

Task 2 analysis

Marcin Stopka, Franciszek Sobała, Grzegorz Dolanowski

2026-01-10

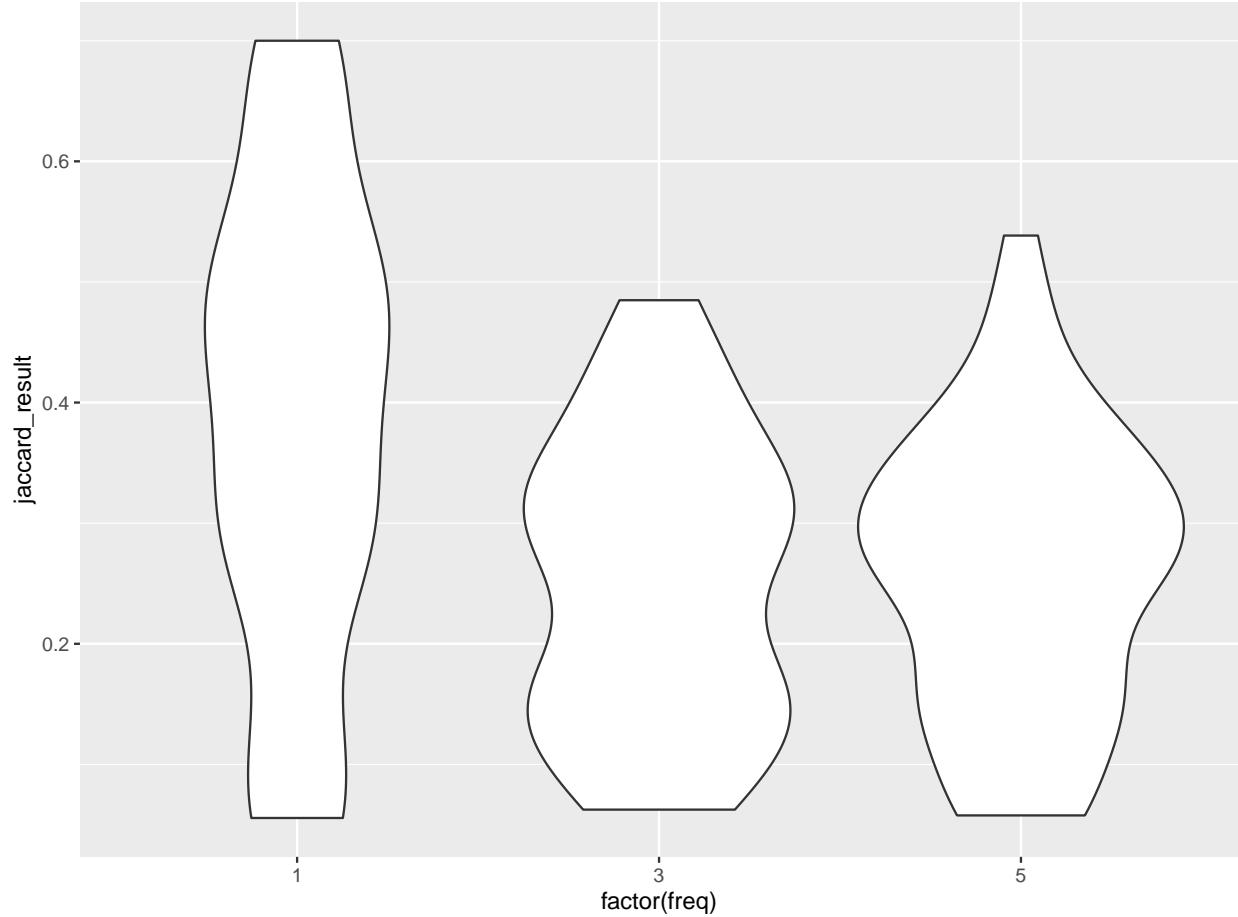
For the second task we've decided to analyze the performance of inference algorithms on 3 validated boolean networks:

1. 8 variable: CARDIAC-DEVELOPMENT
2. 13 variable: METABOLIC-INTERACTIONS-IN-GUT-MICROBIOME
3. 14 variable: NEUROTRANSMITTER-SIGNALING-PATHWAY

We've run exactly the same simulations, including both synchronous and asynchronous manners. In conclusion the results are almost perfectly consistent with our prior findings.

