

Fit Peak Model

William Ruth

31/12/2021

Analysis

In this document, we fit quasi-binomial GLMs in each of the four class size threshold groups to predict peak outbreak size using the epidemiological parameters of our simulation. We compute peak size as a proportion of the overall population size. As such, there is some ambiguity as to whether we should adjust the population size as we change the class size threshold. Unless stated otherwise, we always define peak size relative to the number of students remaining in the network after thresholding, rather than relative to the total number of students enrolled (although in the unthresholded case these numbers coincide). *This is made even more complicated by the fact that we remove all but the largest connected component. All network sizes reported in this document are for the largest remaining connected component after thresholding.*

The extreme outlier in the threshold = 100 group has been removed.

Before doing any model fitting, we plot a histogram of the peak outbreak size, both globally and separately by class size threshold, where the counts are of individual runs of our simulation ($N = 262440$).

```
## Warning: Removed 2 rows containing missing values (geom_bar).

## Warning: Removed 2 rows containing missing values (geom_bar).

## Warning: Removed 2 rows containing missing values (geom_bar).

## Warning: Removed 2 rows containing missing values (geom_bar).

## Saving 6.5 x 4.5 in image
## Saving 6.5 x 4.5 in image
```

Next, we produce the same plots, but with the peak size proportion computed relative to the total number of students, rather than the number of students remaining in the network after thresholding.

```
## Warning: Removed 8 rows containing missing values (geom_bar).

## Warning: Removed 8 rows containing missing values (geom_bar).

## pdf
## 2

## pdf
## 2
```

There are clearly different levels of variability between the different threshold groups. This suggests that we should use an extended quasi-likelihood model with different overdispersion parameters in each group.

The structure of our simulation study includes pure replication within each parameter setting (specifically, we have 10 replicates for each parameter combination). This allows us to investigate the pure replication variability, which informs what we expect to see when we fit our GLM. The following plot gives average CII

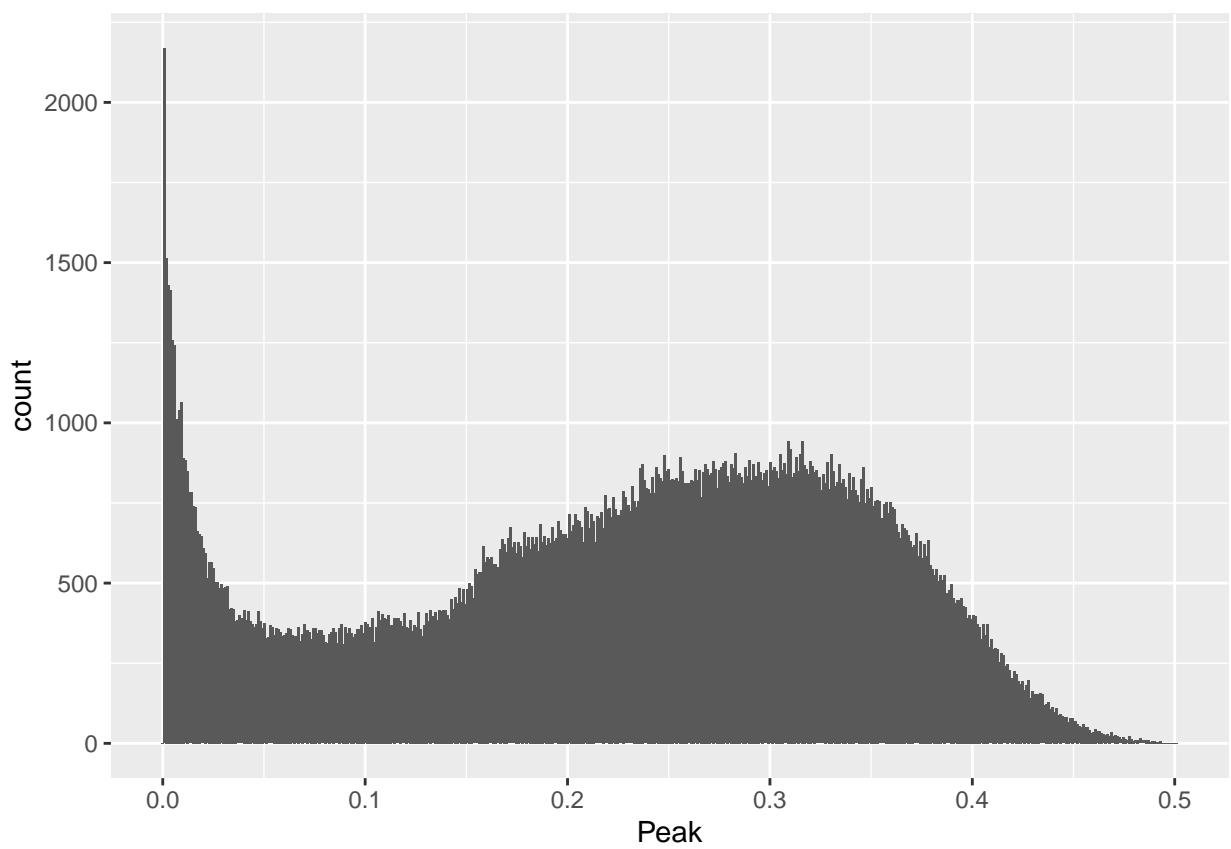


Figure 1: Histogram of peak size.

