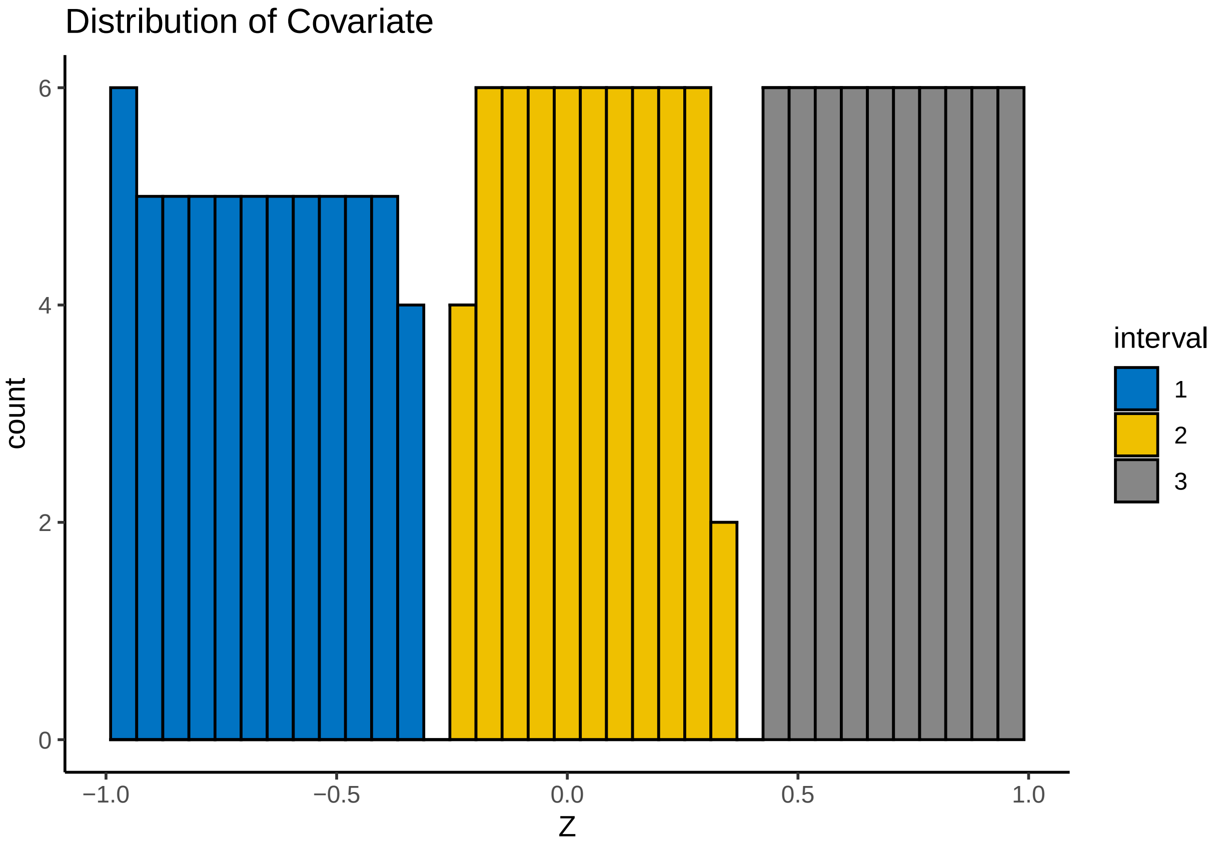
