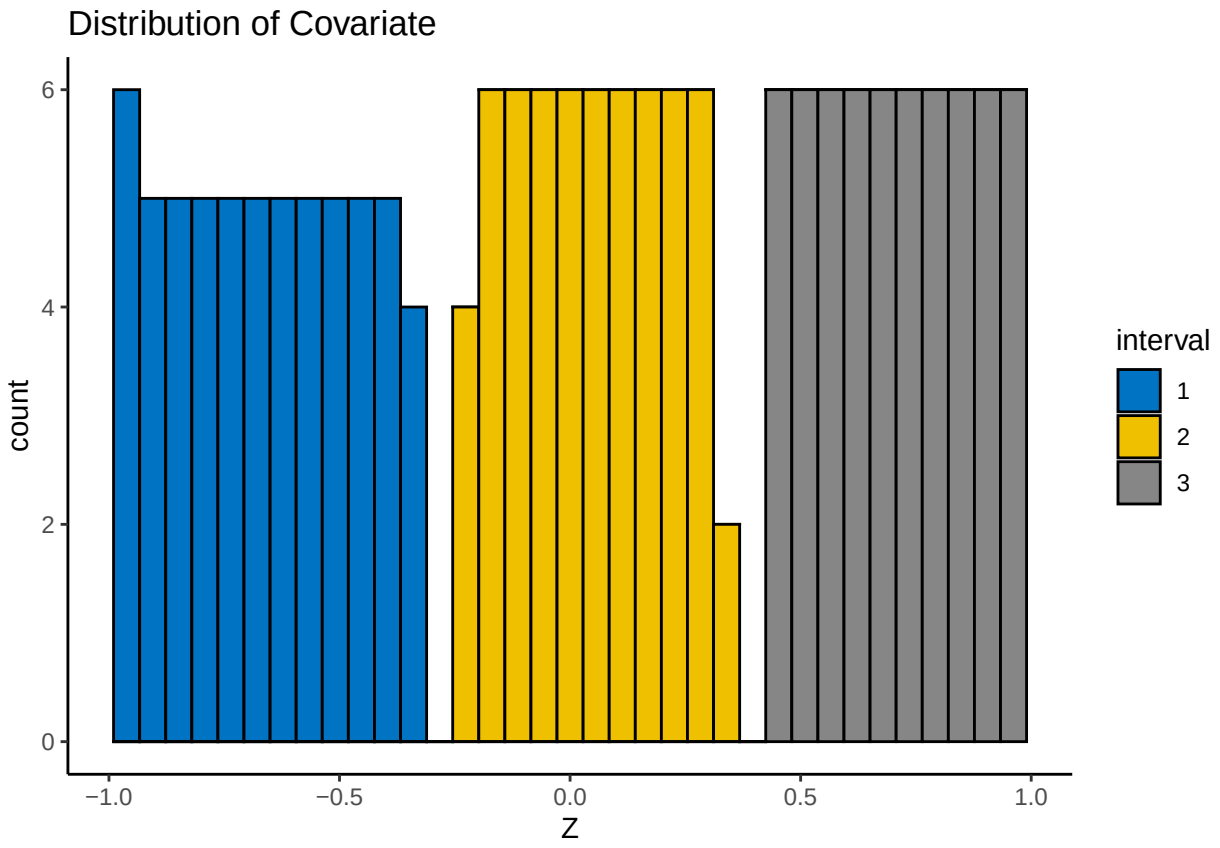
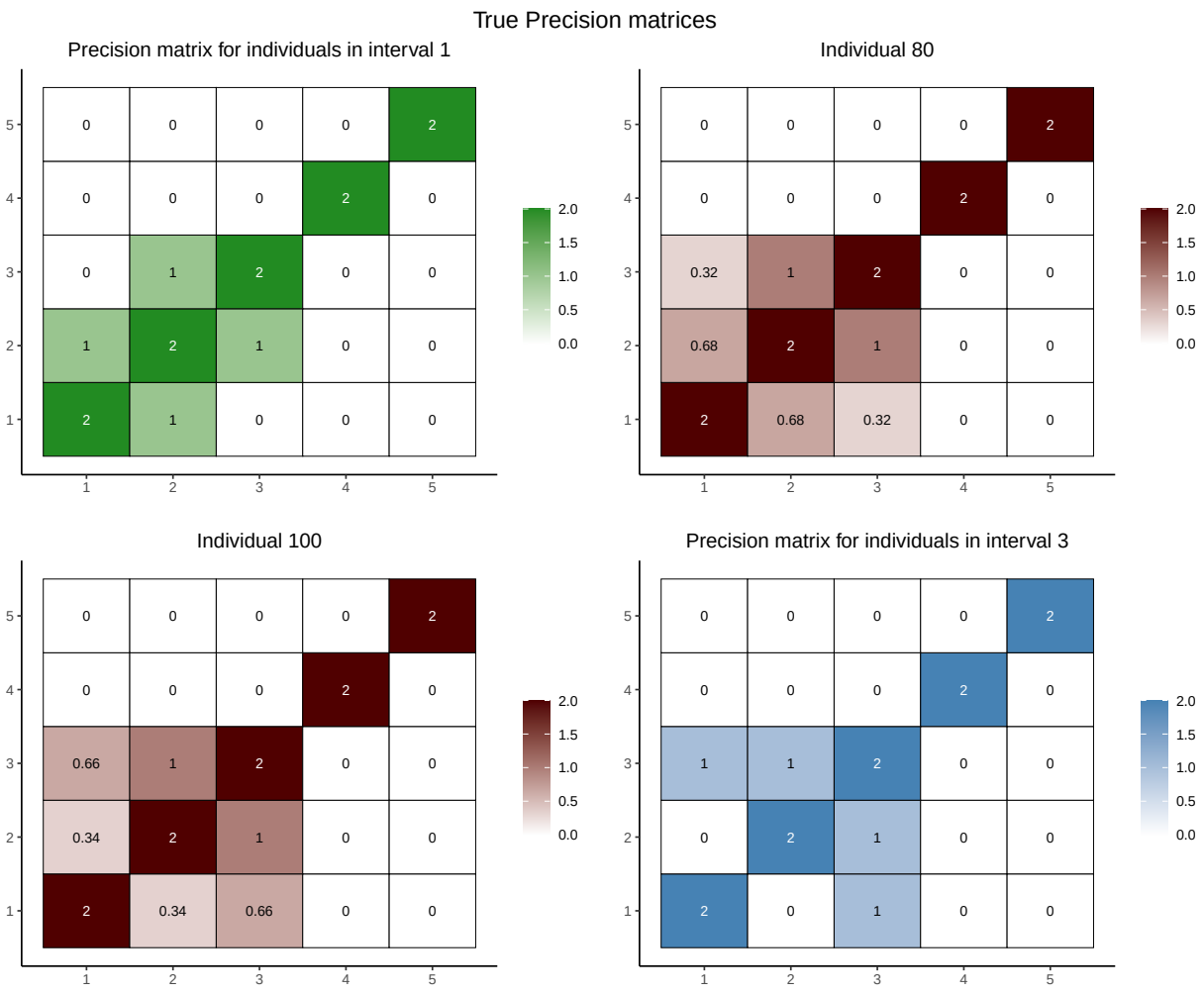
