Assignment 3

Huang Yingshan, Kahlbacher Fabian, Wattenberg Yannick, Aggeler Samuel

11/25/2021

Task 2

(This file only asses the GOF for the model produced from the simSAOMs.r file. For better readability we copied some of the code to produce the plots, this is only in addiotion to and equivalant with the code in the r-file)

```
source('simSAOMs.r')
```

```
## If you use this algorithm object, siena07 will create/use an output file net1_net2.txt .
## [1] 100
## [1] 200
## [1] 300
## [1] 400
## [1] 500
## [1] 600
## [1] 700
## [1] 800
## [1] 900
## [1] 1000
```

```
## -- Attaching packages -------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5      v purrr   0.3.4
## v tibble  3.1.5      v dplyr   1.0.7
## v tidyr   1.1.4      v stringr 1.4.0
## v readr   2.0.2      v forcats 0.5.1
```

```
## -- Conflicts ----------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
ggplot(indegDistDf, aes(degree, nnodes)) +
  geom_violin(trim = FALSE, scale = "width") +
  stat_summary(fun = mean, geom = "point", size = 2) +
  geom_boxplot(width = 0.2, fill = "gray", outlier.shape = 4) +
  geom_point(data = obsIndegData,
             col = "red", size = 2) +
  geom_line(data = obsIndegData, aes(group = 1),
            col = "red", size = 0.5) +
  geom_line(data = percIndeg, mapping = aes(group = quant),
            col = "gray", linetype = "dashed") +
  theme_bw() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
```

```
    axis.title.y = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank()
) +
scale_x_discrete(labels = as.numeric)
```
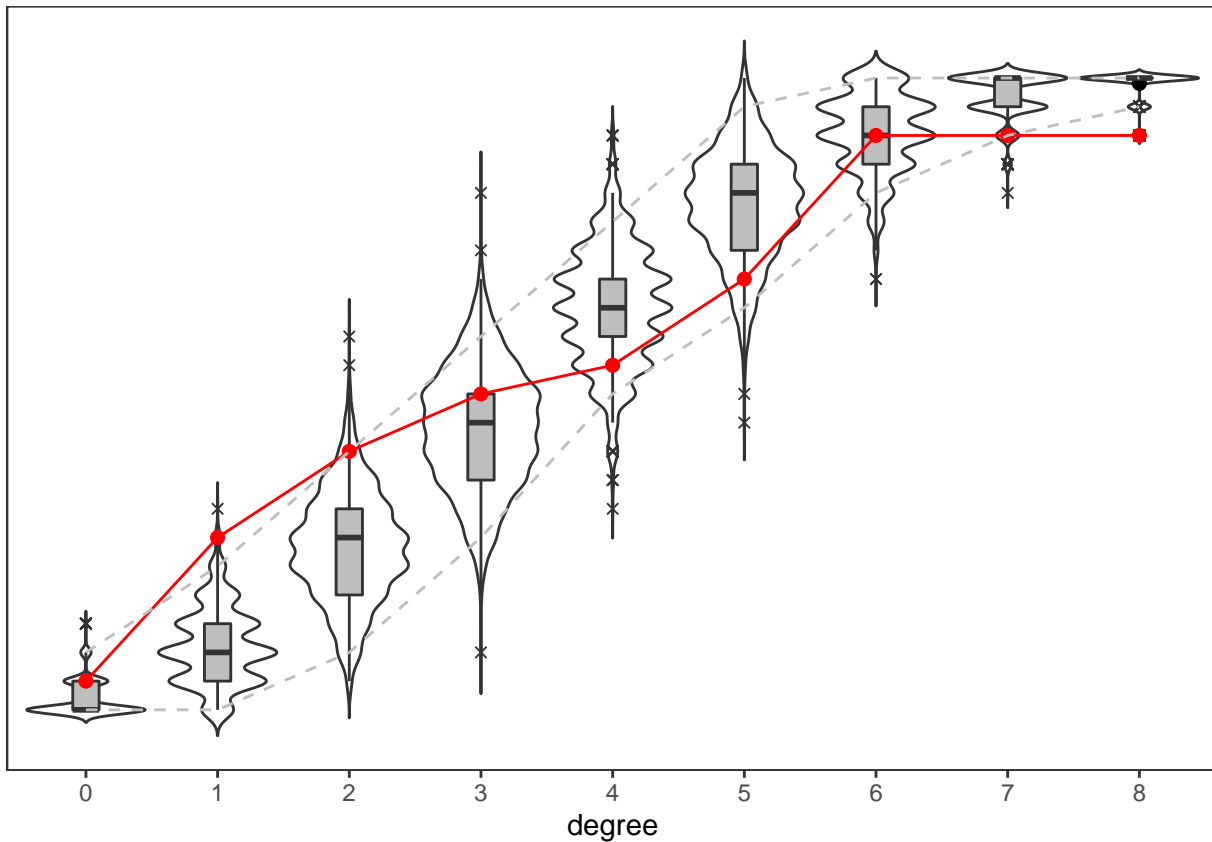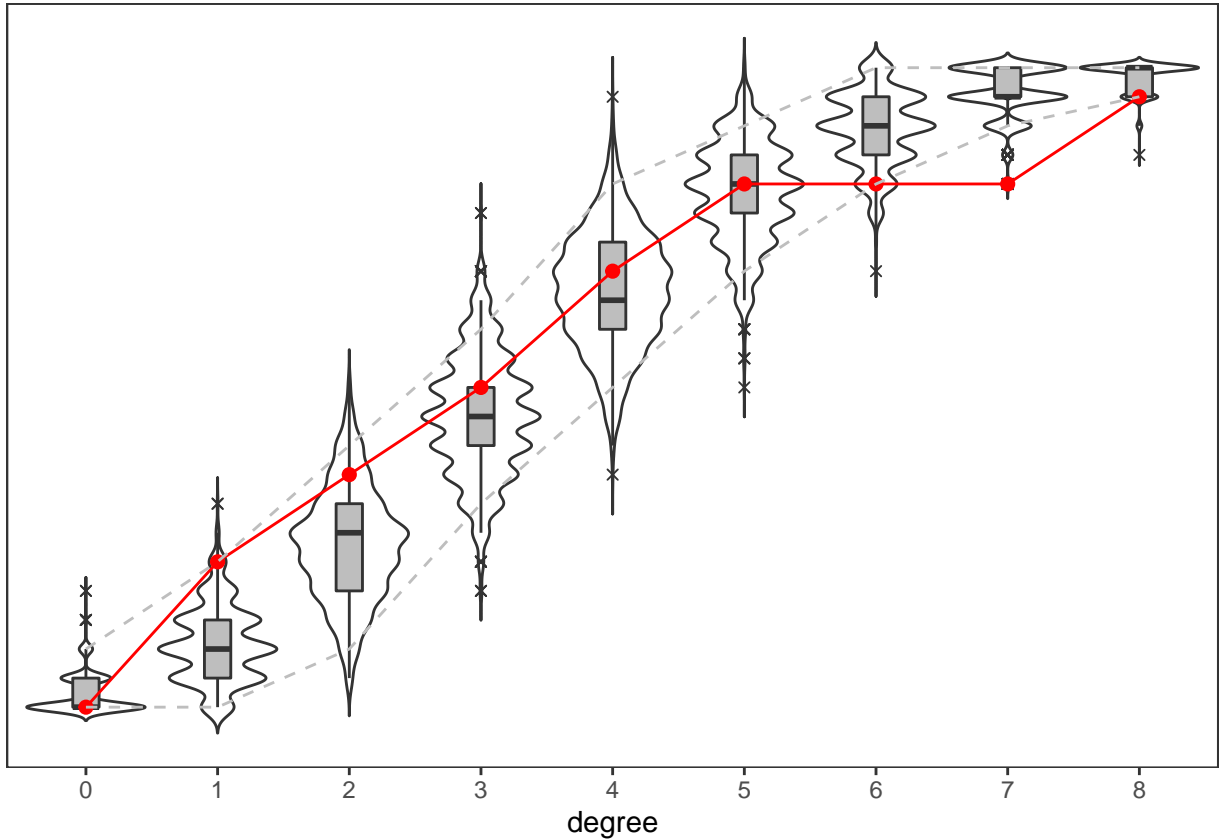

