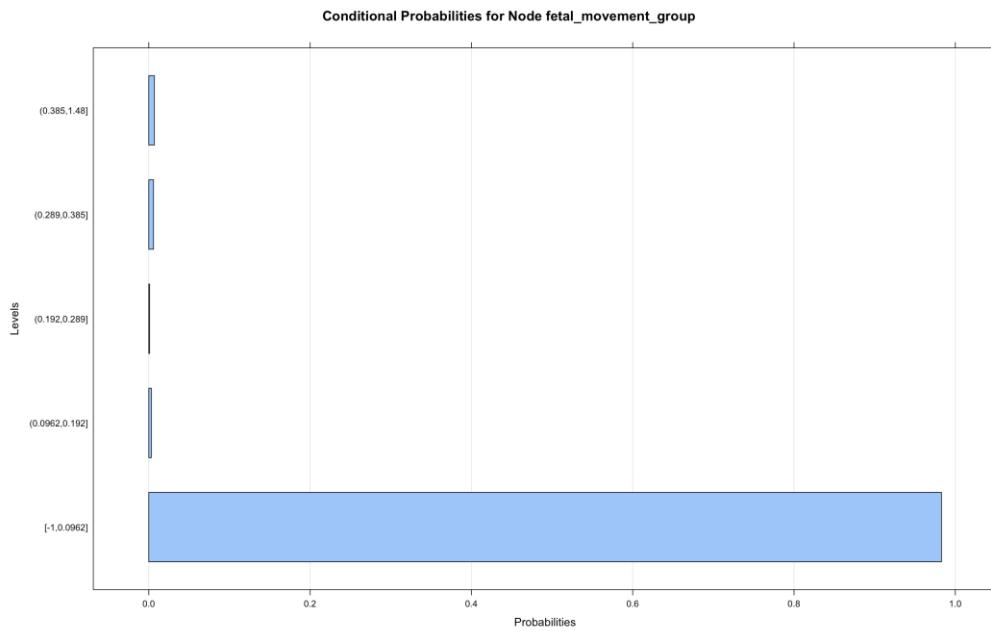
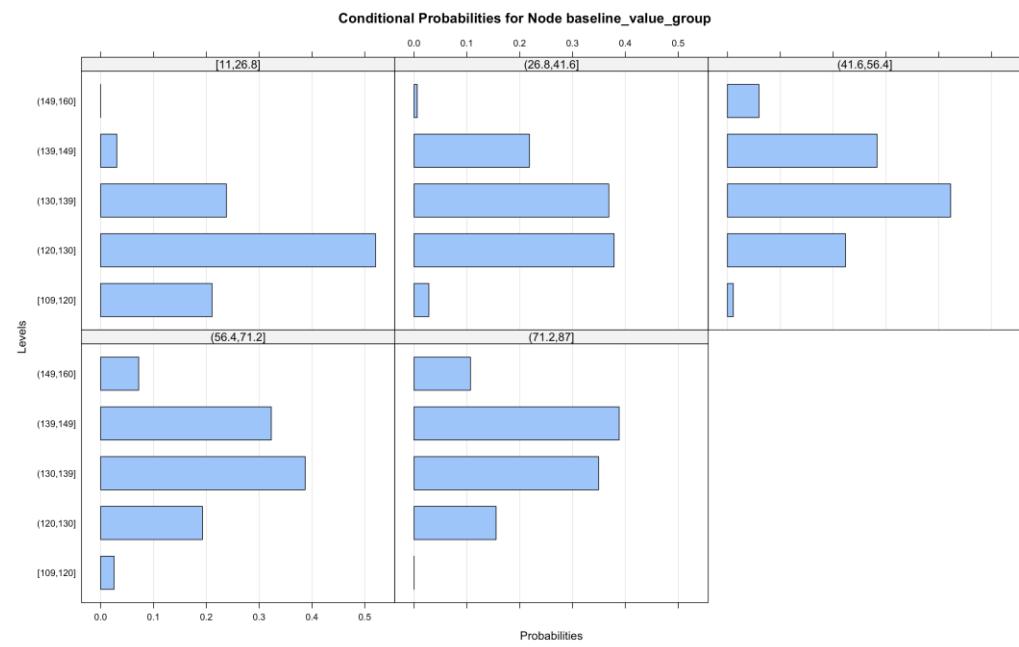
  Conditional probability table:
  [-1,0.0962] (0.0962,0.192] (0.192,0.289] (0.289,0.385] (0.385,1.48]
  0.982814178 0.003222342 0.001074114 0.005907626 0.006981740

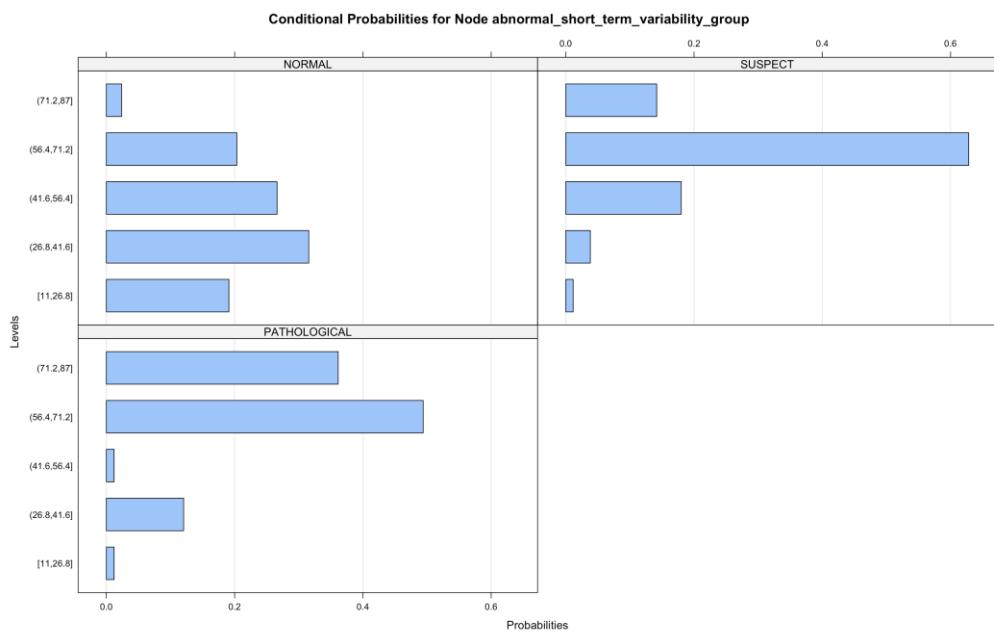
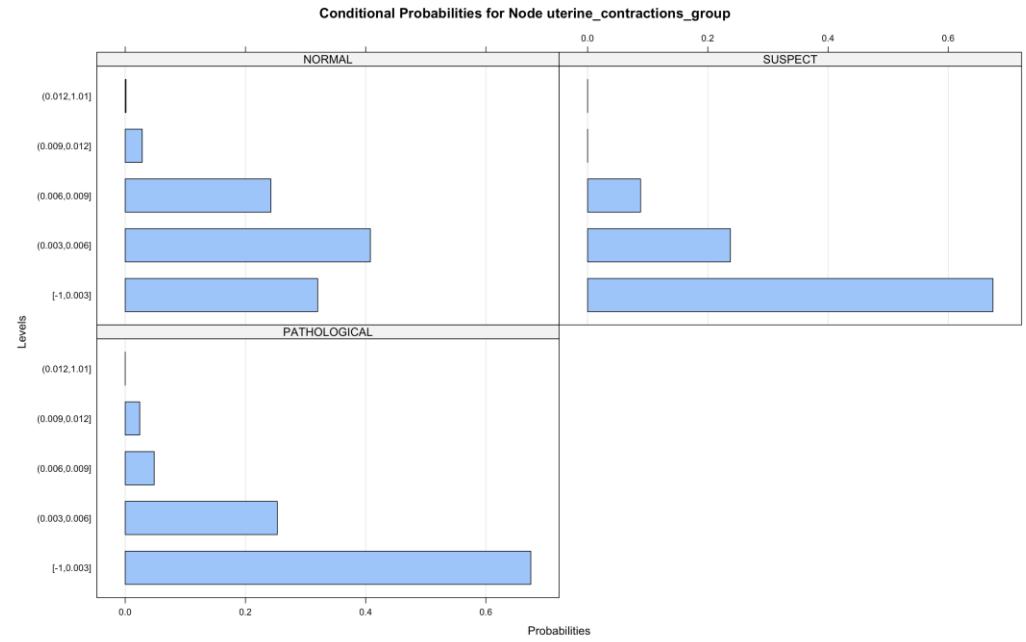
  Parameters of node uterine_contractions_group (multinomial distribution)

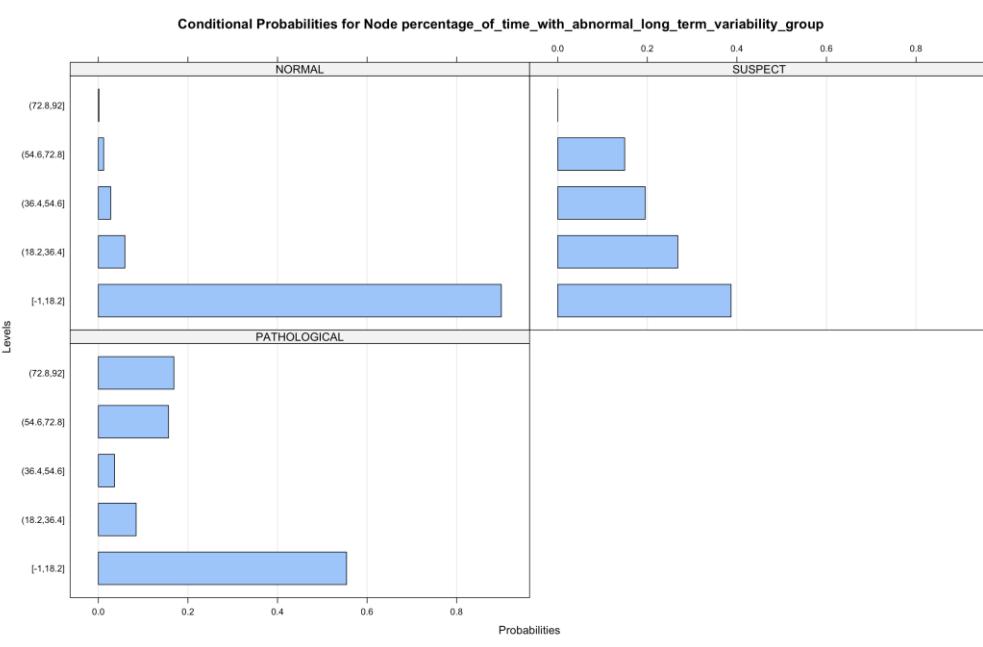
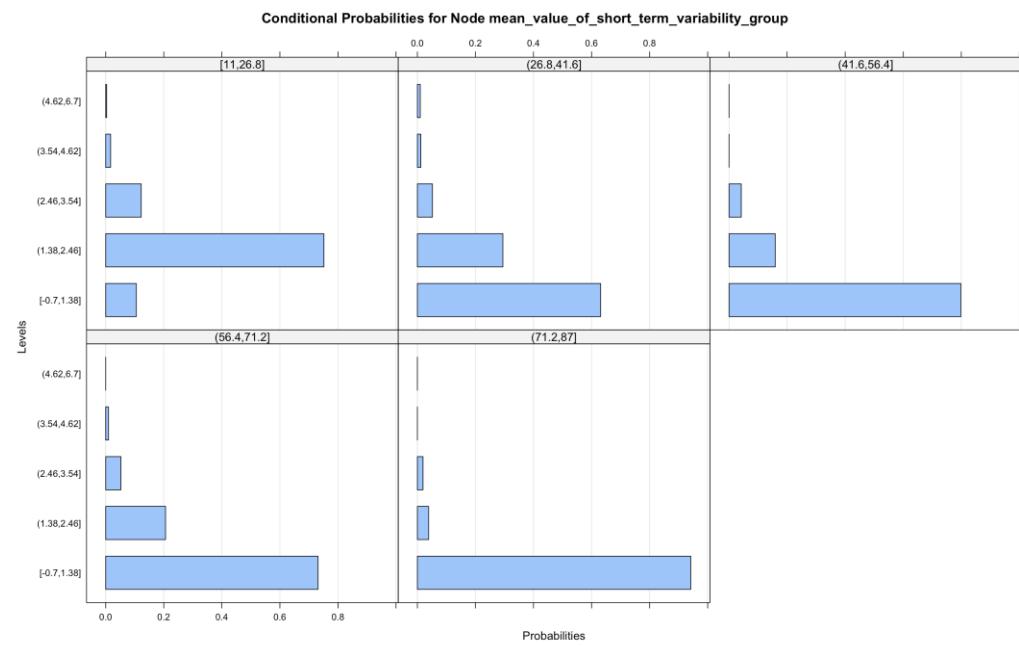
  Conditional probability table:

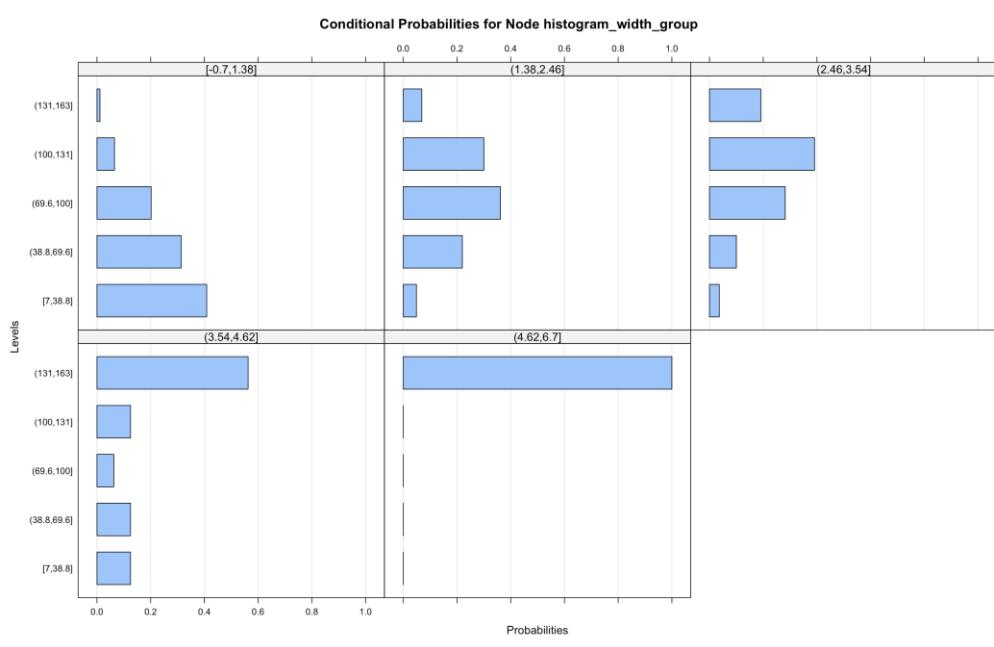
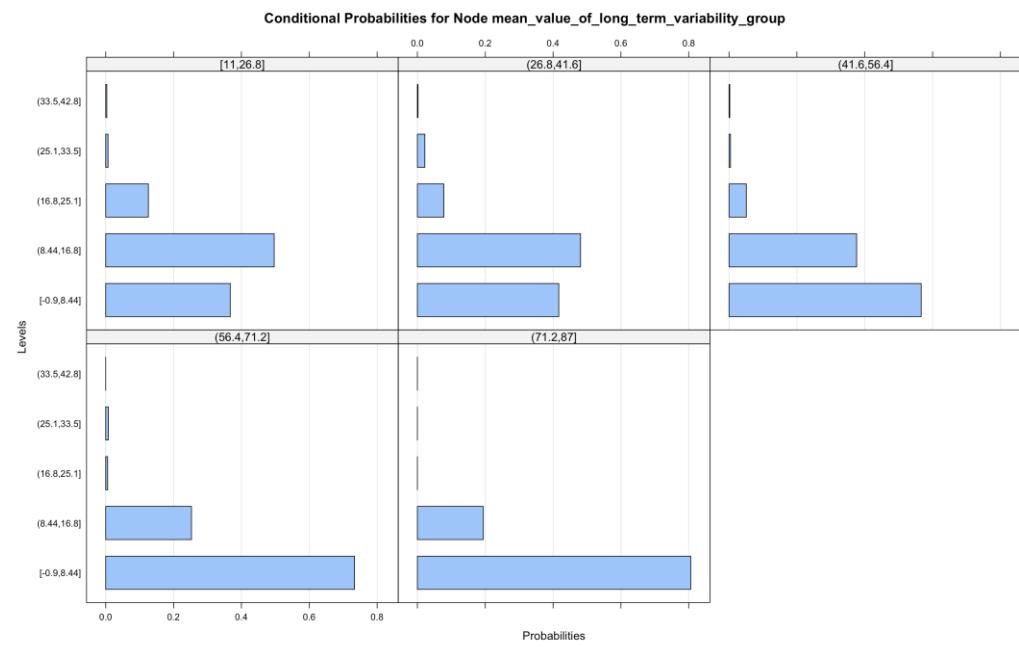
            fetal_health
uterine_contractions_group NORMAL SUSPECT PATHOLOGICAL
[-1,0.003] 0.320158103 0.674329502 0.674698795
(0.003,0.006] 0.407773386 0.237547893 0.253012048
(0.006,0.009] 0.242424242 0.088122605 0.048192771
```

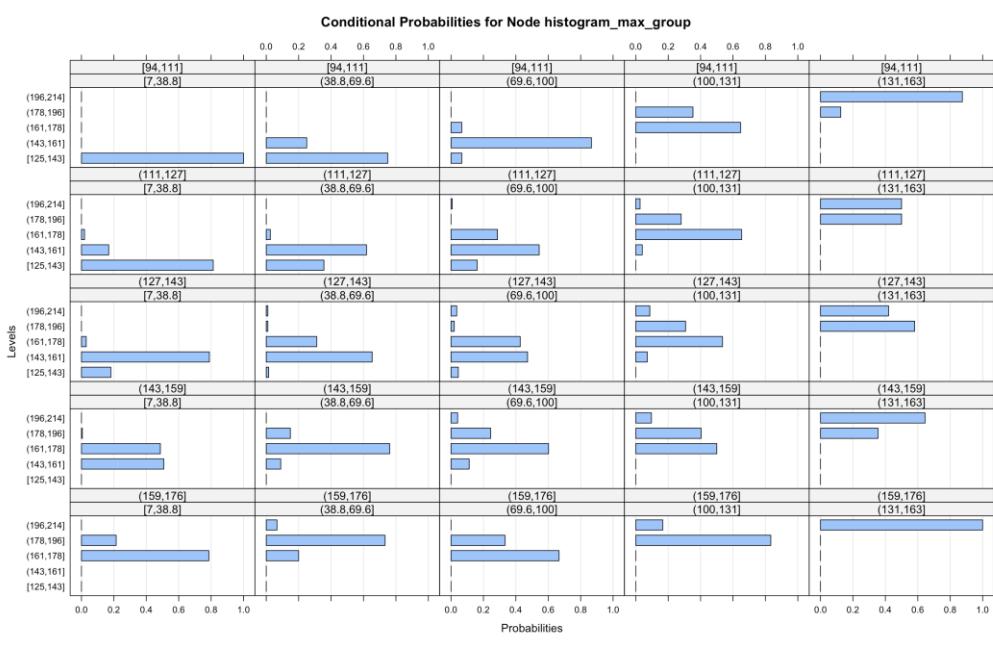
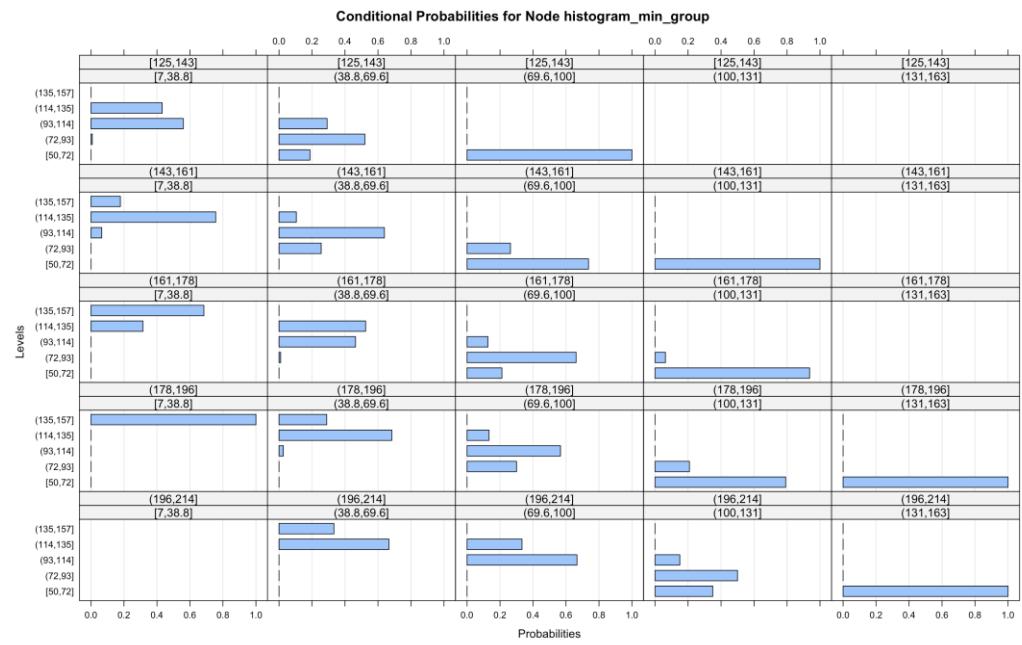
Rozkłady Prawdopodobieństw

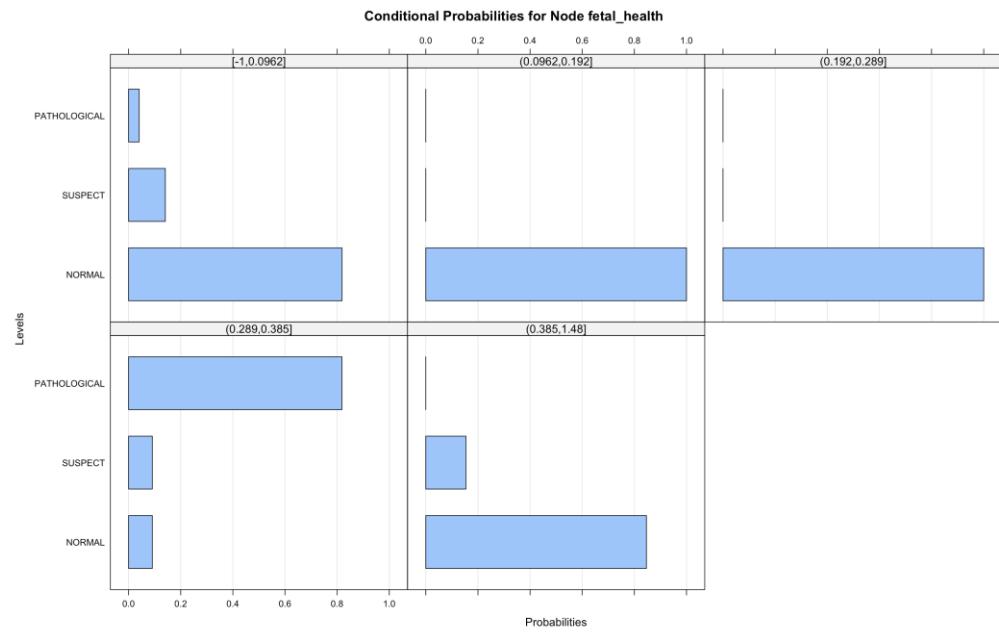
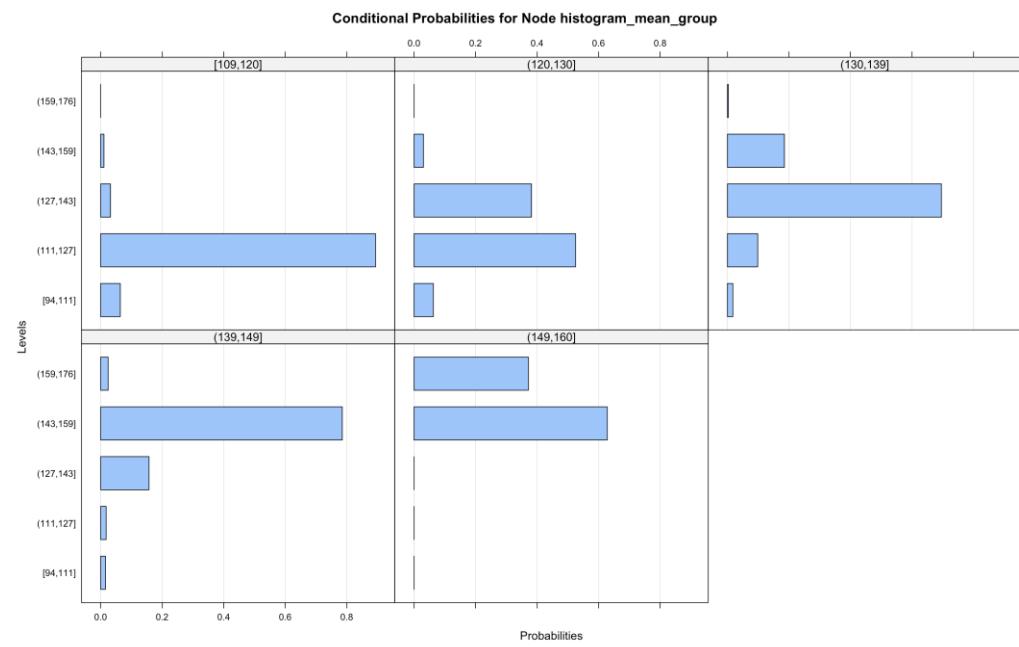












Przykładowe Prawdopodobieństwa

Prawdopodobieństwo 1