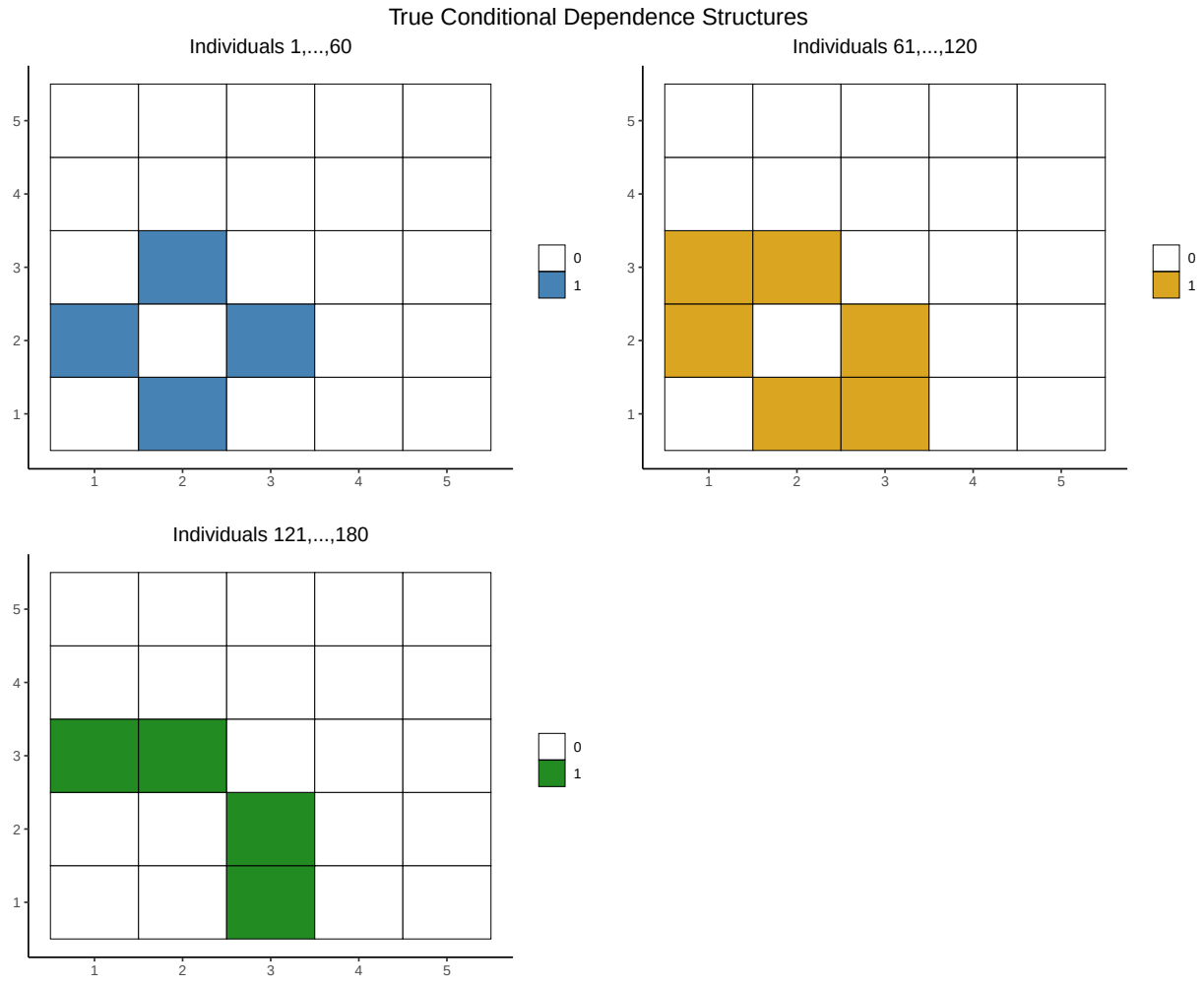
parallelization-demo