[Consistent] Sampling frequency has negative correlation with result

```
ggplot(data, aes(x=factor(freq), y=jaccard_result)) +
  geom_violin()
```



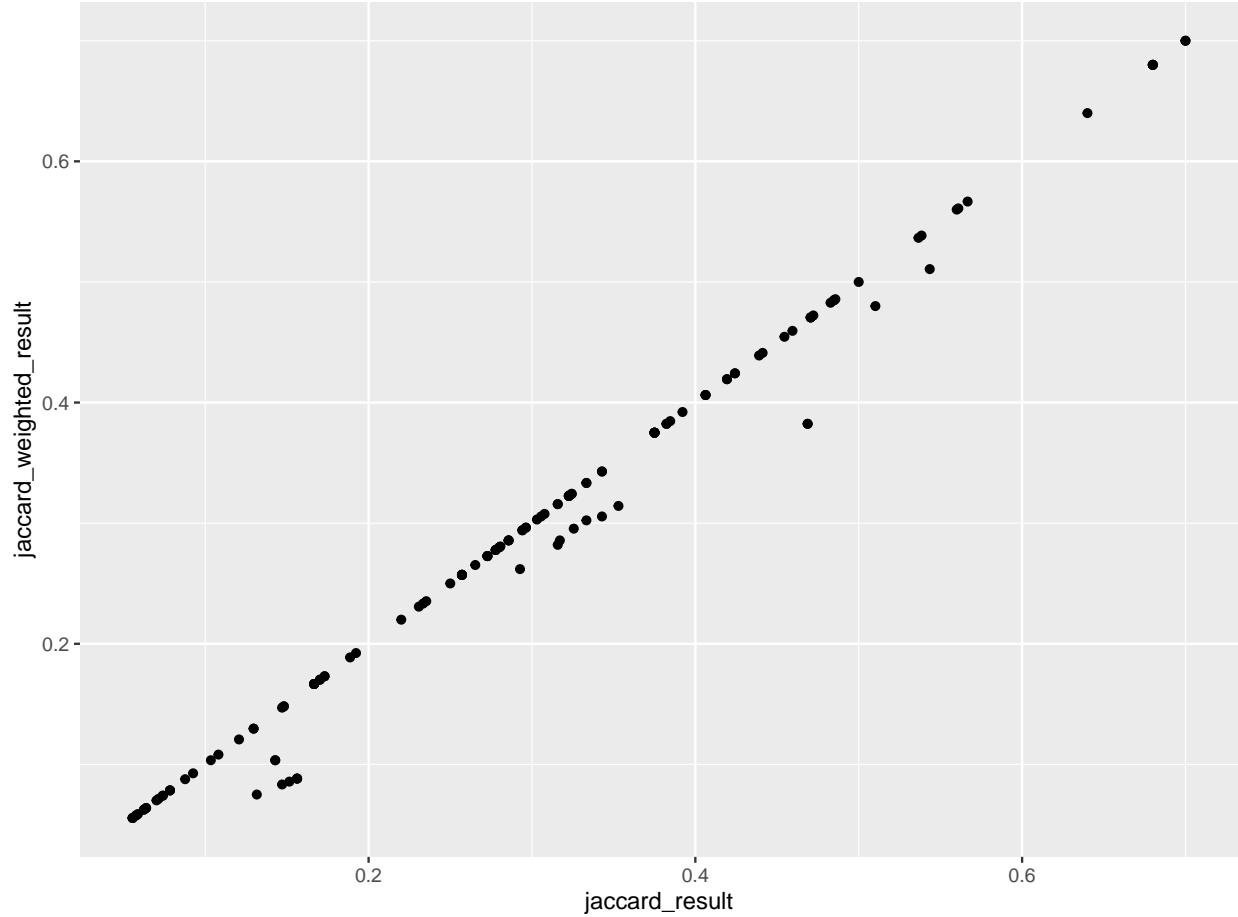
```
tst <- cor.test(data$freq, data$jaccard_result, method = "spearman")

## Warning in cor.test.default(data$freq, data$jaccard_result, method =
## "spearman"): Cannot compute exact p-value with ties
tst

##
## Spearman's rank correlation rho
##
## data: data$freq and data$jaccard_result
## S = 637873, p-value = 0.0006223
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
## -0.2817956
```

Graph Metrics

```
ggplot(data, aes(x=jaccard_result, y=jaccard_weighted_result)) + geom_point()
```