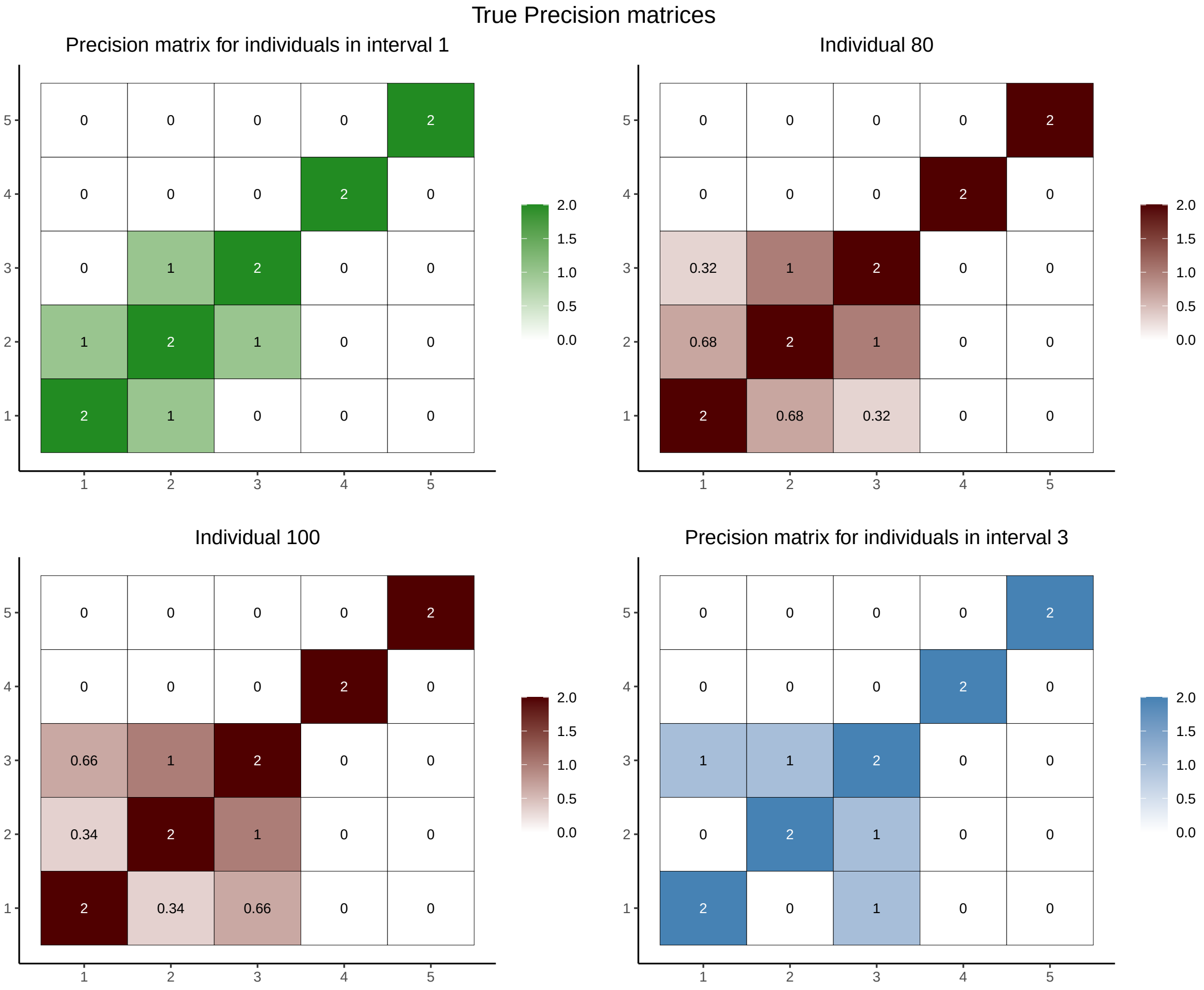
parallelization-demo