Data generation



Interval	Individual Indices
1	$1, \ldots, 60$
2	$61, \ldots, 120$
3	$121, \ldots, 180$


True Precision matrices
Precision matrix for individuals in interval 1
0 0 0 0 2
0 0 0 2 0
0 1 2 0 0
1 2 1 0 0
2 1 0 0 0
Individual 80
0 0 0 0 2
0 0 0 2 0
0.32 1 2 0 0
0.68 2 1 0 0
2 0.68 0.32 0 0
Individual 100
0 0 0 0 2
0 0 0 2 0
0.66 1 2 0 0
0.34 2 1 0 0
2 0.34 0.66 0 0
Precision matrix for individuals in interval 3
0 0 0 0 2
0 0 0 2 0
1 1 2 0 0
0 2 1 0 0
2 0 1 0 0
1 2 3 4 5
2.0 1.5 1.0 0.5 0.0



Parallel CAVI

Setting `parallel = T` in a call to `covdepGE` performs the CAVI for each variable in parallel. Parallel backend may be registered manually by the user, but will otherwise be done automatically. This allows flexibility for the user to configure the parallelization according to their needs.

Manual parallel backend registration:

```
# record time to register parallel backend
start <- Sys.time()
doParallel::registerDoParallel(5)
Sys.time() - start
```

```
## Time difference of 0.4940879 secs
```

```
# run covdepGE in parallel
covdepGE(data_mat, Z, parallel = T, n_param = 5)
```

```
## Detected 5 workers
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -80647.67              Unique conditional dependence structures: 4
## n: 180, variables: 5                        Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 2.02 secs
```

Automatic parallel backend registration

```
covdepGE(data_mat, Z, parallel = T, num_workers = 7, stop_cluster = F, n_param = 5)
```

```
## Warning in covdepGE(data_mat, Z, parallel = T, num_workers = 7, stop_cluster =
## F, : No registered workers detected; registering doParallel with 7 workers
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -80647.67              Unique conditional dependence structures: 4
## n: 180, variables: 5                        Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 2.426 secs
```

By setting `stop_cluster = F`, subsequent parallel calls to `covdepGE` are able to employ the same workers. This avoids the overhead of creating a new cluster.

Efficiency

Large hyperparameter grid

The model in the previous section was relatively simple, with only 5 grid points. In this case, the time to create the cluster and communication from the parent to the children workers outweighs the time savings of parallelizing the CAVI. Thus, sequential execution is faster for this small model.

```
covdepGE(data_mat, Z, n_param = 5)
```

```
##   |                                                                      |
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -80647.67              Unique conditional dependence structures: 4
## n: 180, variables: 5                        Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 2.571 secs
```

However, for a more complex model, the benefits of parallelization become apparent. To increase complexity, I will increase the number of grid points to 200.

```
# sequential
out_seq <- covdepGE(data_mat, Z, n_param = 200, CS = T)
```

```
##   |                                                                      |
```

```
## Warning in covdepGE(data_mat, Z, n_param = 200, CS = T): Variable 3: CAVI did
## not converge in 100 iterations for 1/200 grid search candidates
```

```
out_seq
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -80608.64               Unique conditional dependence structures: 5
## n: 180, variables: 5                       Hyperparameter grid size: 200 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 22.758 secs
```

```
# parallel
out_par <- covdepGE(data_mat, Z, n_param = 200, parallel = T,
                    num_workers = 6, CS = T)
```

```
## Detected 7 workers
```

```
## Warning in covdepGE(data_mat, Z, n_param = 200, parallel = T, num_workers =
## 6, : Variable 3: CAVI did not converge in 100 iterations for 1/200 grid search
## candidates
```