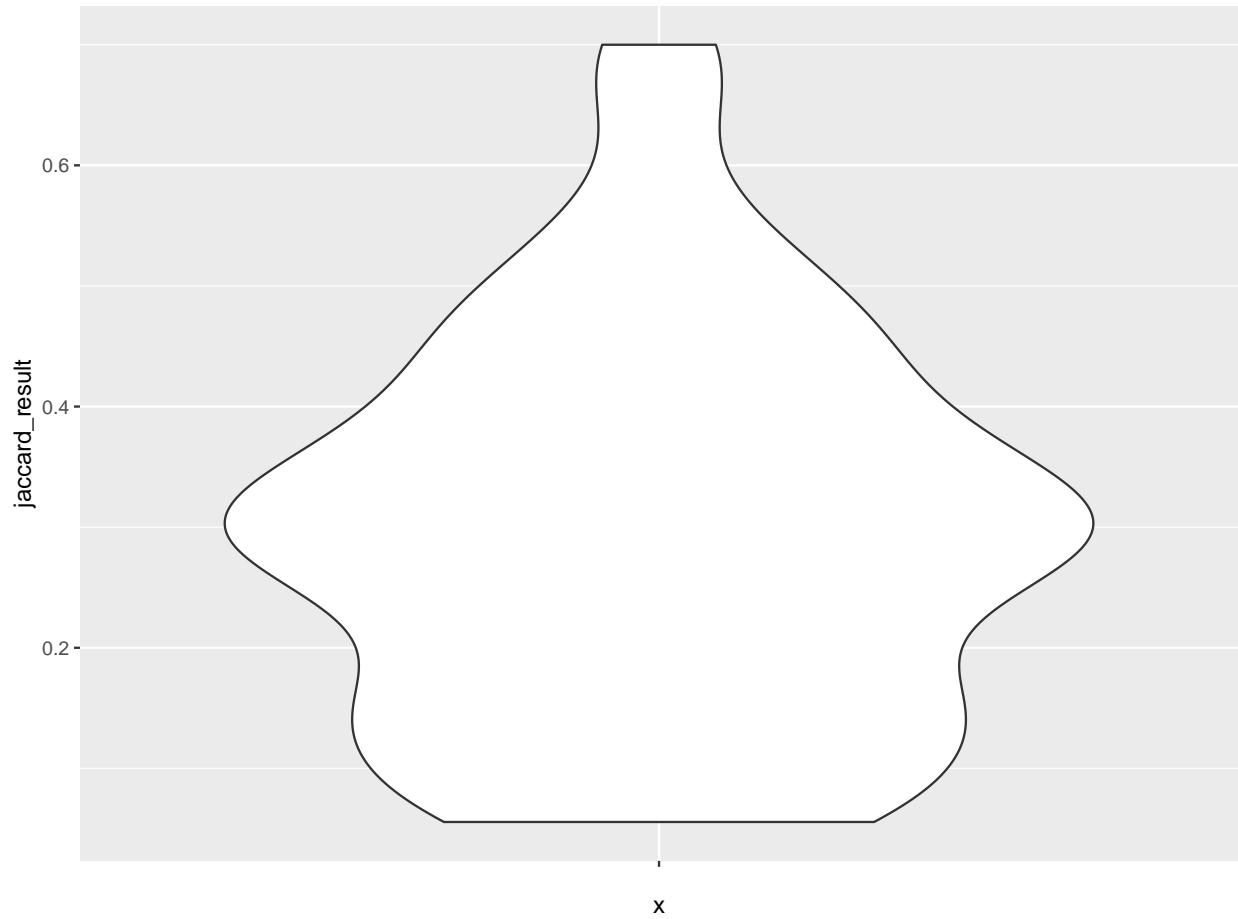
Achieved results

For validated networks the distribution of jaccard metric has similar properties.

```
kable(data.frame(rbind(summary(data$jaccard_result))))
```