Data generation



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

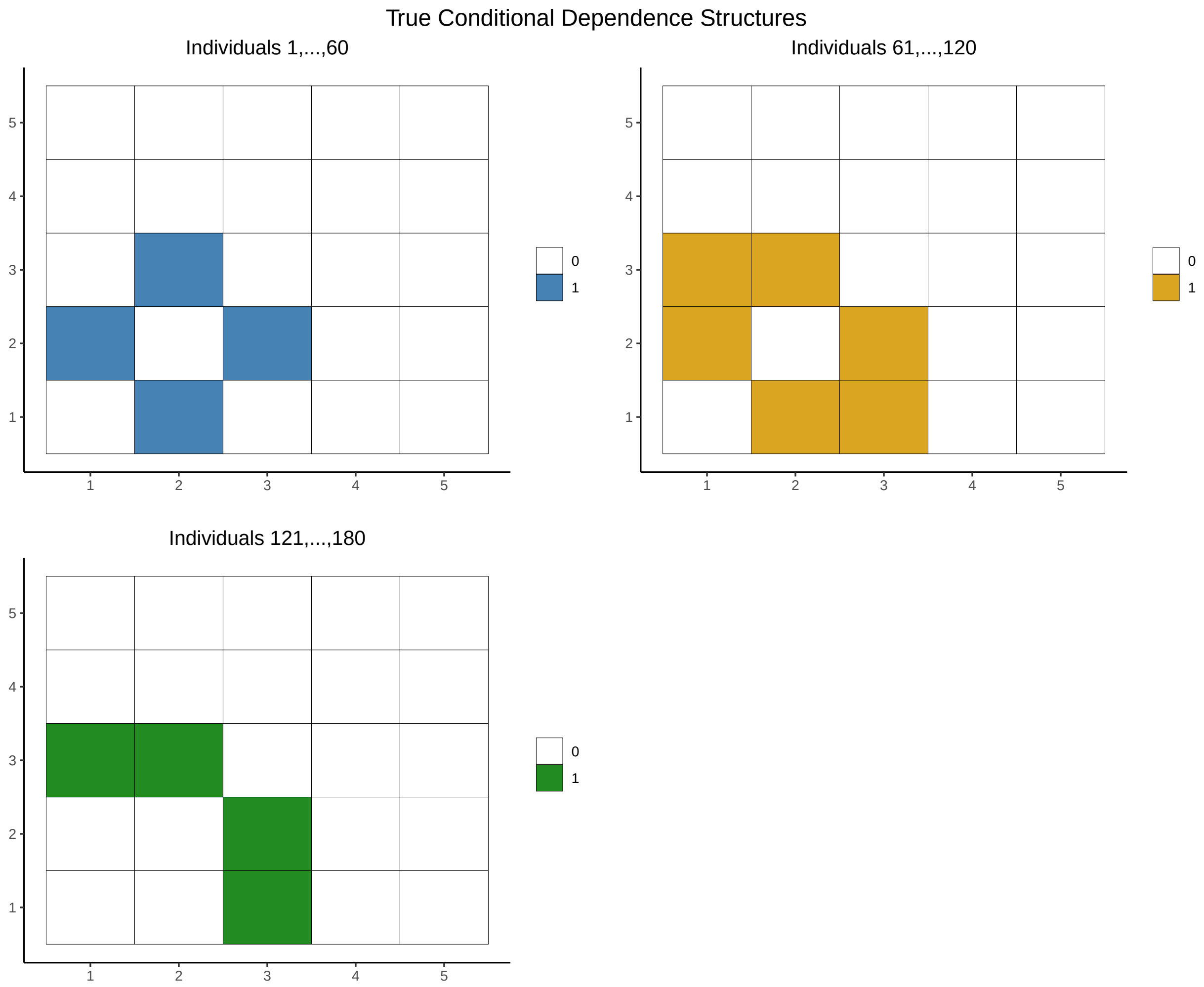


	1	2	3	4	5
5	0	0	0	0	2
4	0	0	0	2	0
3	0	1	2	0	0
2	1	2	1	0	0
1	2	1	0	0	0

	1	2	3	4	5
5	0	0	0	0	2
4	0	0	0	2	0
3	0.32	1	2	0	0
2	0.68	2	1	0	0
1	2	0.68	0.32	0	0

	1	2	3	4	5
5	0	0	0	0	2
4	0	0	0	2	0
3	0.66	1	2	0	0
2	0.34	2	1	0	0
1	2	0.34	0.66	0	0

	1	2	3	4	5
5	0	0	0	0	2
4	0	0	0	2	0
3	1	1	2	0	0
2	0	2	1	0	0
1	2	0	1	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 0.7301049 secs
```

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

```
## Detected 5 workers
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -90160.41               Unique conditional dependence structures: 4
## n: 180, variables: 5                         Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 0.752 secs
```

Automatic parallel backend registration