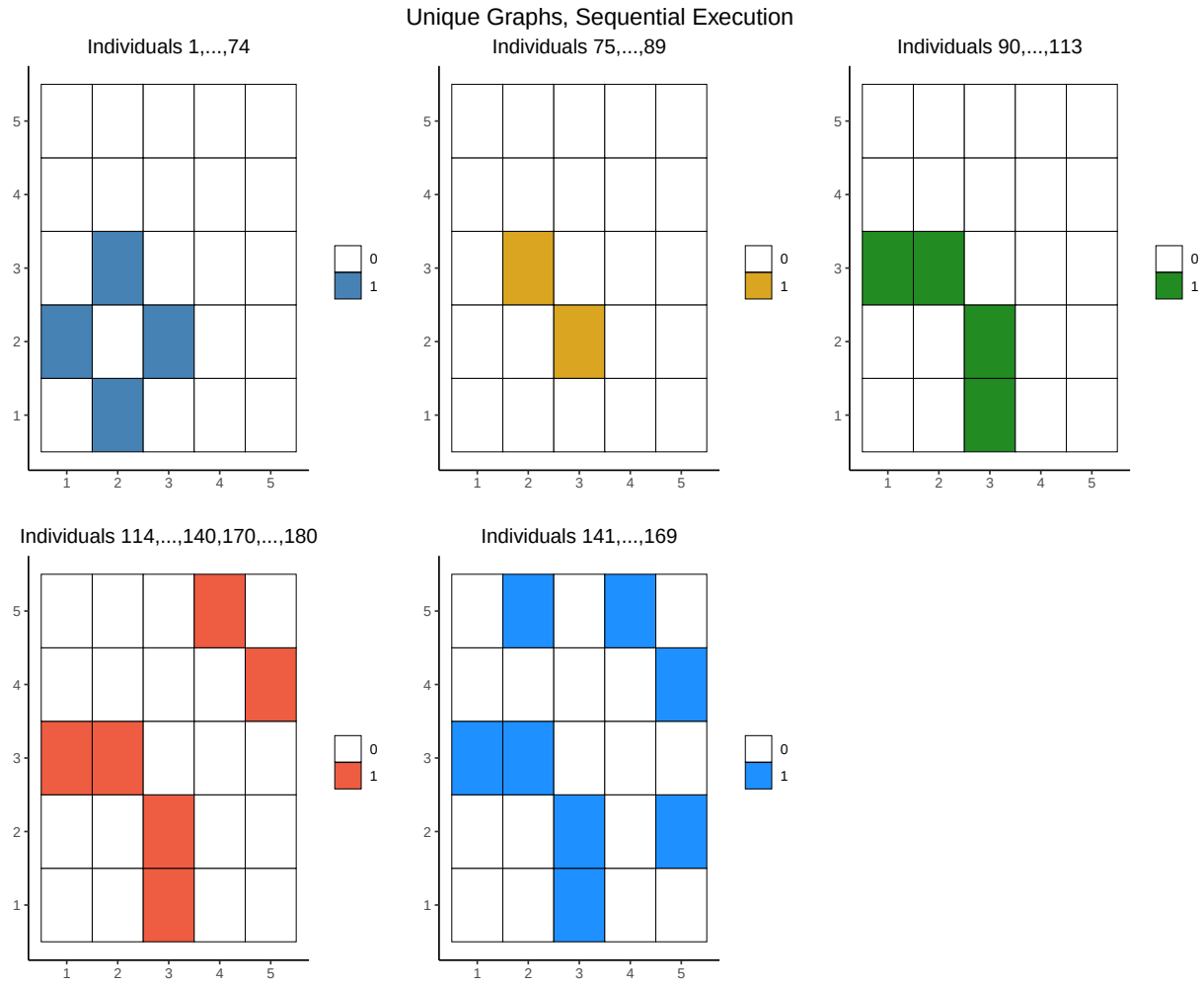
```
out_par
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -80608.64               Unique conditional dependence structures: 5
## n: 180, variables: 5                       Hyperparameter grid size: 200 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 7.277 secs
```

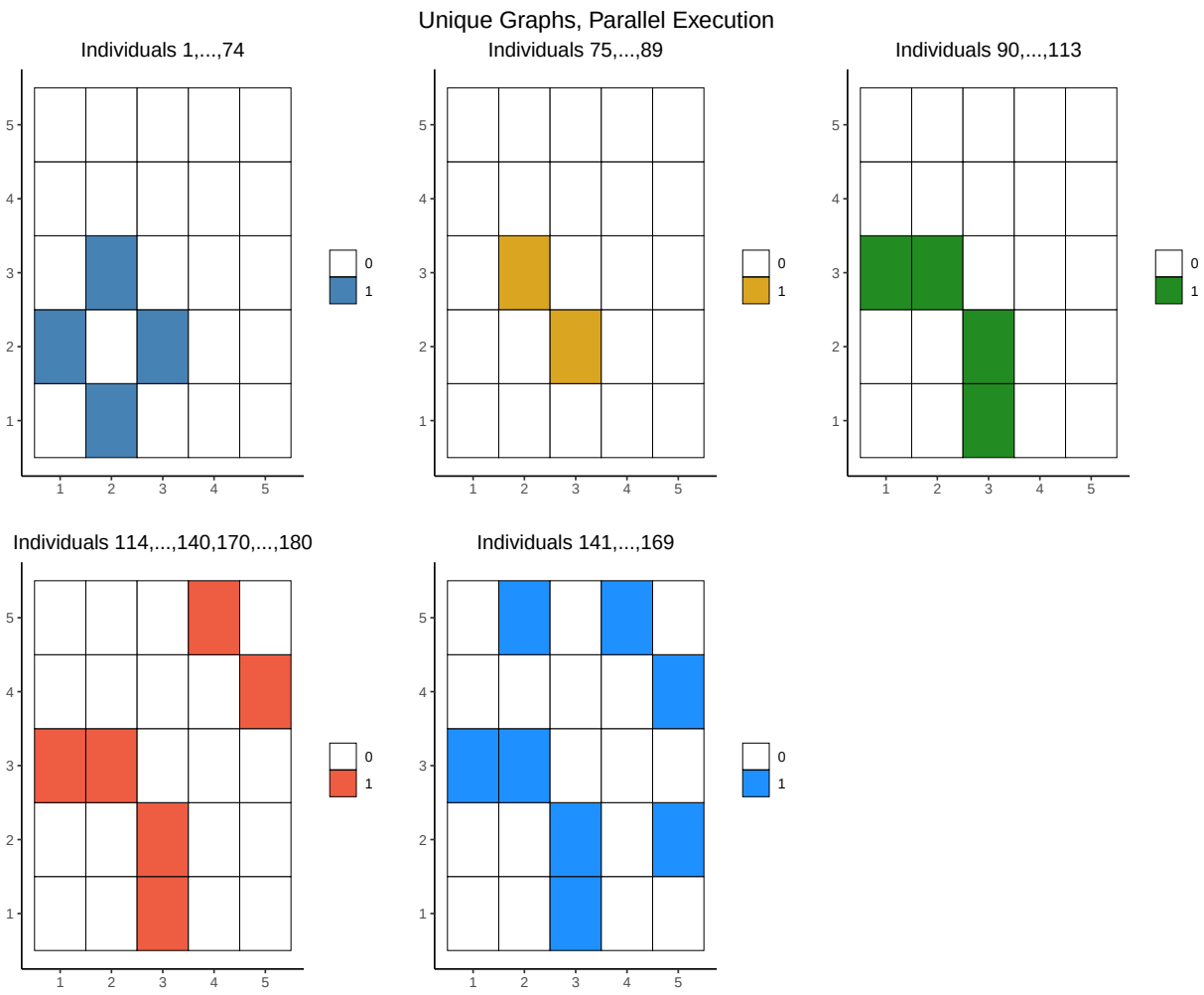
The parallel model outperforms the sequential - additionally, the models produce identical results.