$$P(\text{baseline_value_group} = (130, 139] | \text{fetal_health} = \text{SUSPECT})$$

```
#prob 1
library(gRain)
grain <- compile(as.grain(estimation))

prob1 <- setEvidence(grain,
                      nodes = c("baseline_value_group"),
                      states = c("(130,139]"))
result_prob1 <- querygrain(prob1, nodes = c("fetal_health"))$fetal_health
result_prob1
```

```
> result_prob1
fetal_health
  NORMAL      SUSPECT    PATHOLOGICAL
  0.80677133  0.14807130  0.04515737
> |
```

Prawdopodobieństwo 1 „na papierze”

$$\begin{aligned} P(\text{baseline_value_group} = (130, 139] | \text{fetal_health} = \text{SUSPECT}) &= \\ &= \frac{P(\text{fetal_health} = \text{SUSPECT} | \text{baseline_value_group} = (130, 139]) \cdot P(\text{baseline_value_group})}{P(\text{fetal_health} = \text{SUSPECT})} \\ P(\text{baseline_value_group} = (130, 139]) &= \\ +P(\text{baseline_value_group} = (130, 139] | \text{fetal_health} = \text{NORMAL}) \cdot P(\text{fetal_health} = \text{NORMAL}) & \\ +P(\text{baseline_value_group} = (130, 139] | \text{fetal_health} = \text{SUSPECT}) \cdot