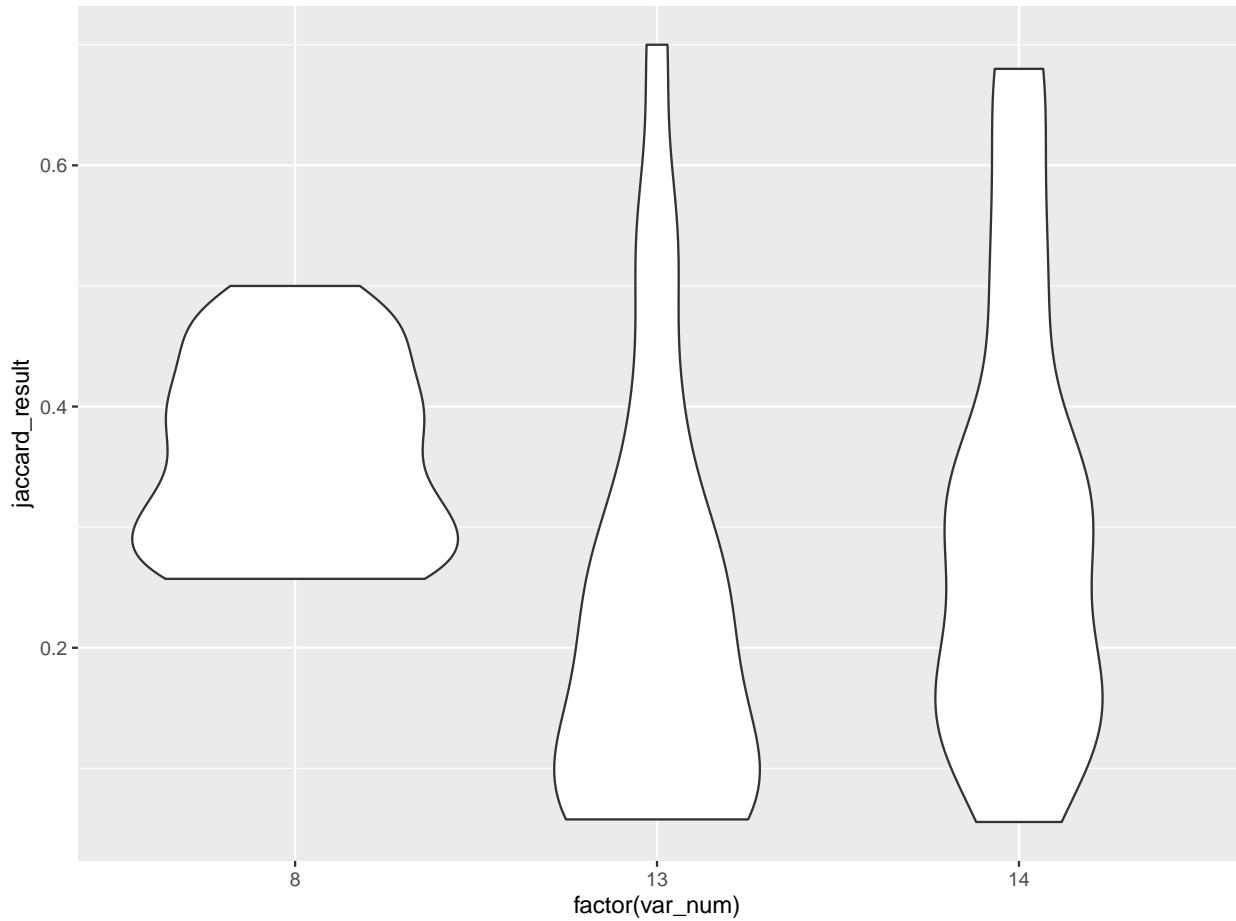
Min.	X1st.Qu.	Median	Mean	X3rd.Qu.	Max.
0.05555556	0.16666667	0.2934003	0.2984055	0.3956801	0.7

```
ggplot(data, aes(x="", y=jaccard_result)) +
  geom_violin()
```