Note the message displayed by the parallel model - it has detected that there are workers on an active cluster from the parallel model with `stop_cluster = F` above. It ignores the `num_workers` argument and re-uses the detected cluster.

```
annotate_figure(ggarrange(plotlist = plot(out_seq, graph_colors = colors)),
                top = text_grob("Unique Graphs, Sequential Execution", size = 15))
```

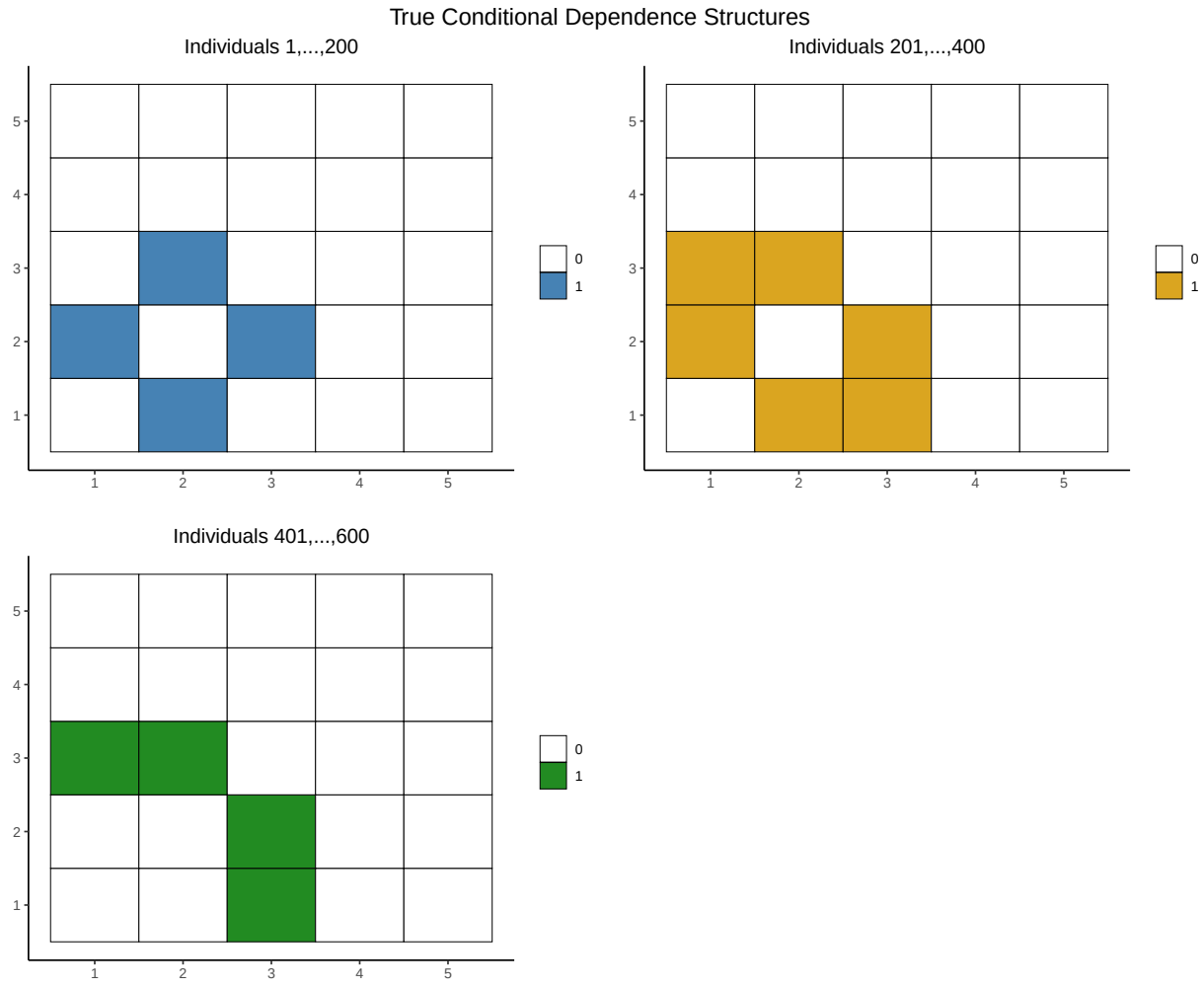



```
annotate_figure(ggarrange(plotlist = plot(out_par, colors)),
                top = text_grob("Unique Graphs, Parallel Execution", size = 15))
```



Large n

An increase in complexity can also be achieved by again choosing the number of grid points to be 5 and increasing the sample size. Again, the parallellized CAVI beats the sequential CAVI while producing the same results.