```r
covdepGE(data_mat, Z, parallel = T, num_workers = 7, stop_cluster = F, n_sigma = 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: -90160.41               Unique conditional dependence structures: 4
## n: 180, variables: 5                         Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 1.598 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.

```r
covdepGE(data_mat, Z, n_sigma = 5)
```

```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -90160.41               Unique conditional dependence structures: 4
## n: 180, variables: 5                         Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 0.839 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_sigma = 200)
out_seq
```

```
##                        Covariate Dependent Graphical Model
##
## Model ELBO: -90084.92               Unique conditional dependence structures: 4
## n: 180, variables: 5                       Hyperparameter grid size: 200 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 24.924 secs
```

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

```
## Detected 7 workers
```

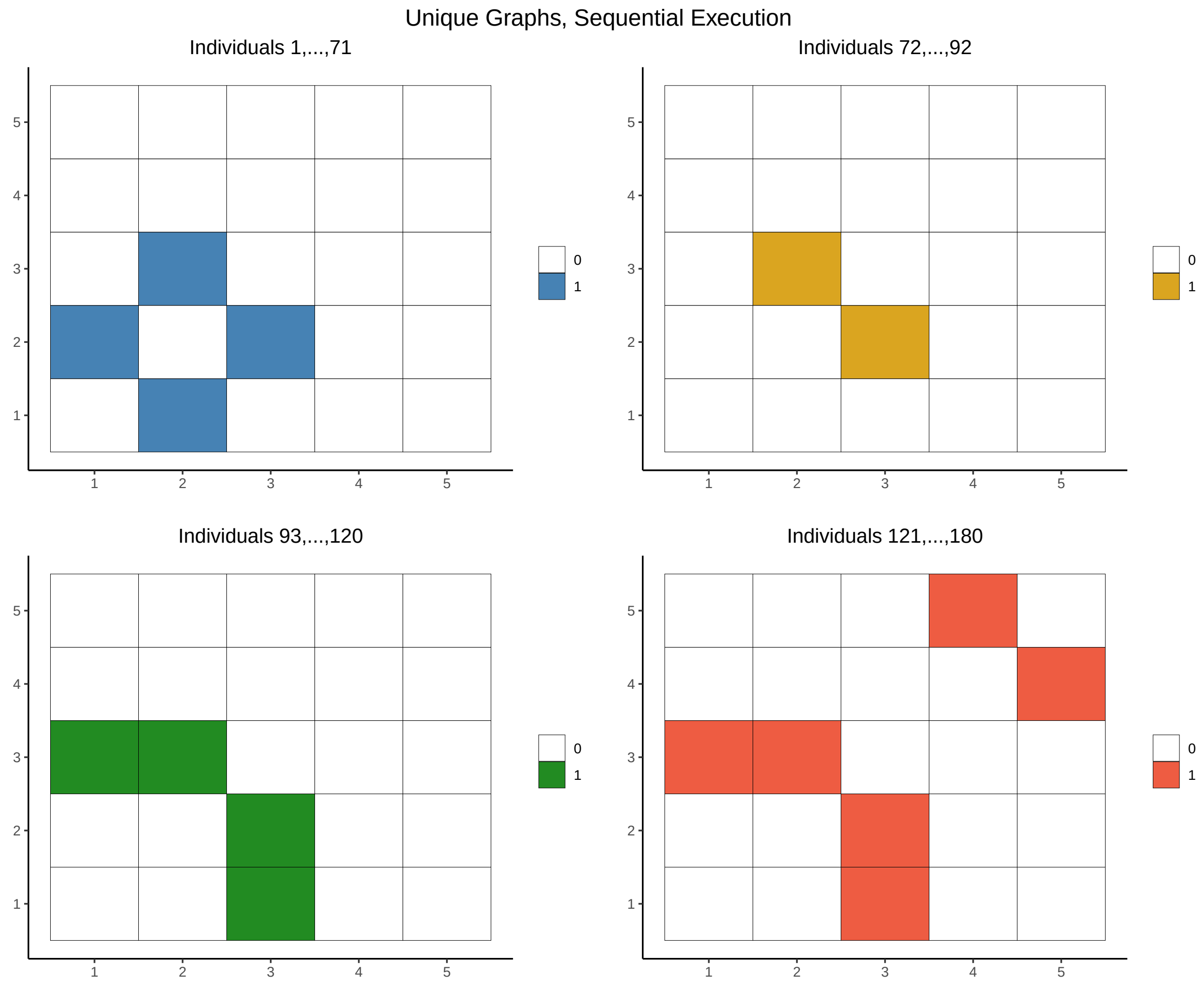
```
out_par
```

```
##                        Covariate Dependent Graphical Model
##
## Model ELBO: -90084.92               Unique conditional dependence structures: 4
## n: 180, variables: 5                       Hyperparameter grid size: 200 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 7.171 secs
```