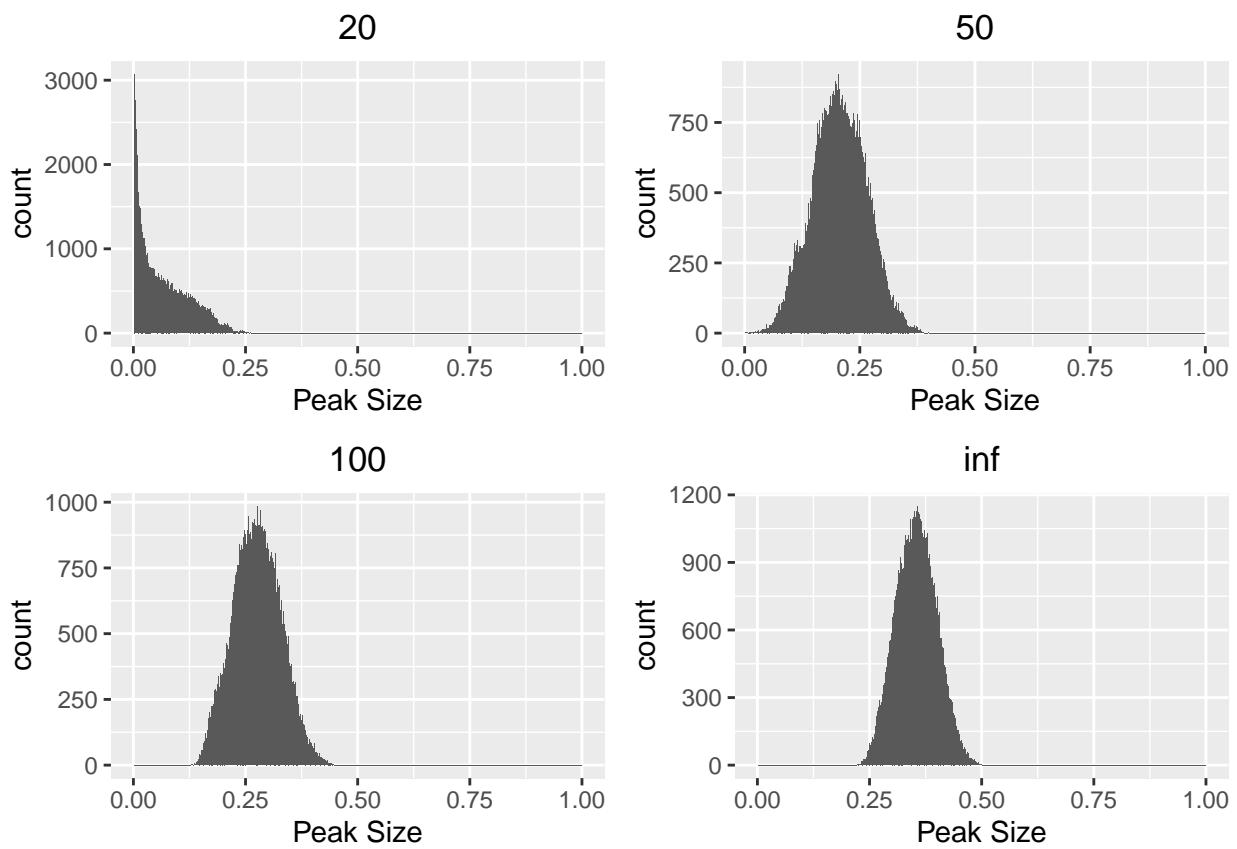


Figure 2: Histograms of peak outbreak size for each class size threshold with uniform axis scaling.

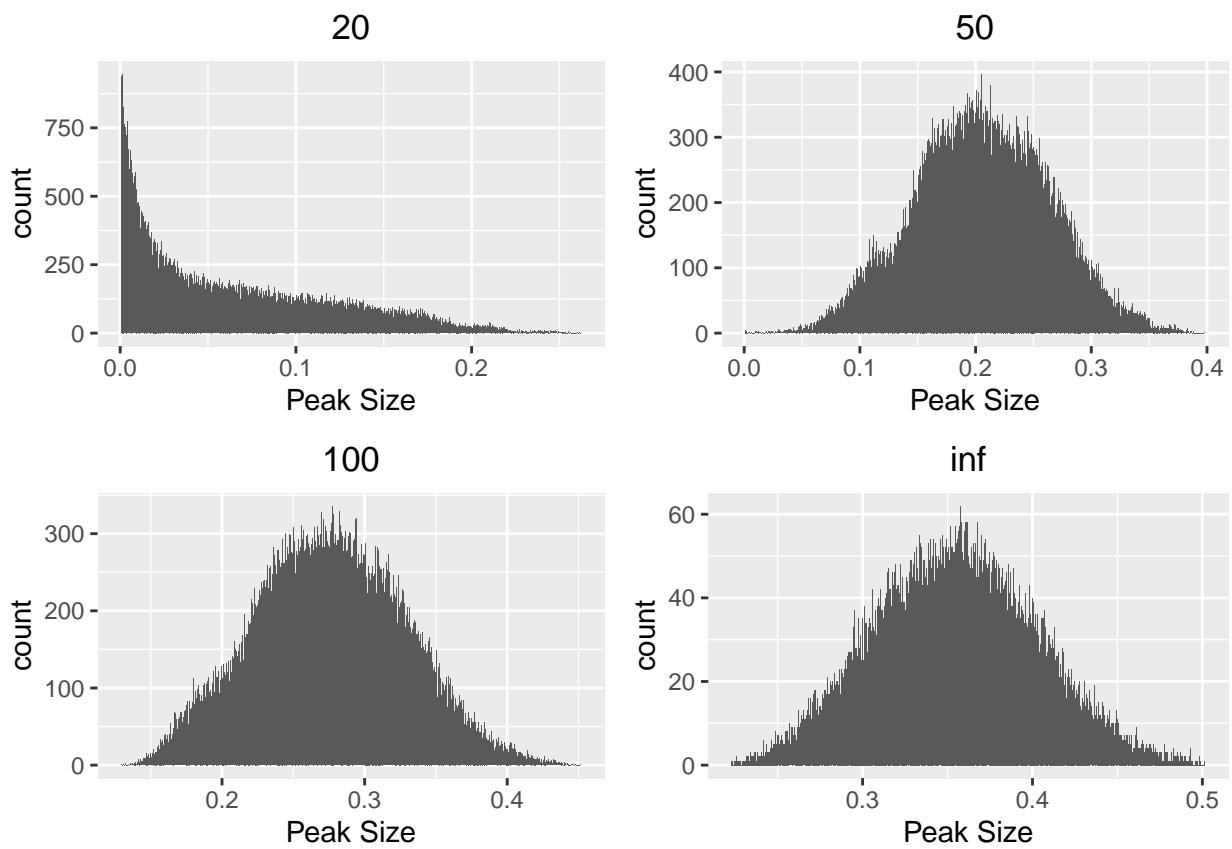


Figure 3: Histograms of peak outbreak size for each class size threshold with heterogeneous axis scaling.

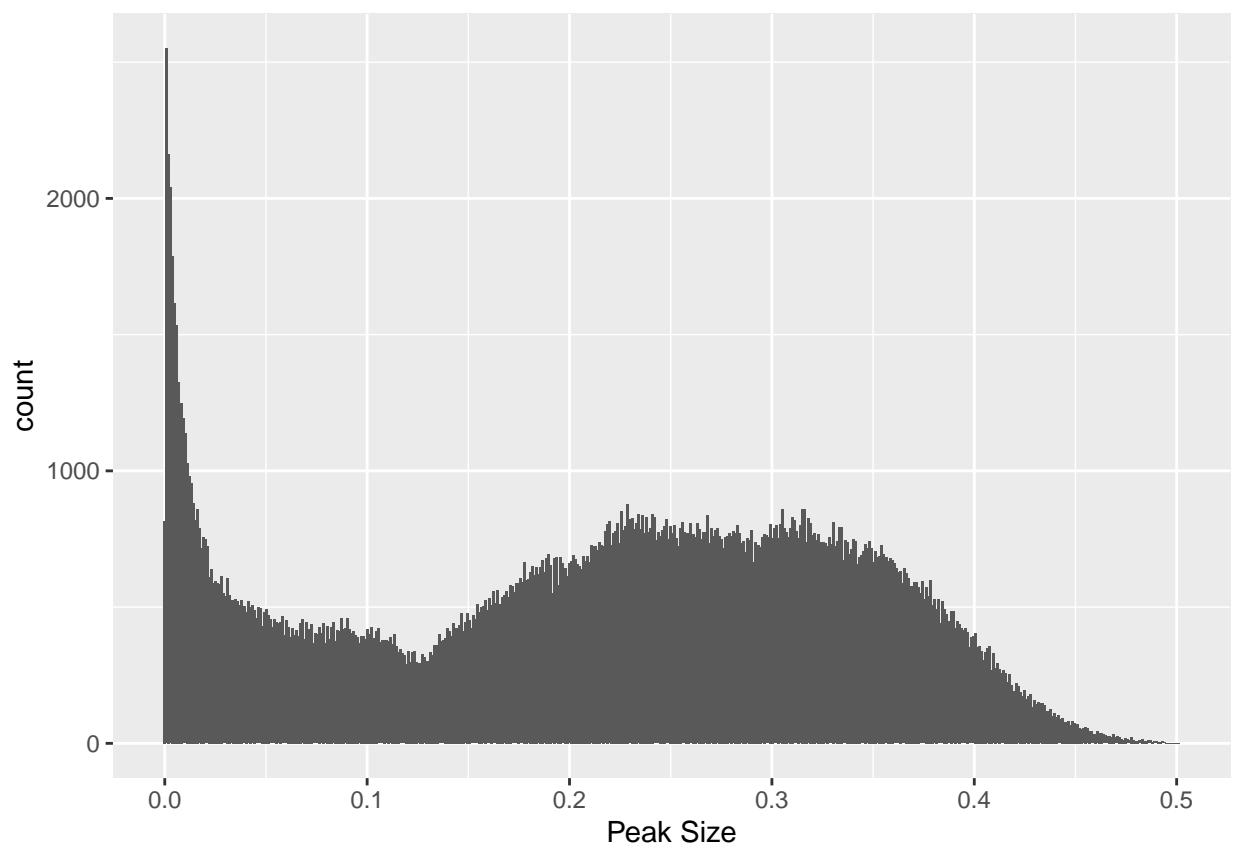


Figure 4: Histogram of peak size relative to total number of students.