[Consistent] Number of variables is negatively correlated with the jaccard metric

```
ggplot(data, aes(x=factor(var_num), y=jaccard_result)) +  
  geom_violin()
```



```
tst <- cor.test(data$var_num, data$jaccard_result, method = "spearman")

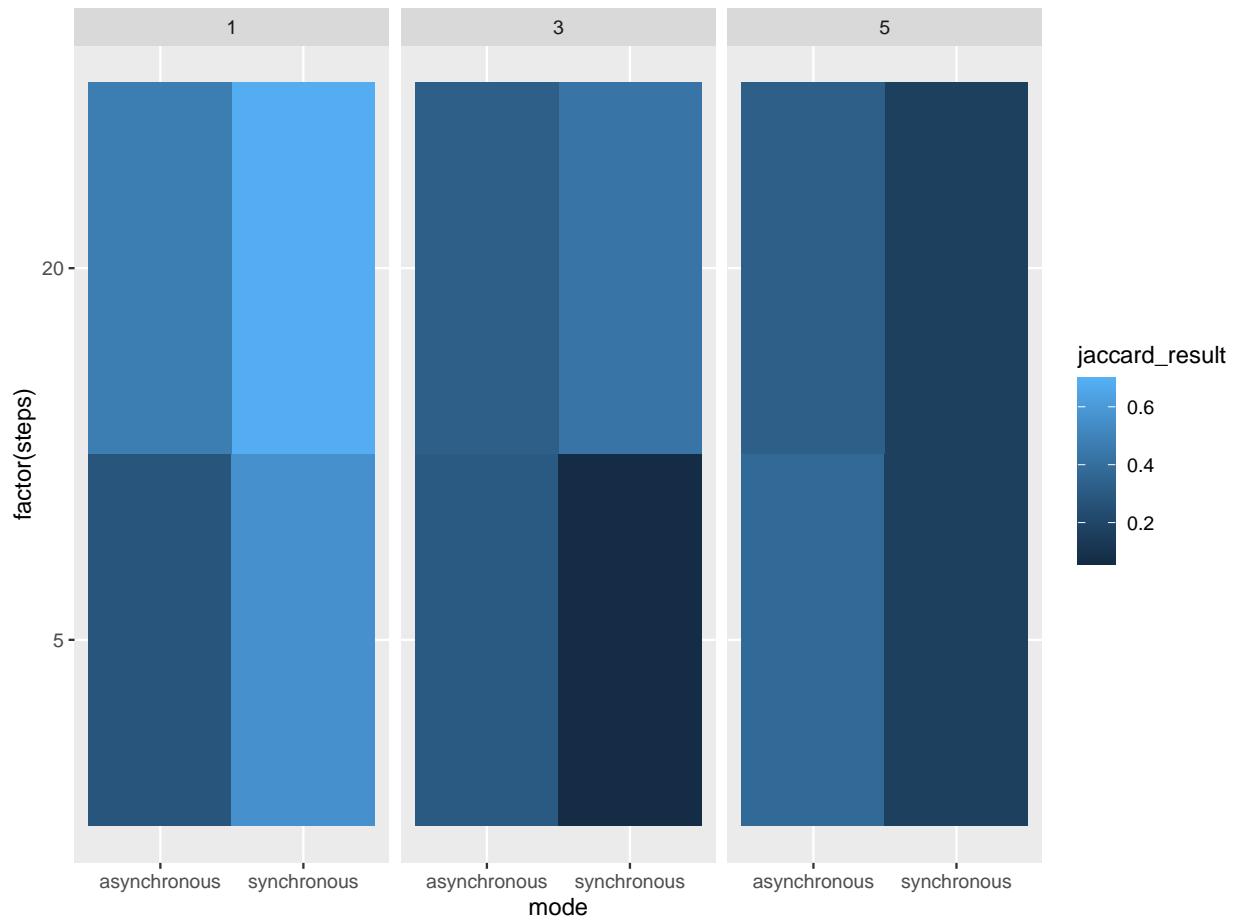
## Warning in cor.test.default(data$var_num, data$jaccard_result, method =
## "spearman"): Cannot compute exact p-value with ties
tst

##
## Spearman's rank correlation rho
##
## data: data$var_num and data$jaccard_result
## S = 594387, p-value = 0.01955
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.1944124
```