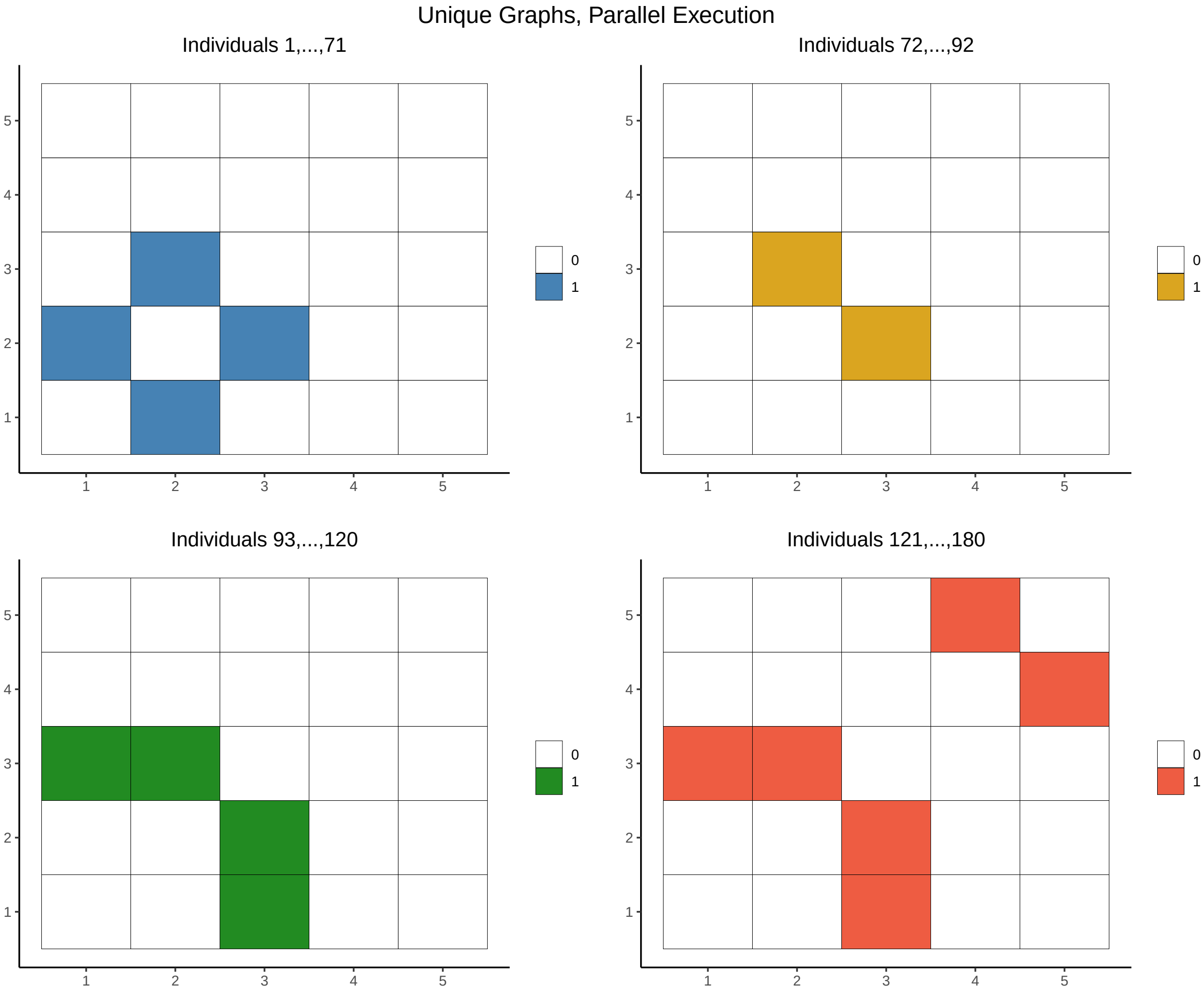
The parallel model outperforms the sequential - additionally, the models produce identical results.