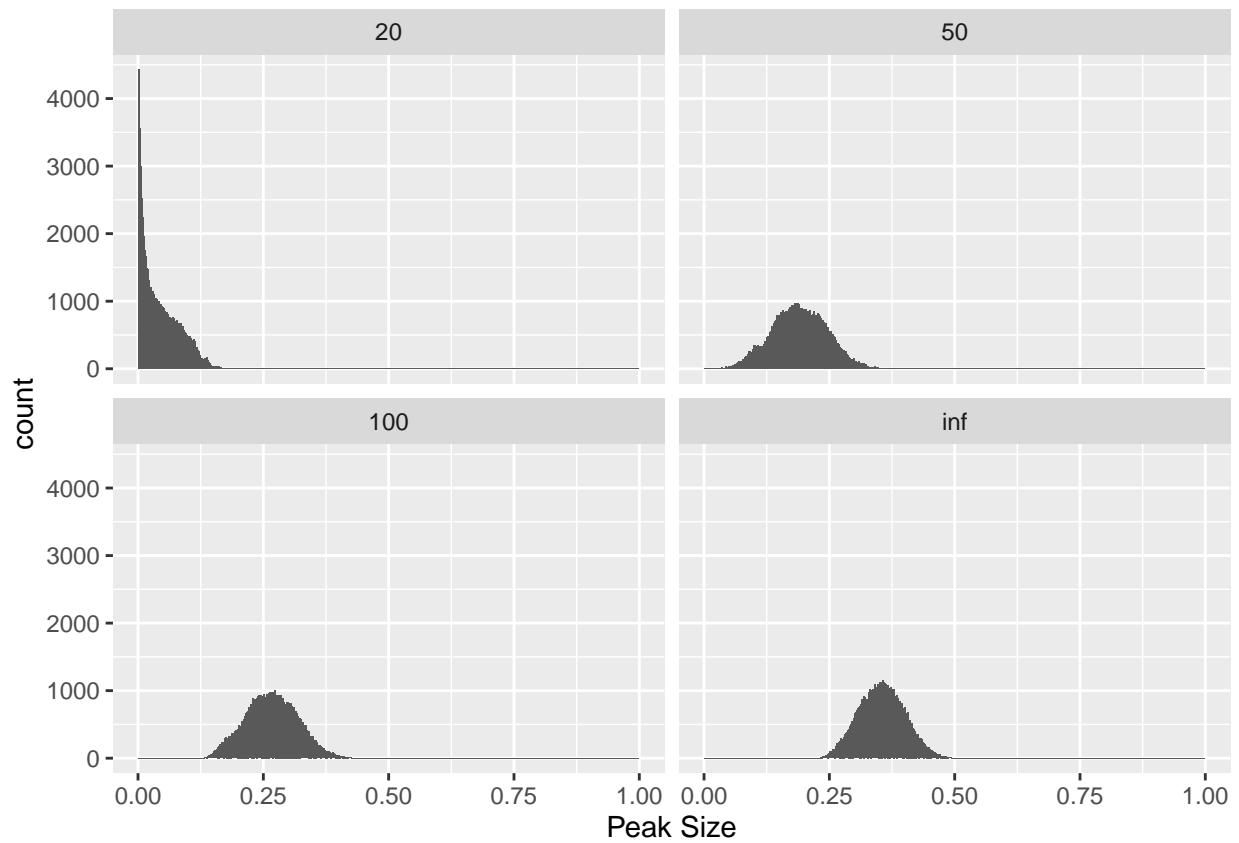


Figure 5: Histograms of peak sizes relative to total number of students for each class size threshold with uniform axis scaling.