[Consistent] Asynchronous mode is “easier”, except for the case of frequency=1

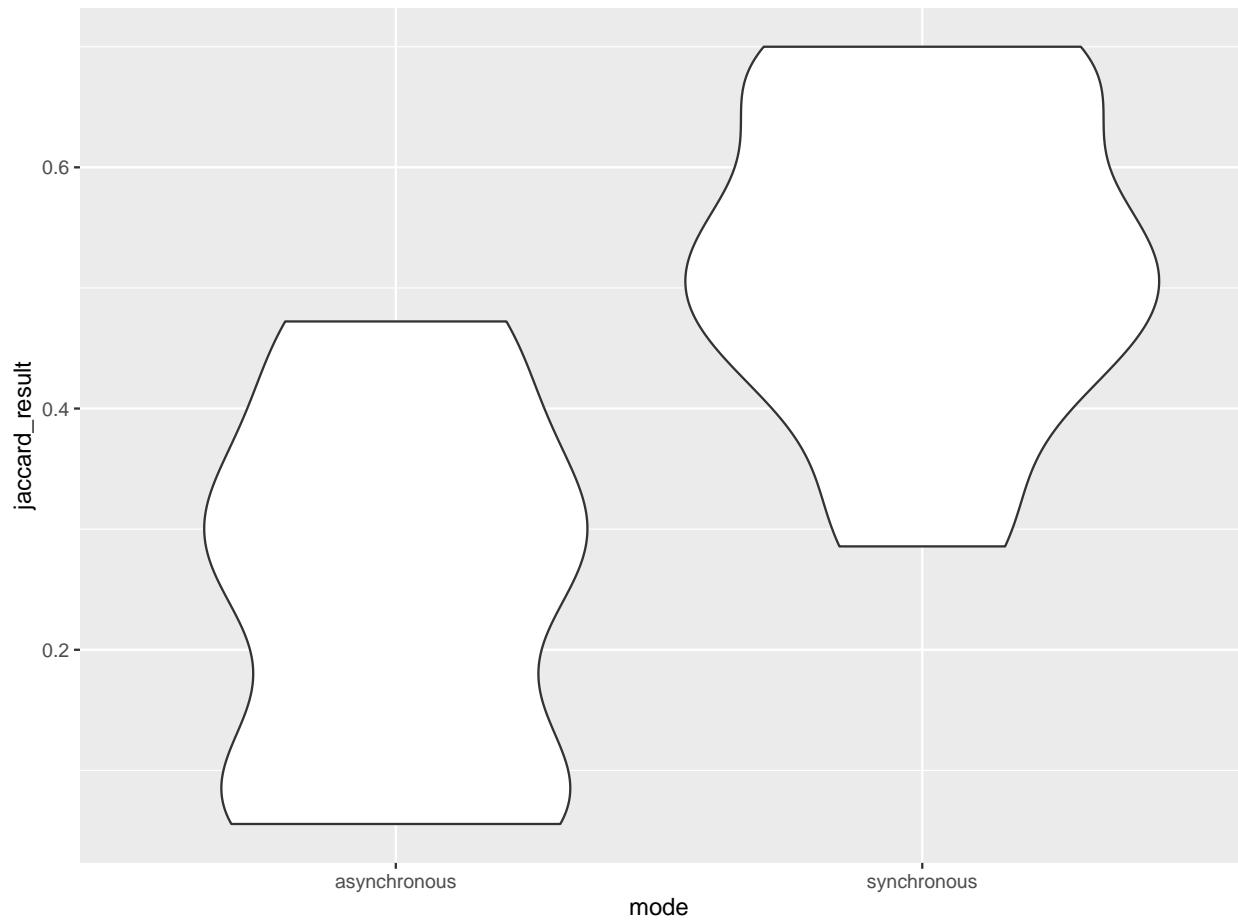
First we can see that the synchronous data is much less resistant to infrequent probes.

```
ggplot(data, aes(x=mode,y=factor(steps),fill=jaccard_result)) + geom_tile() + facet_wrap(~freq)
```