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))
```

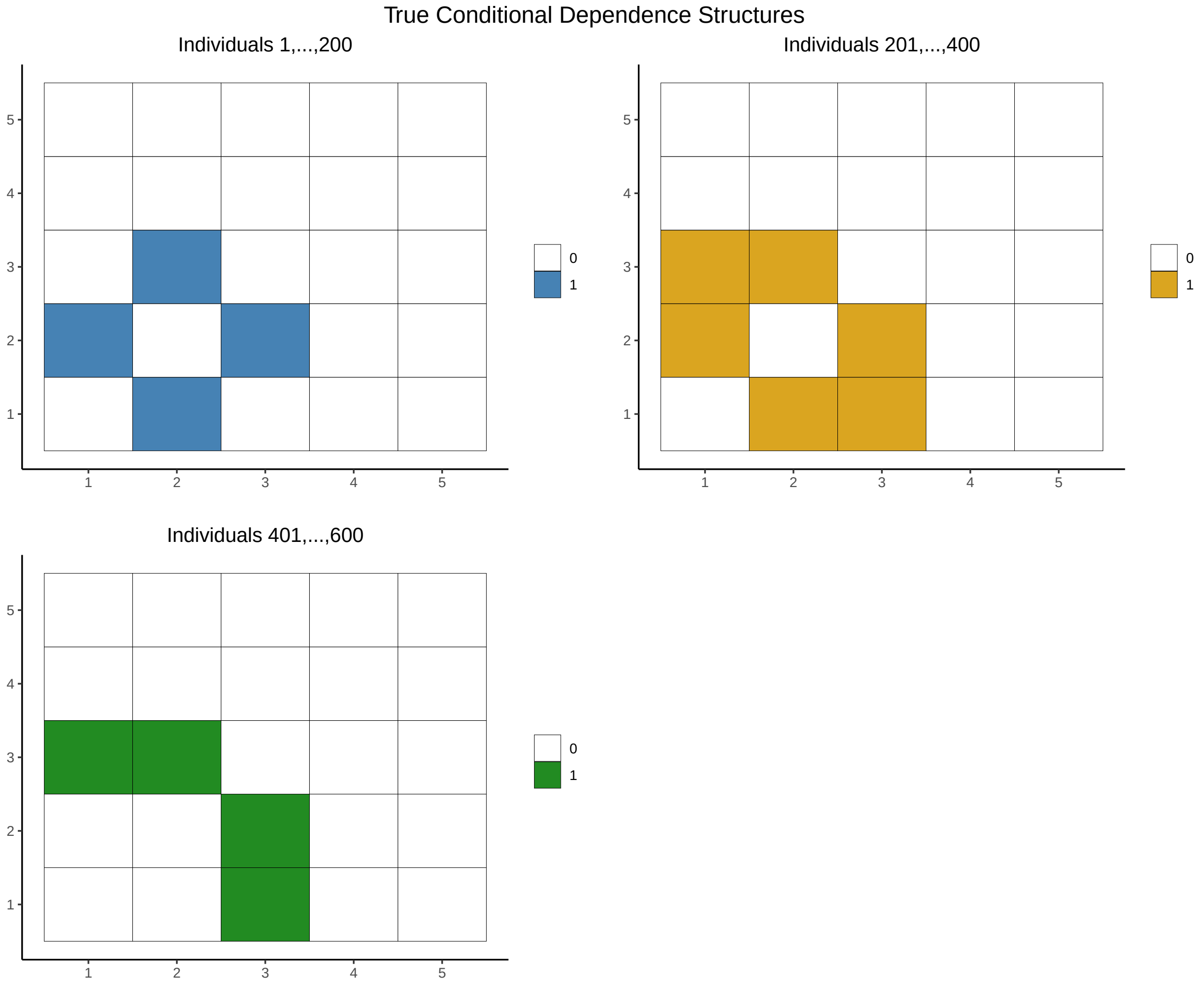