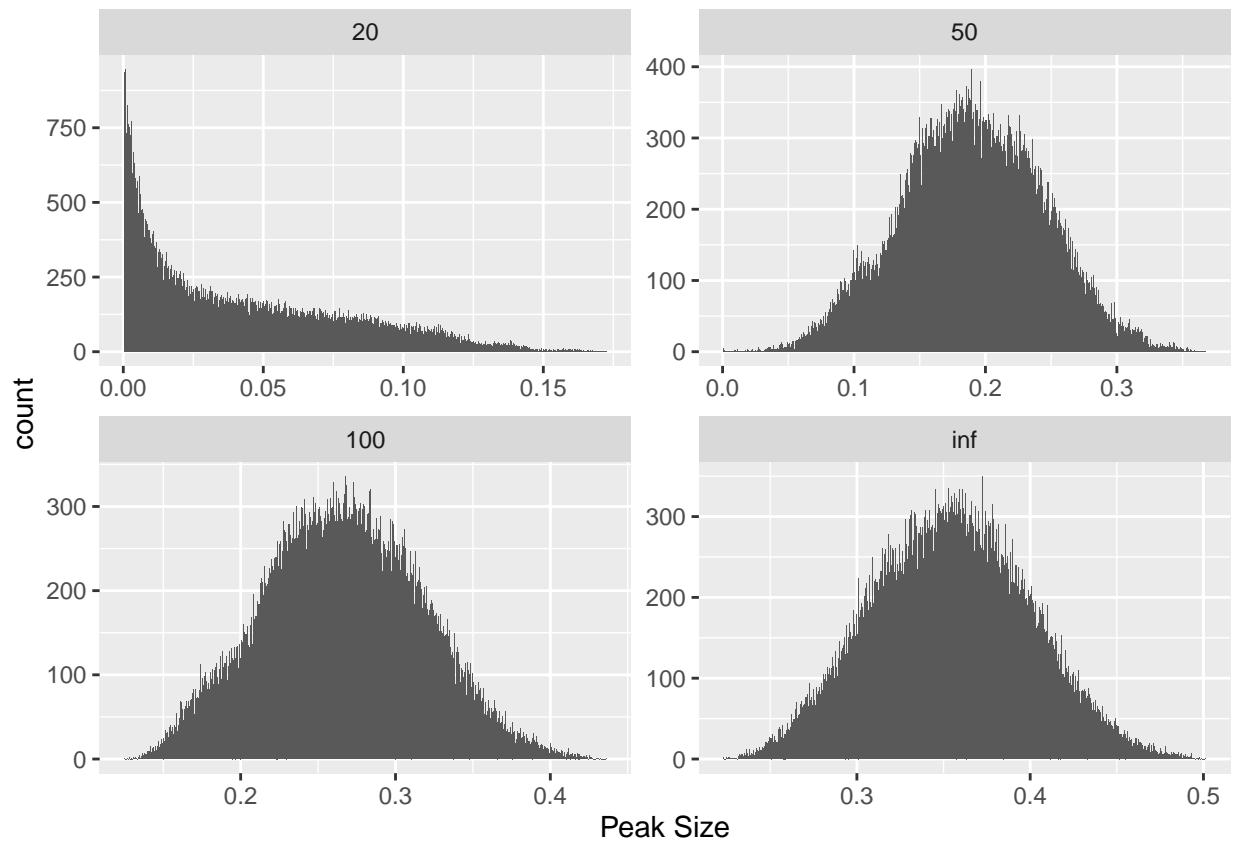
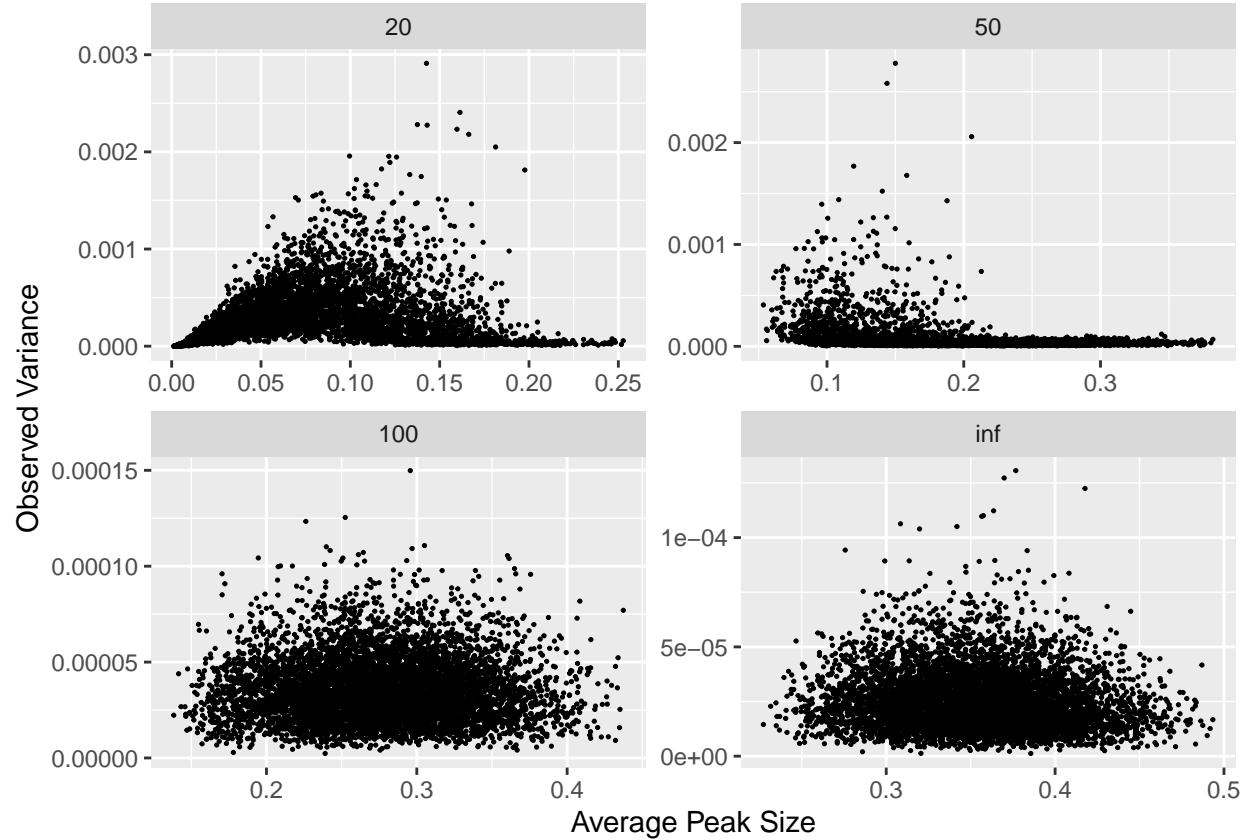
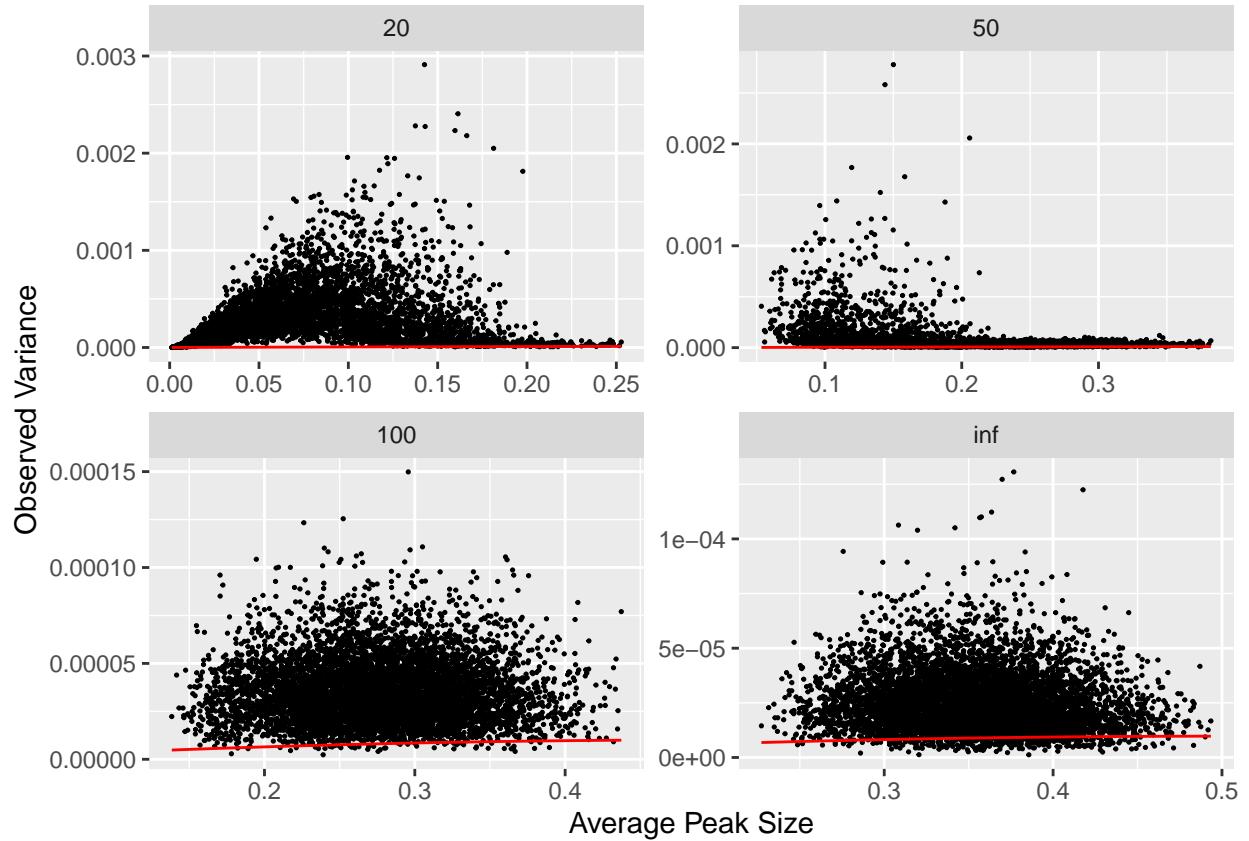


Figure 6: Histograms of peak sizes relative to total number of students for each class size threshold with heterogeneous axis scaling.

vs sample variance for each parameter combination (**Footnote:** the average and variance are computed over the 10 replicates within a single parameter combination).



Next, we reproduce these plots but also add a reference line at the variance which is predicted by the binomial distribution (i.e. at $p(1 - p)/N$, where N is the number of students remaining in the network).



```
## pdf
## 2
```

Our data clearly exhibit overdispersion relative to the predictions of a binomial model.

Investigate interactions

Next, we fit a sequence of models with the first four levels of interaction terms, as well as the null model with only an intercept. We summarize these fits with a few comparisons.