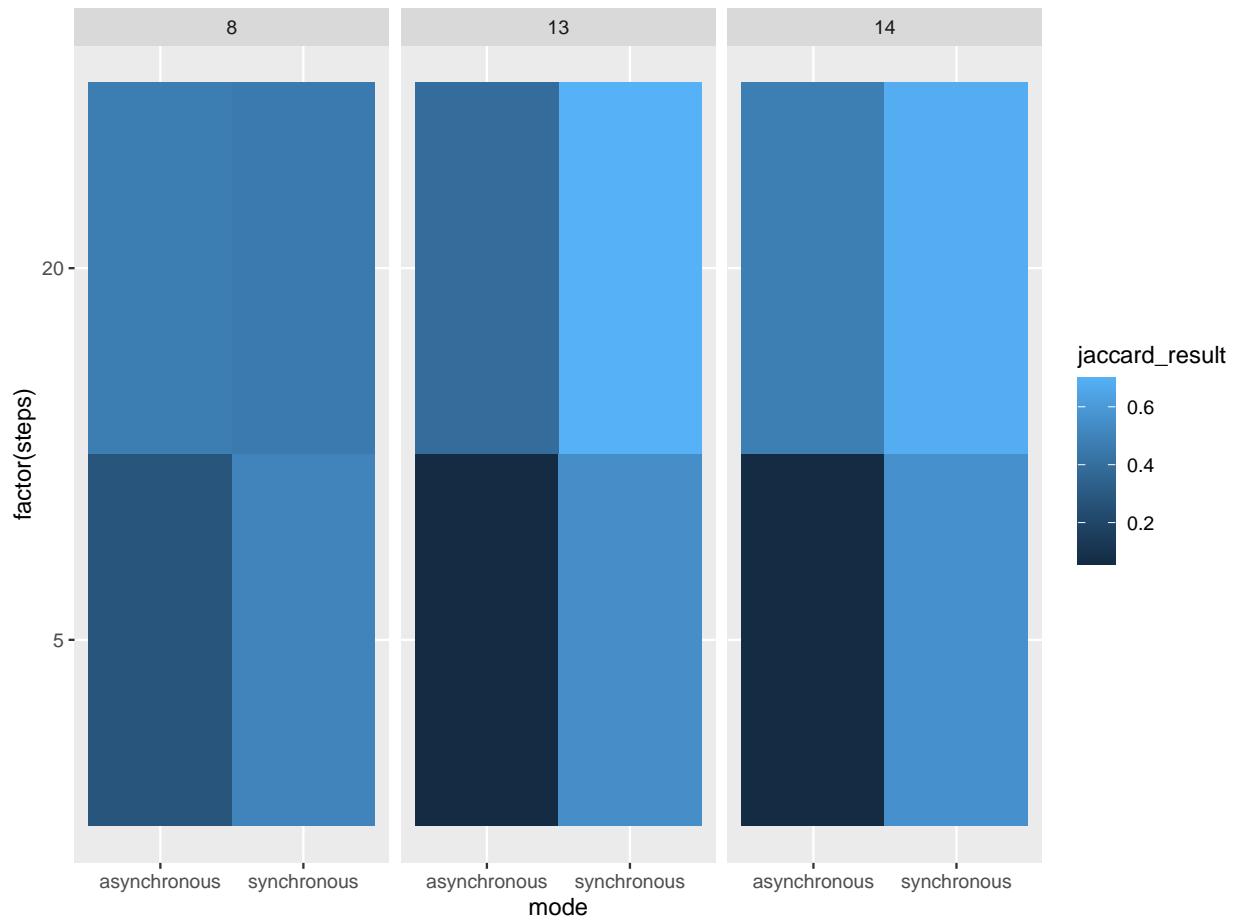


[Consistent] For frequency=1, synchronous mode consistently achieves higher score, for every variable count

```
ggplot(data %>% filter(freq==1), aes(x=mode, y=jaccard_result)) +
  geom_violin()
```

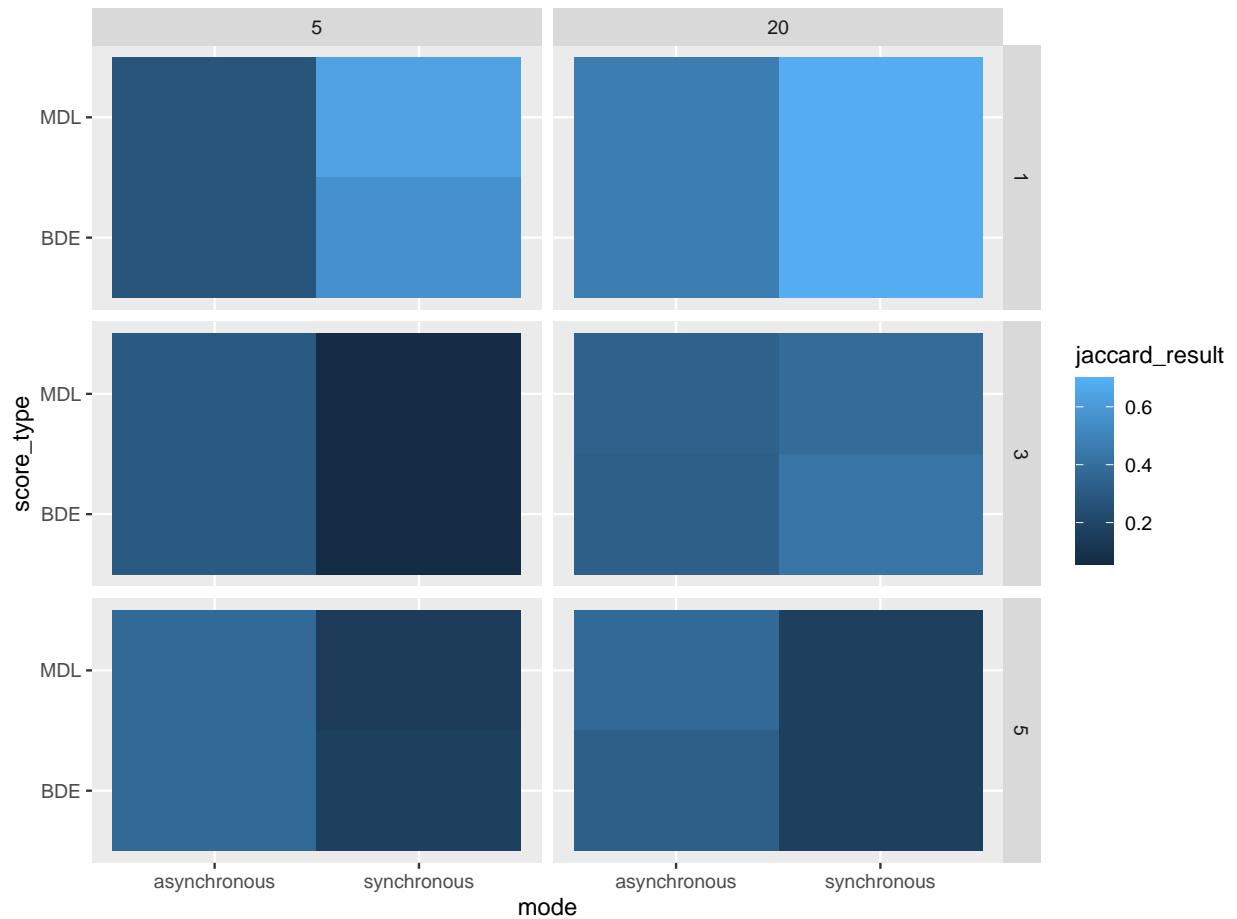


```
ggplot(data %>% filter(freq==1), aes(x=mode,y=factor(steps),fill=jaccard_result)) + geom_tile() +  
  facet_wrap(~var_num)
```