```
annotate_figure(ggarrange(plotlist = plot(out_par, colors)),
                top = text_grob("Unique Graphs, Parallel Execution", size = 15))
```



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

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_sigma = 5)
```

```
## Warning in covdepGE(data_mat, Z, n_sigma = 5): For 1/5 variables, the selected
## value of sigmabeta_sq was on the grid boundary. See return value CAVI_details
```

```
out_seq
```

```
##                        Covariate Dependent Graphical Model
##
## Model ELBO: -980936.13             Unique conditional dependence structures: 6
## n: 600, variables: 5                        Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 42.972 secs
```

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

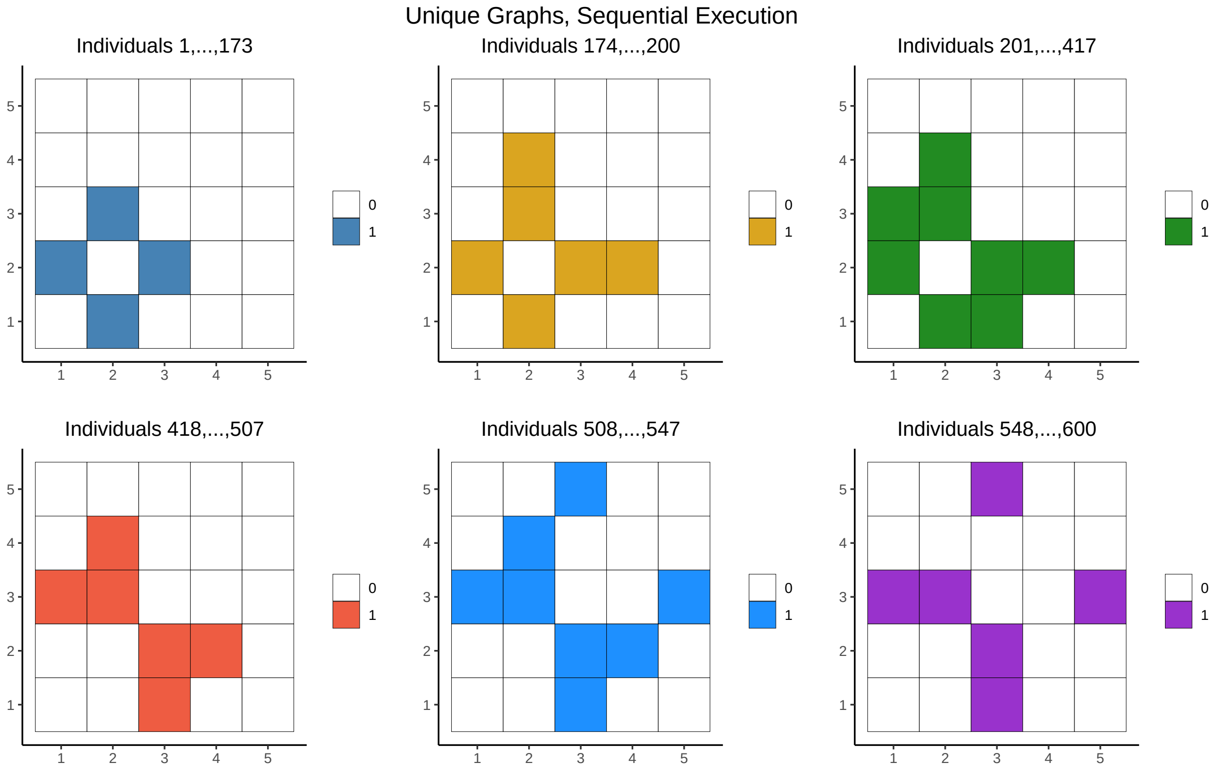
```
## Warning in covdepGE(data_mat, Z, n_sigma = 5, parallel = T, num_workers = 8): No
## registered workers detected; registering doParallel with 8 workers
```

```
## Warning in covdepGE(data_mat, Z, n_sigma = 5, parallel = T, num_workers = 8):
## For 1/5 variables, the selected value of sigmabeta_sq was on the grid boundary.
## See return value CAVI_details
```

```
out_par
```