P(\text{fetal_health} = \text{SUSPECT}) & \\ +P(\text{baseline_value_group} = (130, 139] | \text{fetal_health} = \text{PATHOLOGICAL}) \cdot P(\text{fetal_health} = \text{PATHOLOGICAL}) & \end{aligned}$$

$$\begin{aligned} P(\text{fetal_health} = \text{SUSPECT}) &= \\ P(\text{fetal_health} = \text{SUSPECT} | \text{baseline_value_group} = [109, 120]) * P(\text{baseline_value_group} = [109, 120]) & \\ +P(\text{fetal_health} = \text{SUSPECT} | \text{baseline_value_group} = (120, 130]) * P(\text{baseline_value_group} = (120, 130]) & \\ +P(\text{fetal_health} = \text{SUSPECT} | \text{baseline_value_group} = (130, 139]) * P(\text{baseline_value_group} = (130, 139]) & \\ +P(\text{fetal_health} = \text{SUSPECT} | \text{baseline_value_group} = (139, 149]) * P(\text{baseline_value_group} = (139, 149]) & \\ +P(\text{fetal_health} = \text{SUSPECT} | \text{baseline_value_group} = (149, 160]) * P(\text{baseline_value_group} = (149, 160]) & \end{aligned}$$

Prawdopodobieństwo 2

$$P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = (1.38, 2.46])$$

```
> #prob 2
>
>