First, we give the deviance improvement of each fit relative to the null model divided by the number of fitted parameters in that model.

```
##          Main    2-Way    3-Way    4-Way
## Num Pars 17.0   129.0   577.00  2257.000
## 20        3055752.2 428285.0 96228.31 24643.727
## 50        1872860.3 260298.4 58405.30 14942.919
## 100       1292471.3 178223.9 39882.32 10198.531
## inf       842679.5 116160.4 25994.76  6647.685
```

Next, we give the relative improvement per parameter within each class size threshold.

```
##          Main    2-Way    3-Way    4-Way
## Num Pars 17 129.0000000 577.0000000 2.257000e+03
## 20         1  0.1401570  0.03149087 8.064701e-03
## 50         1  0.1389844  0.03118508 7.978662e-03
## 100        1  0.1378939  0.03085741 7.890721e-03
## inf        1  0.1378465  0.03084774 7.888747e-03
```

We also report the improvement at each step over the next-largest model.

```
##      Main   2-Way   3-Way   4-Way
## 20  51947788 3300981 274964.83 97158.623
## 50  31838625 1739862 121371.64 26309.644
## 100 21972012 1018865 21220.51 5986.123
## inf 14325552  659138 14286.68 4848.807
```

And the corresponding relative improvements within each class size threshold.

```
##      Main   2-Way   3-Way   4-Way
## 20      1 0.06354421 0.0052930999 0.0018703130
## 50      1 0.05464628 0.0038120880 0.0008263436
## 100     1 0.04637103 0.0009657971 0.0002724431
## inf     1 0.04601135 0.0009972865 0.0003384726
```

We now proceed to fit the models to which we have been alluding. First, the four models, with one for each level of class size threshold.

Diagnostics

We now investigate some of the standard diagnostic plots for evaluating GLM fits. Figures 7 and 8 give the deviance and Pearson residuals respectively for each class size threshold.

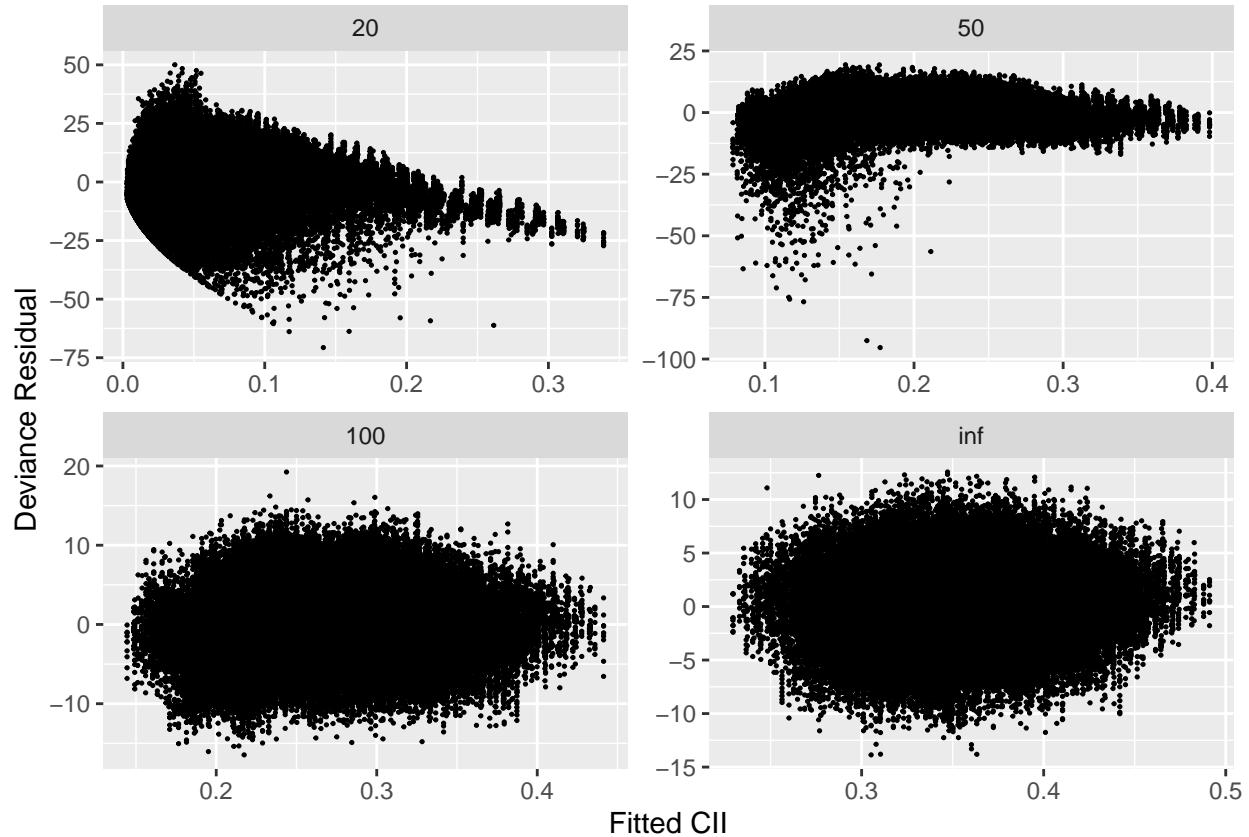


Figure 7: Deviance residuals for each class size threshold.

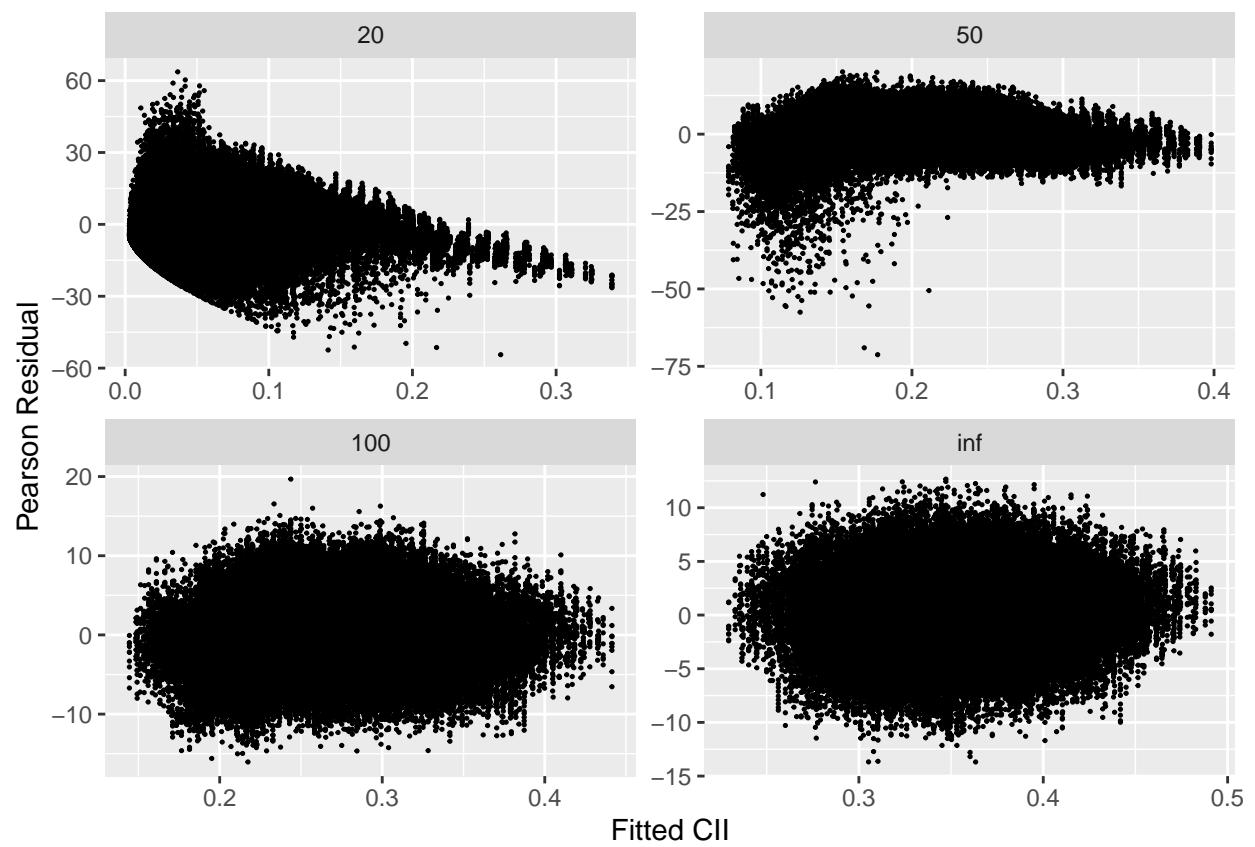


Figure 8: Pearson residuals for each class size threshold.