True Conditional Dependence Structures
Individuals 1,...,200
Individuals 201,...,400
Individuals 401,...,600
0
1
1
2
3
4
5

Note that since the last parallel call to `covdepGE` did not specify `stop_cluster = F`, the cluster must be re-created.

```
# sequential
out_seq <- covdepGE(data_mat, Z, n_param = 5)
```

```
##   |                                                                      |
```

```
## Warning in covdepGE(data_mat, Z, n_param = 5): Variable 1: CAVI did not converge
## in 100 iterations for 20/20 grid search candidates
```

```
## Warning in covdepGE(data_mat, Z, n_param = 5): Variable 2: CAVI did not converge
## in 100 iterations for 20/20 grid search candidates
```

```
out_seq
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -866063.03              Unique conditional dependence structures: 6
## n: 600, variables: 5                          Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 39.226 secs
```

```
# parallel
out_par <- covdepGE(data_mat, Z, n_param = 5, parallel = T, num_workers = 8)
```

```
## Warning in covdepGE(data_mat, Z, n_param = 5, parallel = T, num_workers = 8): No
## registered workers detected; registering doParallel with 8 workers
```

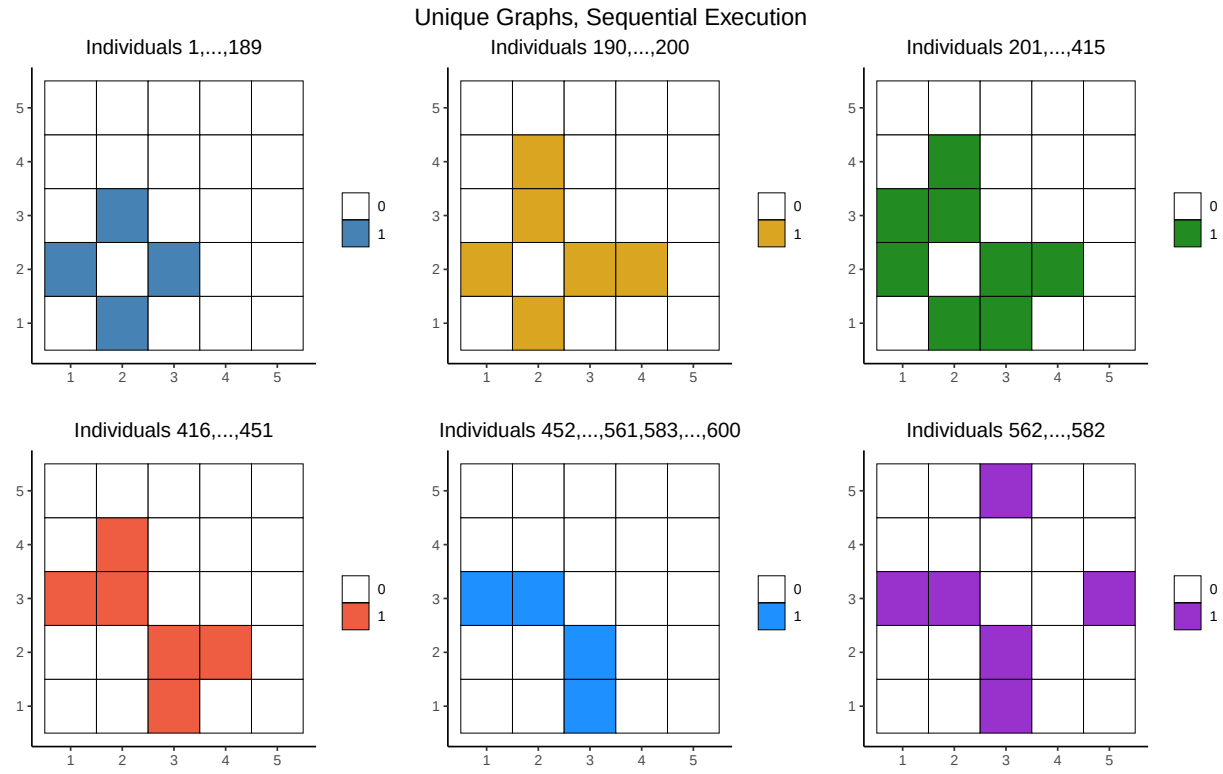
```
## Warning in covdepGE(data_mat, Z, n_param = 5, parallel = T, num_workers =
## 8): Variable 1: CAVI did not converge in 100 iterations for 20/20 grid search
## candidates
```

```
## Warning in covdepGE(data_mat, Z, n_param = 5, parallel = T, num_workers =
## 8): Variable 2: CAVI did not converge in 100 iterations for 20/20 grid search
## candidates
```