> prob2 <- setEvidence(grain,
+                         nodes = c("histogram_width_group"),
+                         states = c("(100,131]"))
> result_prob2 <- querygrain(prob2, nodes = c("mean_value_of_short_term_variability_group"))$mean_value_of_short_term_variability_group
> result_prob2
mean_value_of_short_term_variability_group
[-0.7,1.38] [1.38,2.46] [2.46,3.54] [3.54,4.62] [4.62,6.7]
0.26829268 0.574912892 0.149825784 0.006968641 0.000000000
```

$$P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = (1.38, 2.46]) =$$

$$\frac{P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46] \mid \text{histogram_width_group} = (100, 131]) \cdot P(\text{histogram_width_group} = (100, 131])}{P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46])}$$

$$P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46]) =$$

$$+P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46] \mid \text{histogram_width_group} = [7, 38.8]) \cdot P(\text{histogram_width_group} = [7, 38.8])$$

$$+P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46] \mid \text{histogram_width_group} = (38.8, 69.6]) \cdot P(\text{histogram_width_group} = (38.8, 69.6])$$

$$+P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46] \mid \text{histogram_width_group} = (69.6, 100]) \cdot P(\text{histogram_width_group} = (69.6, 100])$$

$$+P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46] \mid \text{histogram_width_group} = (100, 131]) \cdot P(\text{histogram_width_group} = (100, 131])$$

$$+P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46] \mid \text{histogram_width_group} = (131, 163]) \cdot P(\text{histogram_width_group} = (131, 163))$$

$$P(\text{histogram_width_group} = (100, 131]) =$$

$$P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = [-0.7, 1.38]) \cdot P(\text{mean_value_of_short_term_variability_group} = [-0.7, 1.38])$$

$$+P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = (1.38, 2.46]) \cdot P(\text{mean_value_of_short_term_variability_group} = (1.38, 2.46])$$

$$+P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = (2.46, 3.54]) \cdot P(\text{mean_value_of_short_term_variability_group} = (2.46, 3.54])$$