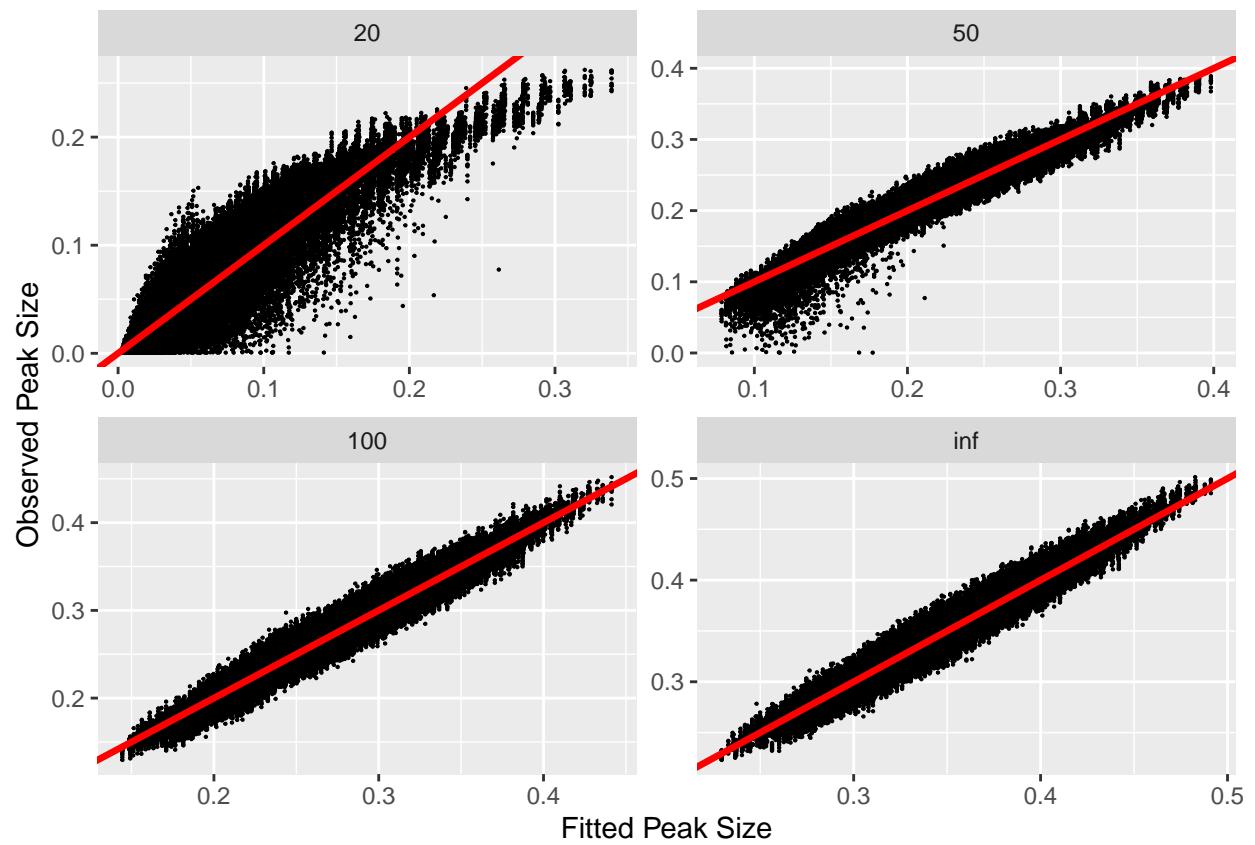


Figure 9: Observed peak size vs fitted values.

Results

We now report some summaries.

Deviance Changes

To start, we extract the deviance improvement provided by each variable when added to a model already containing the other predictors.

Table 1: Deviance improvements within each class size threshold.

	20	50	100	inf
infect_prop_A	5.17E+05	2.97E+05	1.60E+05	6.95E+04
infect_prop_I1	6.63E+06	3.26E+06	2.33E+06	1.47E+06
infect_param_I2	3.30E+07	1.36E+07	7.22E+06	2.75E+06
advance_prob_E	2.06E+05	2.11E+05	2.75E+05	4.38E+05
advance_prob_A	1.79E+05	2.80E+05	2.91E+05	3.12E+05
advance_prob_I1	2.45E+06	2.76E+06	2.70E+06	2.56E+06
advance_prob_I2	7.63E+06	9.46E+06	7.29E+06	5.29E+06
E_to_A_prob	2.08E+06	2.22E+06	1.83E+06	1.49E+06
Residual	9.22E+06	2.73E+06	1.33E+06	8.63E+05

```
## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Wed Feb 23 10:55:45 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{ccccc}
##   \hline
##   & 20 & 50 & 100 & inf \\
##   \hline
##   infect\_prop\_A & 5.17E+05 & 2.97E+05 & 1.60E+05 & 6.95E+04 \\
##   infect\_prop\_I1 & 6.63E+06 & 3.26E+06 & 2.33E+06 & 1.47E+06 \\
##   infect\_param\_I2 & 3.30E+07 & 1.36E+07 & 7.22E+06 & 2.75E+06 \\
##   advance\_prob\_E & 2.06E+05 & 2.11E+05 & 2.75E+05 & 4.38E+05 \\
##   advance\_prob\_A & 1.79E+05 & 2.80E+05 & 2.91E+05 & 3.12E+05 \\
##   advance\_prob\_I1 & 2.45E+06 & 2.76E+06 & 2.70E+06 & 2.56E+06 \\
##   advance\_prob\_I2 & 7.63E+06 & 9.46E+06 & 7.29E+06 & 5.29E+06 \\
##   E\_to\_A\_prob & 2.08E+06 & 2.22E+06 & 1.83E+06 & 1.49E+06 \\
##   Residual & 9.22E+06 & 2.73E+06 & 1.33E+06 & 8.63E+05 \\
##   \hline
## \end{tabular}
## \caption{Deviance increase caused by omitting each variable from a model containing all others. Model}
## \label{tab:dev_peak}
## \end{table}
```

For reference, we give the relative change in deviance compared to the largest in each group.

Table 2: Relative deviance improvements within each class size threshold.

	20	50	100	inf
infect_prop_A	0.016	0.022	0.022	0.013
infect_prop_I1	0.2	0.24	0.32	0.28

	20	50	100	inf
infect_param_I2	1	1	0.99	0.52
advance_prob_E	0.0062	0.016	0.038	0.083
advance_prob_A	0.0054	0.021	0.04	0.059
advance_prob_I1	0.074	0.2	0.37	0.48
advance_prob_I2	0.23	0.7	1	1
E_to_A_prob	0.063	0.16	0.25	0.28
Residual	0.28	0.2	0.18	0.16

```

## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Wed Feb 23 10:55:45 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{ccccc}
##   \hline
## & 20 & 50 & 100 & inf \\
##   \hline
## infect\_prop\_A & 0.016 & 0.022 & 0.022 & 0.013 \\
## infect\_prop\_I1 & 0.2 & 0.24 & 0.32 & 0.28 \\
## infect\_param\_I2 & 1 & 1 & 0.99 & 0.52 \\
## advance\_prob\_E & 0.0062 & 0.016 & 0.038 & 0.083 \\
## advance\_prob\_A & 0.0054 & 0.021 & 0.04 & 0.059 \\
## advance\_prob\_I1 & 0.074 & 0.2 & 0.37 & 0.48 \\
## advance\_prob\_I2 & 0.23 & 0.7 & 1 & 1 \\
## E\_to\_A\_prob & 0.063 & 0.16 & 0.25 & 0.28 \\
## Residual & 0.28 & 0.2 & 0.18 & 0.16 \\
##   \hline
## \end{tabular}
## \caption{Relative deviance increases within each class size threshold.}
## \label{tab:rel_dev_peak}
## \end{table}

```

We also rank the predictors within each group in decreasing order of deviance improvement.