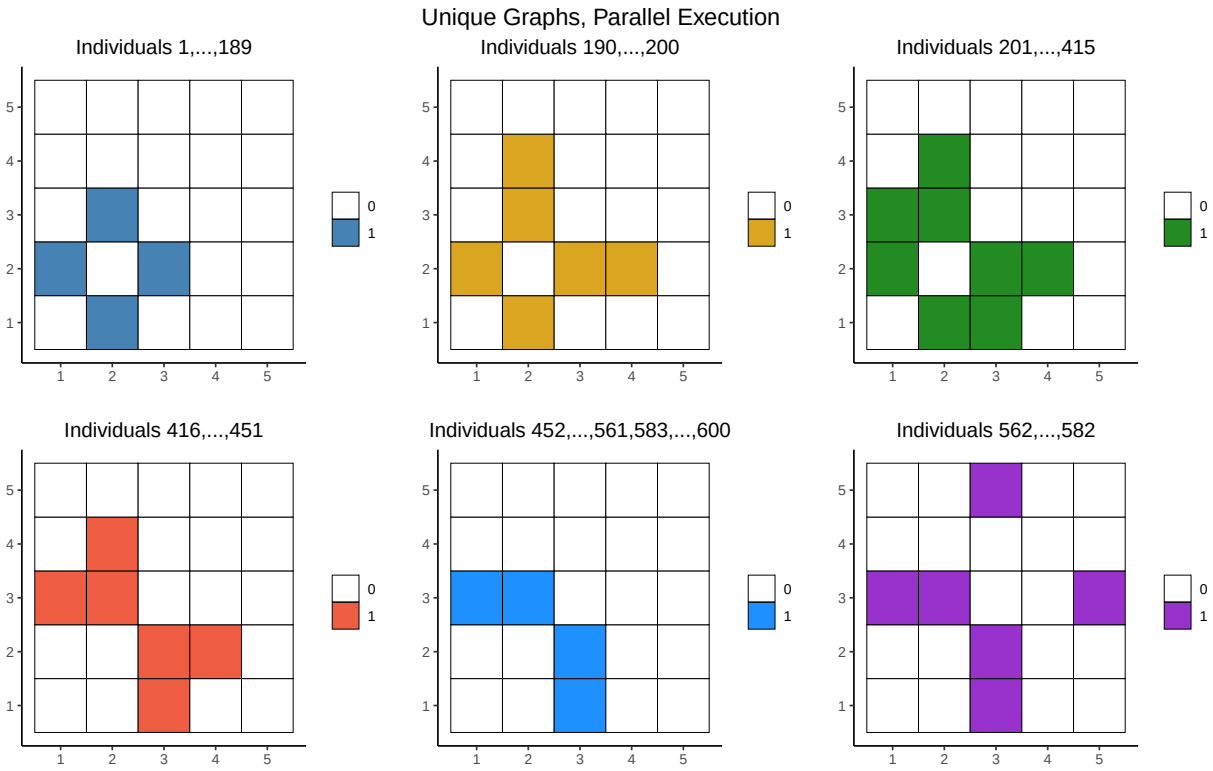
```
out_par
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -866063.03              Unique conditional dependence structures: 6
## n: 600, variables: 5                          Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 17.937 secs
```

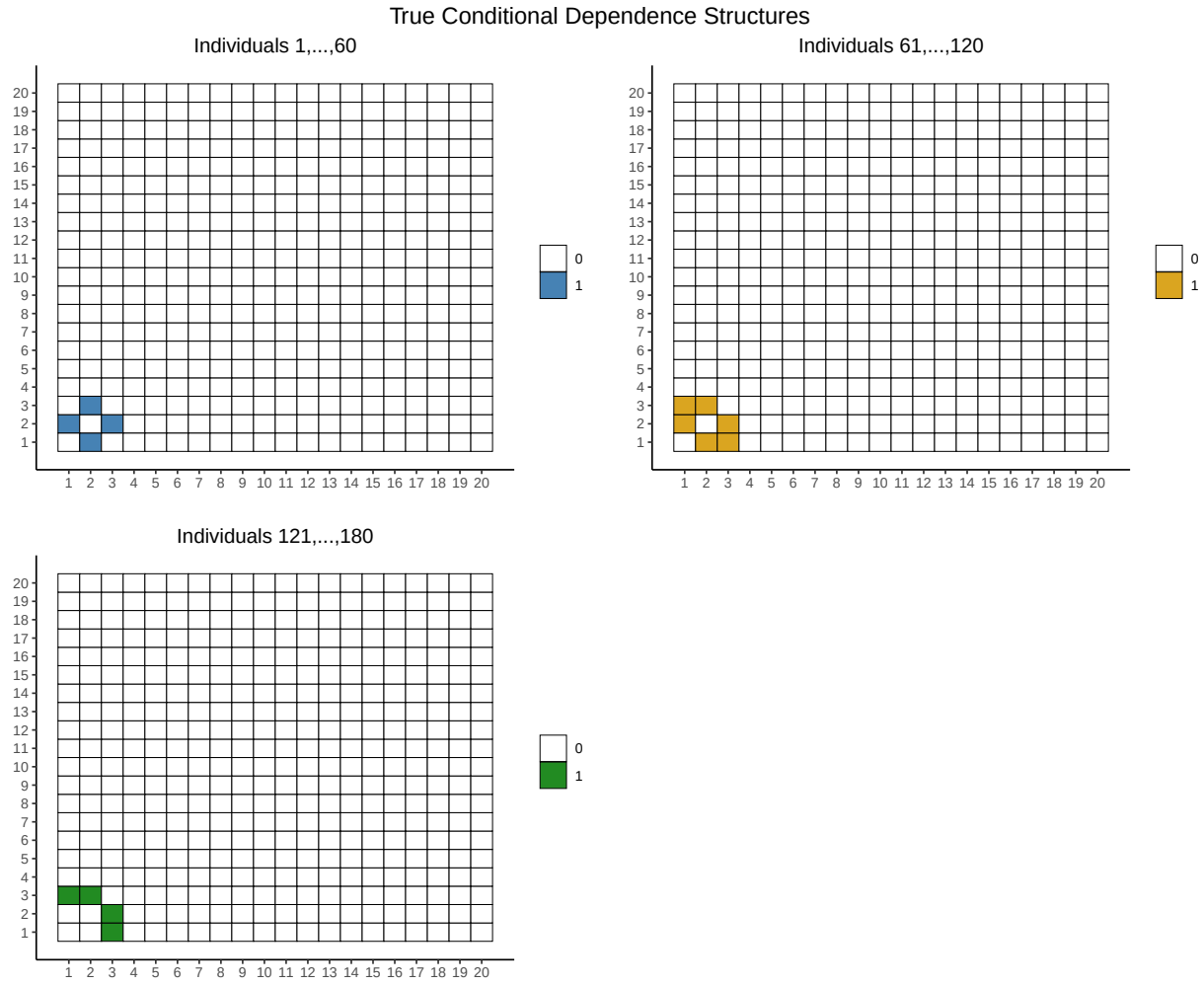
```
annotate_figure(ggarrange(plotlist = plot(out_seq, colors)),
                top = text_grob("Unique Graphs, Sequential Execution", size = 15))
```




```
annotate_figure(ggarrange(plotlist = plot(out_par, colors)),
                top = text_grob("Unique Graphs, Parallel Execution", size = 15))
```