$$+P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = (3.54, 4.62]) \cdot P(\text{mean_value_of_short_term_variability_group} = (3.54, 4.62])$$

$$+P(\text{histogram_width_group} = (100, 131] \mid \text{mean_value_of_short_term_variability_group} = (4.62, 6.7]) \cdot P(\text{mean_value_of_short_term_variability_group} = (4.62, 6.7])$$

Prawdopodobieństwo 2 „na papierze”

Prawdopodobieństwo 3

$$P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = [7, 38.8])$$

```
> # prob 3
>
> prob3 <- setEvidence(grain,
+                         nodes = c("fetal_movement_group"),
+                         states = c("[-1,0.0962]"))
> result_prob3 <- querygrain(prob3, nodes = c("histogram_width_group"))$histogram_width_group
> result_prob3
histogram_width_group
[7,38.8] (38.8,69.6] (69.6,100] (100,131] (131,163]
0.27635413 0.27060548 0.25197056 0.15428811 0.04678173
```

Prawdopodobieństwo 3 „na papierze”

$$\begin{aligned} & P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = [7, 38.8]) = \\ &= \frac{P(\text{histogram_width_group} = [7, 38.8] \mid \text{fetal_movement_group} = [-1, 0.0962]) \cdot P(\text{fetal_movement_group} = [-1, 0.0962])}{P(\text{histogram_width_group} = [7, 38.8])} \\ & P(\text{histogram_width_group} = [7, 38.8]) = \\ & + P(\text{histogram_width_group} = [7, 38.8] \mid \text{fetal_movement_group} = [-1, 0.0962]) \cdot P(\text{fetal_movement_group} = [-1, 0.0962]) \\ & + P(\text{histogram_width_group} = [7, 38.8] \mid \text{fetal_movement_group} = (0.0962, 0.192)) \cdot P(\text{fetal_movement_group} = (0.0962, 0.192)) \\ & + P(\text{histogram_width_group} = [7, 38.8] \mid \text{fetal_movement_group} = (0.192, 0.289)) \cdot P(\text{fetal_movement_group} = (0.192, 0.289)) \\ & + P(\text{histogram_width_group} = [7, 38.8] \mid \text{fetal_movement_group} = (0.289, 0.385)) \cdot P(\text{fetal_movement_group} = (0.289, 0.385)) \\ & + P(\text{histogram_width_group} = [7, 38.8] \mid \text{fetal_movement_group} = (0.385, 1.48)) \cdot P(\text{fetal_movement_group} = (0.385, 1.48)) \end{aligned}$$

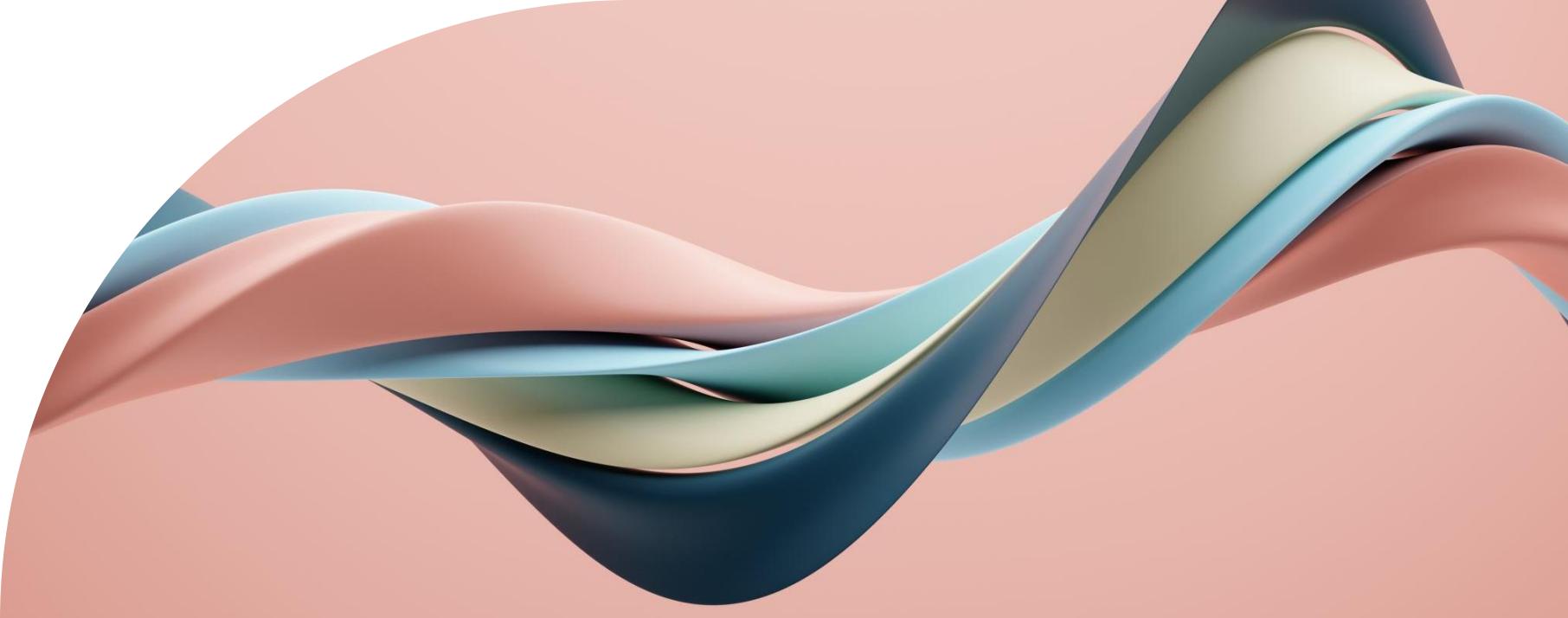