Table 3: Ranked deviance improvements within each class size threshold. 1 has the greatest improvement and 8 has the least.

	20	50	100	inf
infect_prop_A	6	6	8	8
infect_prop_I1	3	3	4	5
infect_param_I2	1	1	2	2
advance_prob_E	7	8	7	6
advance_prob_A	8	7	6	7
advance_prob_I1	4	4	3	3
advance_prob_I2	2	2	1	1
E_to_A_prob	5	5	5	4

```

## % latex table generated in R 4.1.2 by xtable 1.8-4 package
## % Wed Feb 23 10:55:45 2022
## \begin{table}[ht]
## \centering
## \begin{tabular}{ccccc}
##   \hline
## infect\_prop\_A & 6 & 6 & 8 & 8 \\
## infect\_prop\_I1 & 3 & 3 & 4 & 5 \\
## infect\_param\_I2 & 1 & 1 & 2 & 2 \\
## advance\_prob\_E & 7 & 8 & 7 & 6 \\
## advance\_prob\_A & 8 & 7 & 6 & 7 \\
## advance\_prob\_I1 & 4 & 4 & 3 & 3 \\
## advance\_prob\_I2 & 2 & 2 & 1 & 1 \\
## E\_to\_A\_prob & 5 & 5 & 5 & 4
##   \hline
## \end{tabular}

```

```

## & 20 & 50 & 100 & inf \\
## \hline
## infect\_prop\_A & 6 & 6 & 8 & 8 \\
## infect\_prop\_I1 & 3 & 3 & 4 & 5 \\
## infect\_param\_I2 & 1 & 1 & 2 & 2 \\
## advance\_prob\_E & 7 & 8 & 7 & 6 \\
## advance\_prob\_A & 8 & 7 & 6 & 7 \\
## advance\_prob\_I1 & 4 & 4 & 3 & 3 \\
## advance\_prob\_I2 & 2 & 2 & 1 & 1 \\
## E\_to\_A\_prob & 5 & 5 & 5 & 4 \\
## \hline
## \end{tabular}
## \caption{Deviance increase ranks within each class size threshold. 1 has the greatest increase and 8 1
## \label{tab:dev_peak_ranks}
## \end{table}

```

Overdispersion Parameter

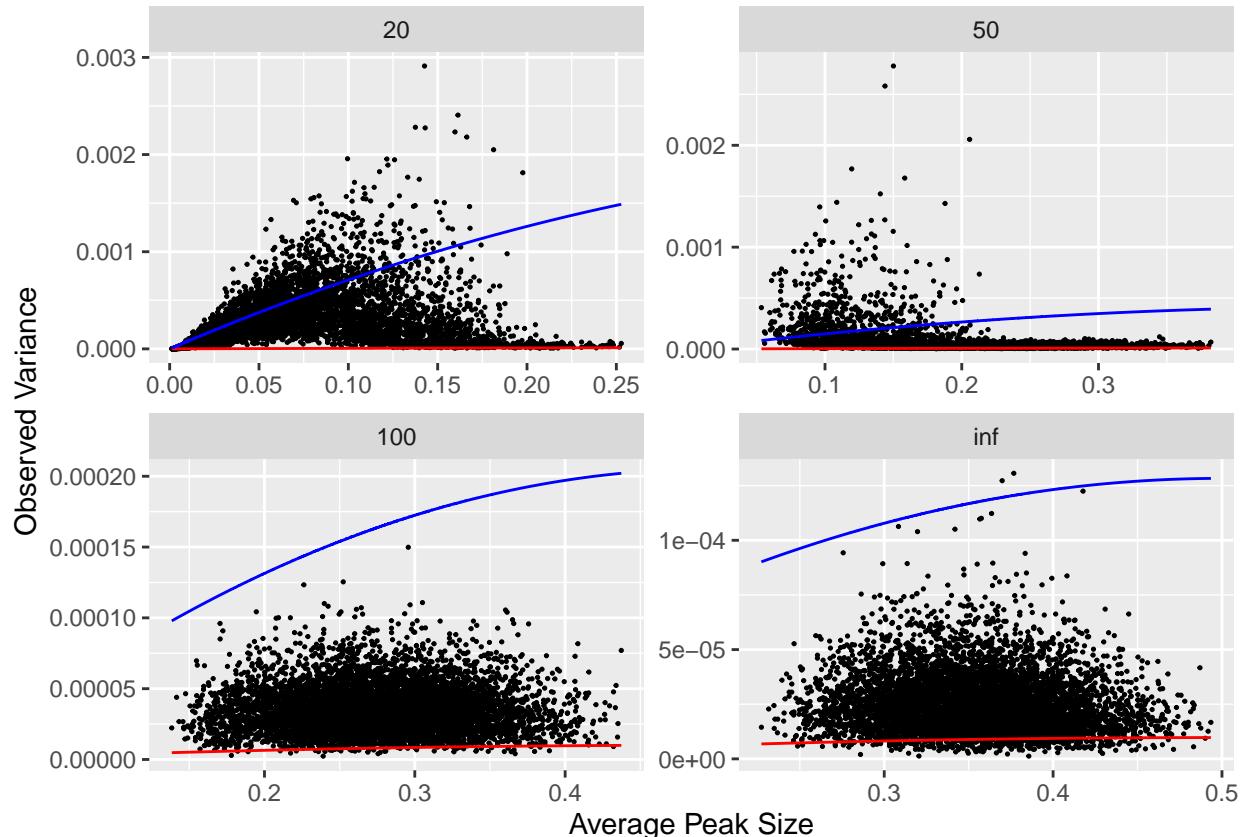
Next, we present the fitted overdispersion parameter from each group.

```

##    20      50     100     inf
## "133"   "39"   "20"   "13"

```

Rescaled Overdispersion Plots



```

## pdf
## 2

```

Global Model

Caution: The analyses in this section are all of proportions relative to the total number of students, rather than the number in the largest remaining connected component as in the rest of this document.

For reference, we repeat the above analysis on the ordinary quasi-likelihood model. That is, we retain the interaction terms for the mean model, but use a single global overdispersion parameter.

The deviance changes for excluding each variable individually from our model (while retaining all others) are as follows.