Unique Graphs, Parallel Execution
Individuals 1,...,189
Individuals 190,...,200
Individuals 201,...,415
Individuals 416,...,451
Individuals 452,...,561,583,...,600
Individuals 562,...,582
0
1

Large p


True Conditional Dependence Structures
Individuals 1,...,60
Individuals 61,...,120
Individuals 121,...,180
0
1

```r
# sequential
out_seq <- covdepGE(data_mat, Z, n_param = 5, warnings = F)
```

```
##   |                                                                      |
```

```r
out_seq
```

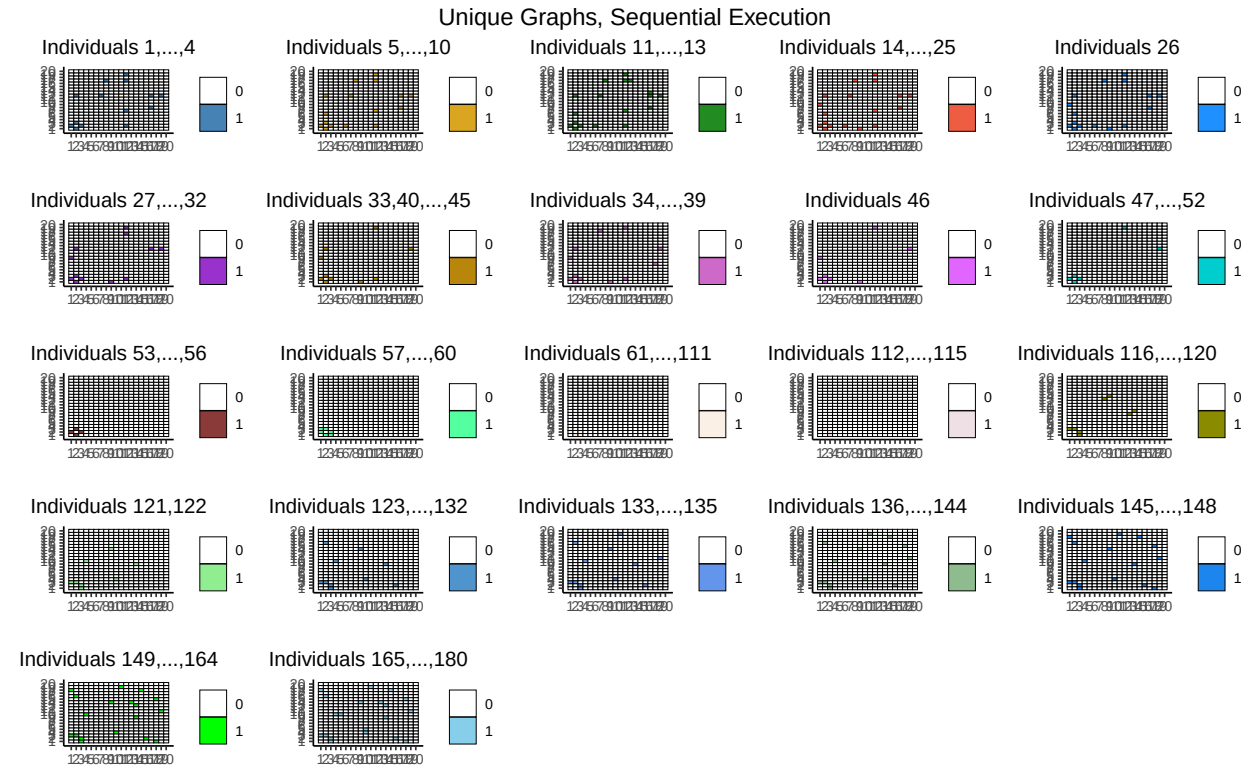
```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -329404.34             Unique conditional dependence structures: 22
## n: 180, variables: 20                        Hyperparameter grid size: 5 points
## CAVI converged for 19/20 variables
##
## Model fit completed in 1.116 mins
```

```r
# parallel
out_par <- covdepGE(data_mat, Z, n_param = 5, parallel = T, num_workers = 16, warnings = F)
out_par
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -329404.34             Unique conditional dependence structures: 22
## n: 180, variables: 20                        Hyperparameter grid size: 5 points
## CAVI converged for 19/20 variables
##
## Model fit completed in 16.996 secs
```

```r
set.seed(4)
colors <- c(colors, sample(colors()[sapply(colors(), function(color) !(substr(color, 1, 4)) %in% c("gre

annotate_figure(ggarrange(plotlist = plot(out_seq, colors)),
                top = text_grob("Unique Graphs, Sequential Execution", size = 15))
```




```
annotate_figure(ggarrange(plotlist = plot(out_par, colors)),
                top = text_grob("Unique Graphs, Parallel Execution", size = 15))
```