$$\begin{aligned} & P(\text{fetal_movement_group} = [-1, 0.0962]) = \\ & P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = [7, 38.8]) \cdot P(\text{histogram_width_group} = [7, 38.8]) \\ & + P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = (38.8, 69.6)) \cdot P(\text{histogram_width_group} = (38.8, 69.6)) \\ & + P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = (69.6, 100)) \cdot P(\text{histogram_width_group} = (69.6, 100)) \\ & + P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = (100, 131)) \cdot P(\text{histogram_width_group} = (100, 131)) \\ & + P(\text{fetal_movement_group} = [-1, 0.0962] \mid \text{histogram_width_group} = (131, 163)) \cdot P(\text{histogram_width_group} = (131, 163)) \end{aligned}$$

Wnioski

Sieci Bayesa pozwalają na zbadanie połączeń między danymi oraz ujrzanie połączeń niewidocznych na pierwszy rzut oka.

Projekt umożliwia zbadanie tych połączeń oraz prawdopodobieństw wystąpienia danej wartości pod warunkiem wystąpienia innej.

Niestety nie wszystkie automatycznie wygenerowane sieci są pełne i funkcjonalne, lecz można uzupełnić je korzystając z macierzy niezależności.



Dziękuję za uwagę.

Mateusz Pindyk

169831