[Consistent] Scoring mode has no statistically significant effect on result.

```
ggplot(data, aes(x=mode,y=score_type,fill=jaccard_result)) +
  geom_tile() +
  facet_grid(freq~steps)
```



```
ks.test((data %>% filter(score_type=='MDL'))$jaccard_result,
       (data %>% filter(score_type=='BDE'))$jaccard_result)
```

```
##
##  Exact two-sample Kolmogorov-Smirnov test
##
## data:  (data %>% filter(score_type == "MDL"))$jaccard_result and (data %>% filter(score_type == "BDE"))
## D = 0.041667, p-value = 1
## alternative hypothesis: two-sided
```

[Inconsistent] Correlation of attractor frequency when using synchronous mode is statistically insignificant.

```
data_synch <- data %>% filter(mode=="synchronous")
tst <- cor.test(data_synch$attper, data_synch$jaccard_result, method = "spearman")

## Warning in cor.test.default(data_synch$attper, data_synch$jaccard_result, :
## Cannot compute exact p-value with ties
tst

##
##  Spearman's rank correlation rho
##
```

```

## data: data_synch$attper and data_synch$jaccard_result
## S = 70310, p-value = 0.2747
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## -0.1304561

```

[Consistent] For asynchronous mode, attractor frequency is positively correlated with the score.

```

data_asynch <- data %>% filter(mode=="asynchronous")
tst <- cor.test(data_asynch$attper, data_asynch$jaccard_result, method = "spearman")

## Warning in cor.test.default(data_asynch$attper, data_asynch$jaccard_result, :
## Cannot compute exact p-value with ties
tst

##
## Spearman's rank correlation rho
##
## data: data_asynch$attper and data_asynch$jaccard_result
## S = 29812, p-value = 2.751e-06
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.5206826

```

Linear regression

If we run linear regression with all the data, it doesn't work the best, gives adjusted R squared of 0.1298414. On the other hand, if we filter only by those results that have frequency 1, we get a high R squared.

```

data_with_dummies <- dummy_cols(data) %>% filter(freq==1)
linear_model <- lm(jaccard_result ~
                     var_num + steps + numtraj + attper + mode_synchronous + score_type_BDE,
                     data_with_dummies)

kable(tidy(linear_model))

```

term	estimate	std.error	statistic	p.value
(Intercept)	0.0072878	0.1501476	0.0485379	0.9615233
var_num	0.0035245	0.0107933	0.3265502	0.7456693
steps	0.0115109	0.0052642	2.1866489	0.0345301
numtraj	0.0017841	0.0006555	2.7218770	0.0094861
attper	-0.0542584	0.1725653	-0.3144223	0.7547948
mode_synchronous	0.2840250	0.0446732	6.3578359	0.0000001
score_type_BDE	-0.0024669	0.0326757	-0.0754976	0.9401860

And with data filtered by frequency equal to 1, we get adjusted R squared of 0.6674927.