As one can see for the in-degree distribution the observation is not in the 95% confidence interval for the degrees of 1 and 8 and on the border of the intervall for degrees 2 and 7.

```
ggplot(outdegDistDf, aes(degree, nnodes)) +
  geom_violin(trim = FALSE, scale = "width") +
  stat_summary(fun = mean, geom = "point", size = 2) +
  geom_boxplot(width = 0.2, fill = "gray", outlier.shape = 4) +
  geom_point(data = obsOutdegData,
             col = "red", size = 2) +
  geom_line(data = obsOutdegData, aes(group = 1),
            col = "red", size = 0.5) +
  geom_line(data = percOutdeg, mapping = aes(group = quant),
            col = "gray", linetype = "dashed") +
  theme_bw() +
  theme(
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.title.y = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks.y = element_blank()
```

```
) +
scale_x_discrete(labels = as.numeric)
```


For the out-degree distribution the observation outside the 95% confidence interval only for a degree of 7 and on the border of the intervall for degrees 0,1,6 and 8.

```
indeg_precentage
```

```
## [1] 0.002
```

```
outdeg_precentage
```

```
## [1] 0.015
```

We use the Mahalanobis distance to asses the GOF as proposed in 'Goodness of fit for stochastic actor-oriented models' by Josh Lospinoso et. al. The p-values here address the hypothesis that the SAOM with the parameters -1.3178 for outdegree and 1.7292 for the reciprocity generated the observed network against the hypothesis that some other model generated it. As the p-values are both below the 5% significance level and very small (0.015 out-degree and 0.002 in-degree) the fit is thought to be poor (see Josh Lospinoso et. al.). This can also partly be seen in the violin plots as there are a number of degrees in both plots where the observations is outside or just on the 95% confidence intevall.