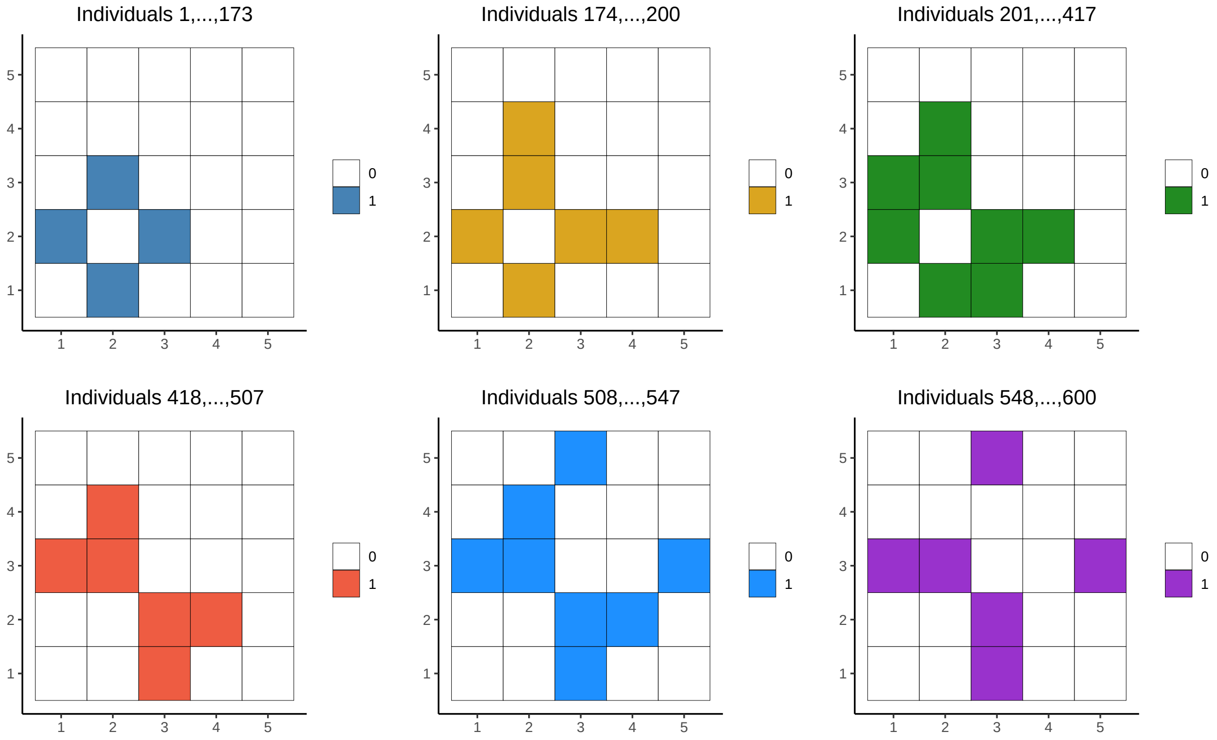
```
##                       Covariate Dependent Graphical Model
##
## Model ELBO: -980936.13             Unique conditional dependence structures: 6
## n: 600, variables: 5                         Hyperparameter grid size: 5 points
## CAVI converged for 5/5 variables
##
## Model fit completed in 25.754 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,...,173

Individuals 174,...,200

Individuals 201,...,417

Individuals 418,...,507

Individuals 508,...,547

Individuals 548,...,600