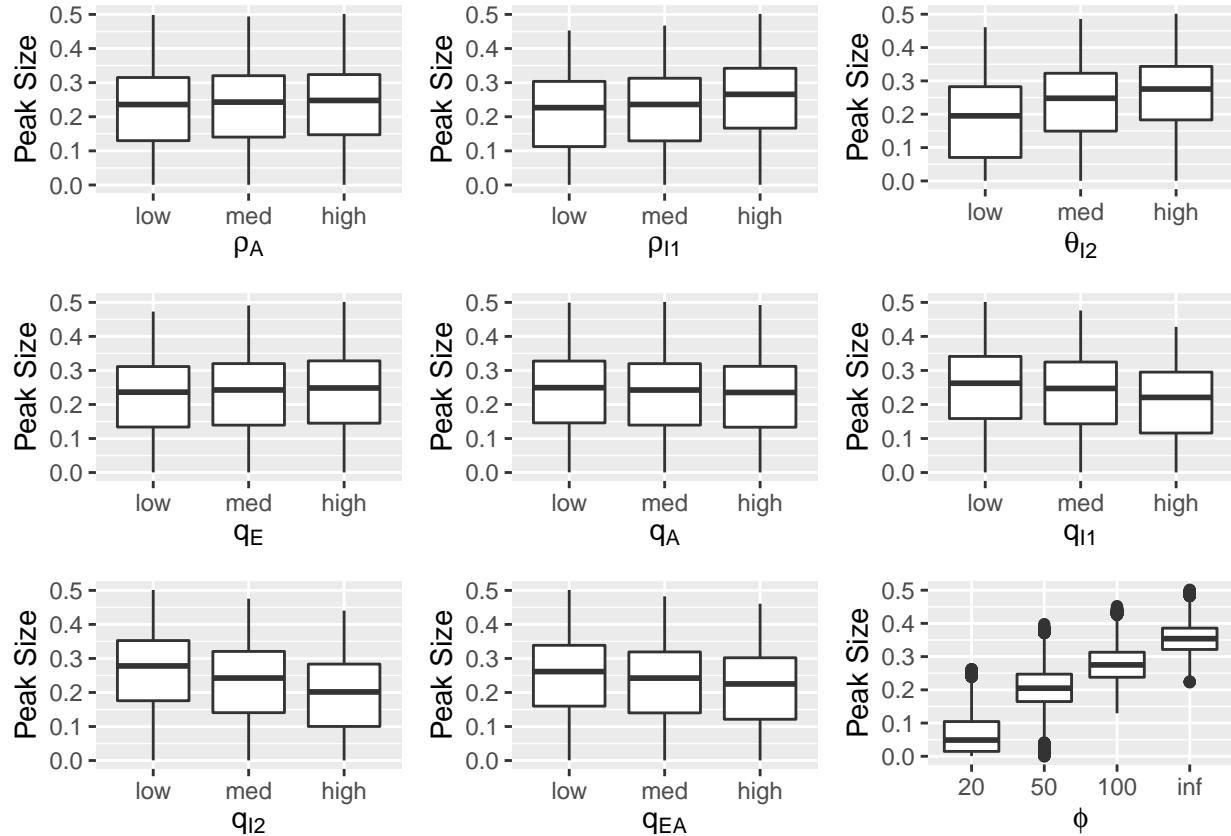
```
##   infect_prop_A  infect_prop_I1  infect_param_I2  advance_prob_E  advance_prob_A
##   "4.77E+05"      "5.35E+06"      "3.38E+07"      "6.37E+04"      "4.91E+04"
##  advance_prob_I1  advance_prob_I2    E_to_A_prob
##   "1.07E+06"      "4.41E+06"      "1.14E+06"
```

Similarly, the relative deviance improvements are:

Finally, the fitted global overdispersion parameter is 70, which is the mean of the individual group models' overdispersion parameters.

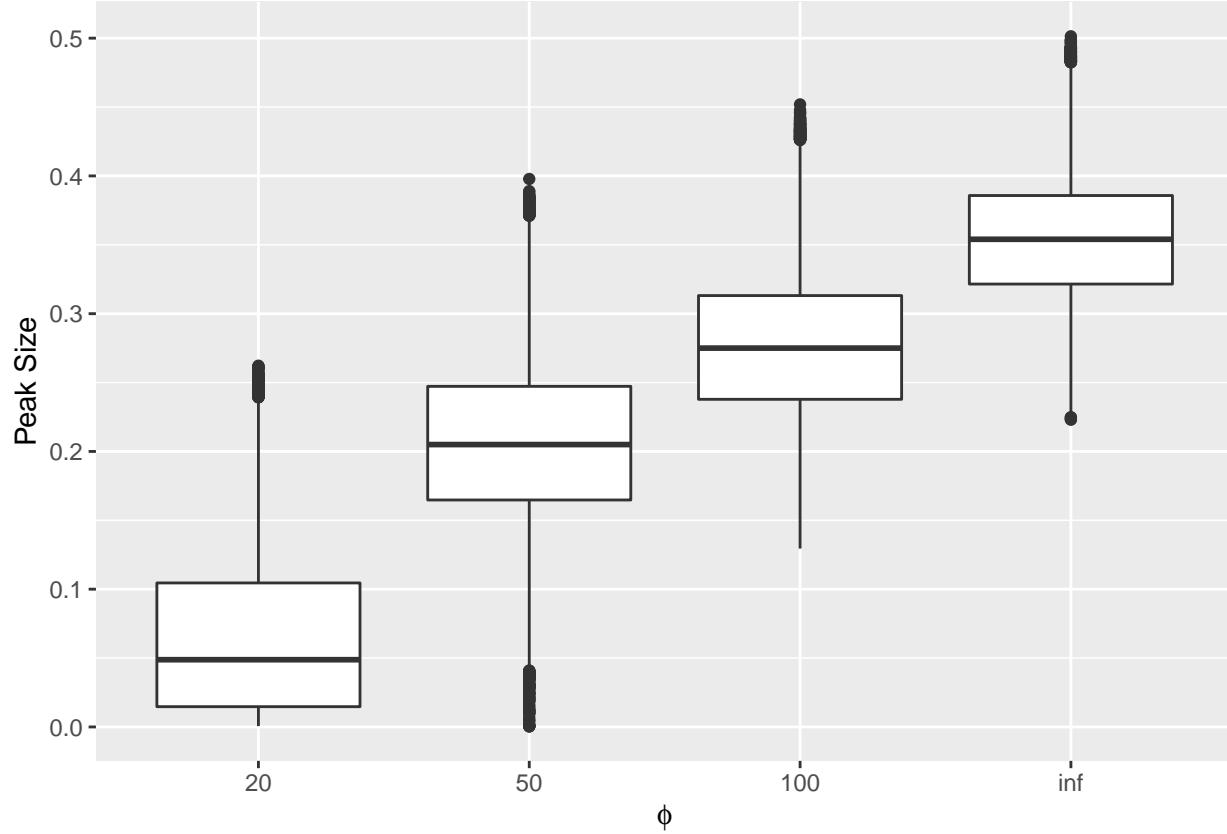
Marginal Behaviour of Each Predictor

The following figure gives a sequence of boxplots for the CII in our simulation. Each plot corresponds to a single predictor, and each box contains all simulation runs with that level of the predictor.



```
## pdf
## 2
```

The threshold variable clearly has the greatest association with CII, so we reproduce that boxplot **with greater resolution**.



Discussion

In Tables 1-3, we see that the infectiousness parameter for the symptomatic group has the largest deviance improvement in every model. Recall that the infectiousness parameter for the symptomatic group gives the absolute proportionality constant for infection probability (proportional to $1 / \sqrt{\text{class size}}$). However, the parameter for the presymptomatic and asymptomatic groups reflect their relative infectiousness compared to the symptomatic group. Thus, increasing or decreasing the infectiousness parameter for the symptomatic group changes the infection probability for all infectious compartments, while the parameters for presymptomatic and asymptomatic only affect their own compartment.

Based on the relative deviance contributions across variables given in Tables 2 and 3, we see that the infectiousness of symptomatic individuals is most strongly associated with CII. As discussed above, this parameter must be interpreted carefully in our model, so we

The behaviour of the threshold = 20 group is qualitatively different from the other threshold levels. In Figures 2 and 3, we see that the CIIs clump around 0 with a long right tail when threshold is 20, while the other groups have their mass concentrated above 0.5.