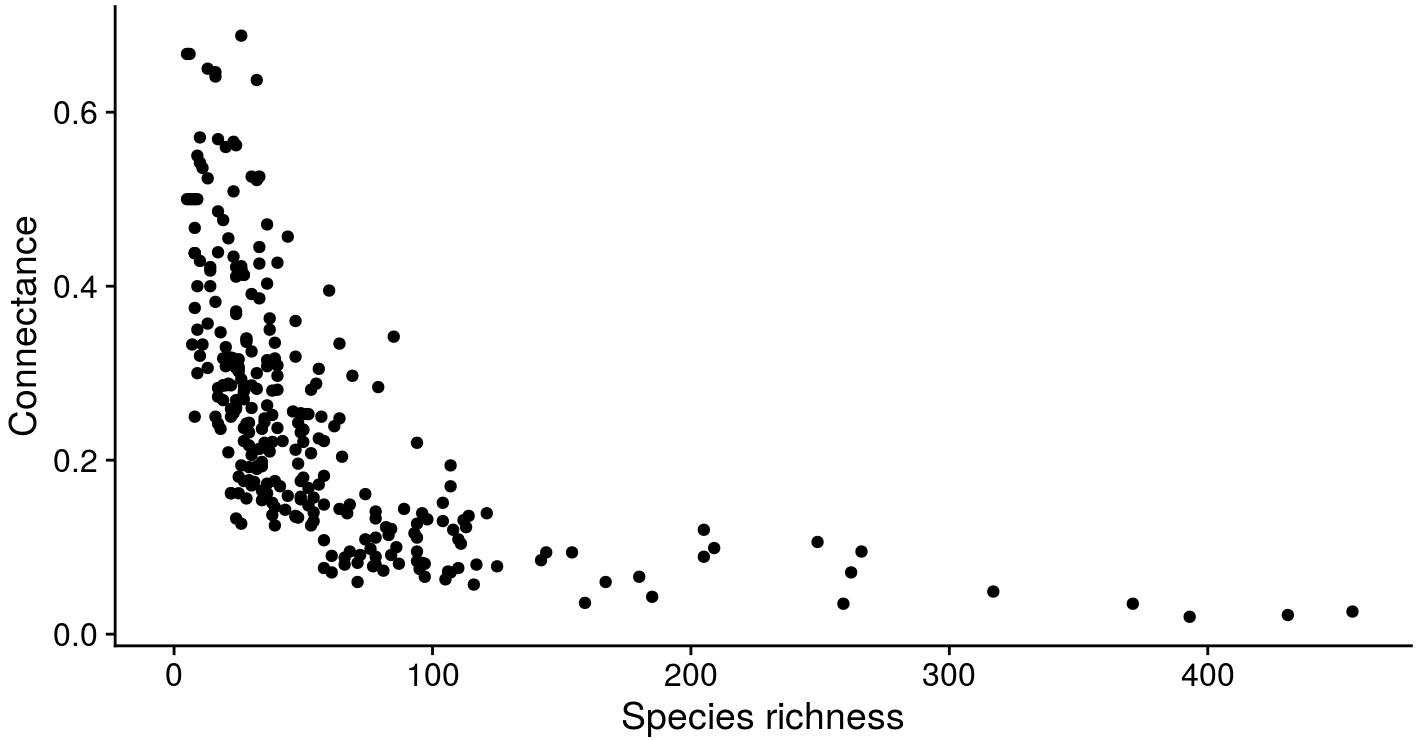
Statistical modeling of ecological networks

Matt Barbour

3/31/2021

What is a statistical model?

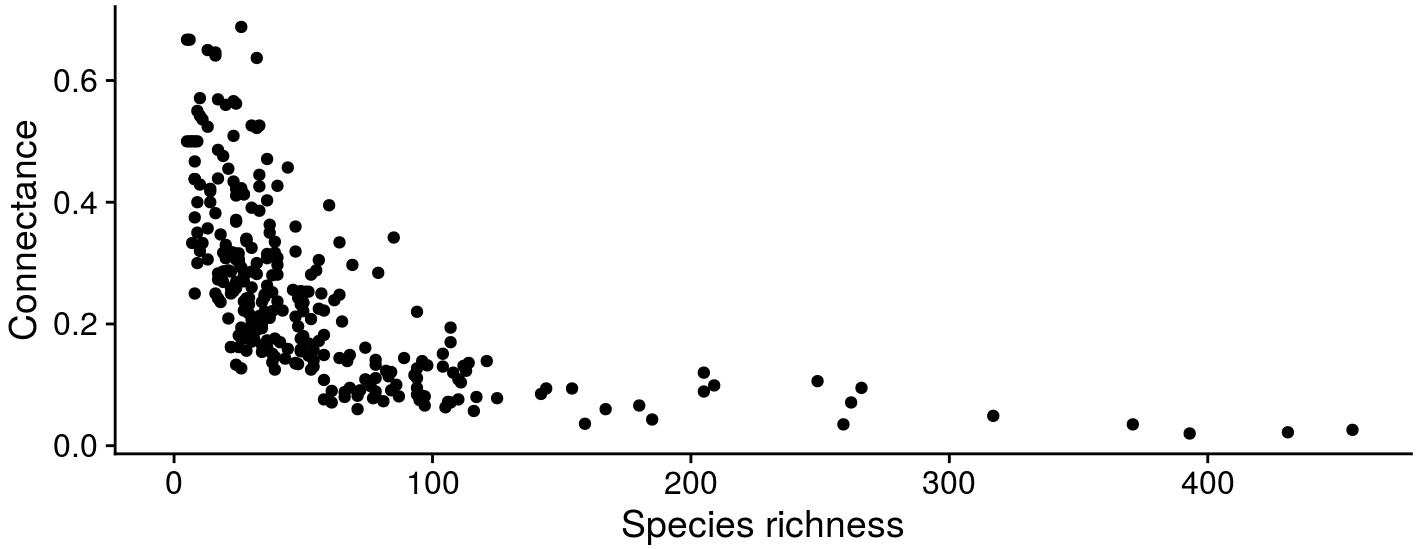
- Describes the mathematical relationships between some explanatory/predictor variables and a response variable
- Seek a parsimonious model to explain the data



$$y = mx + b$$

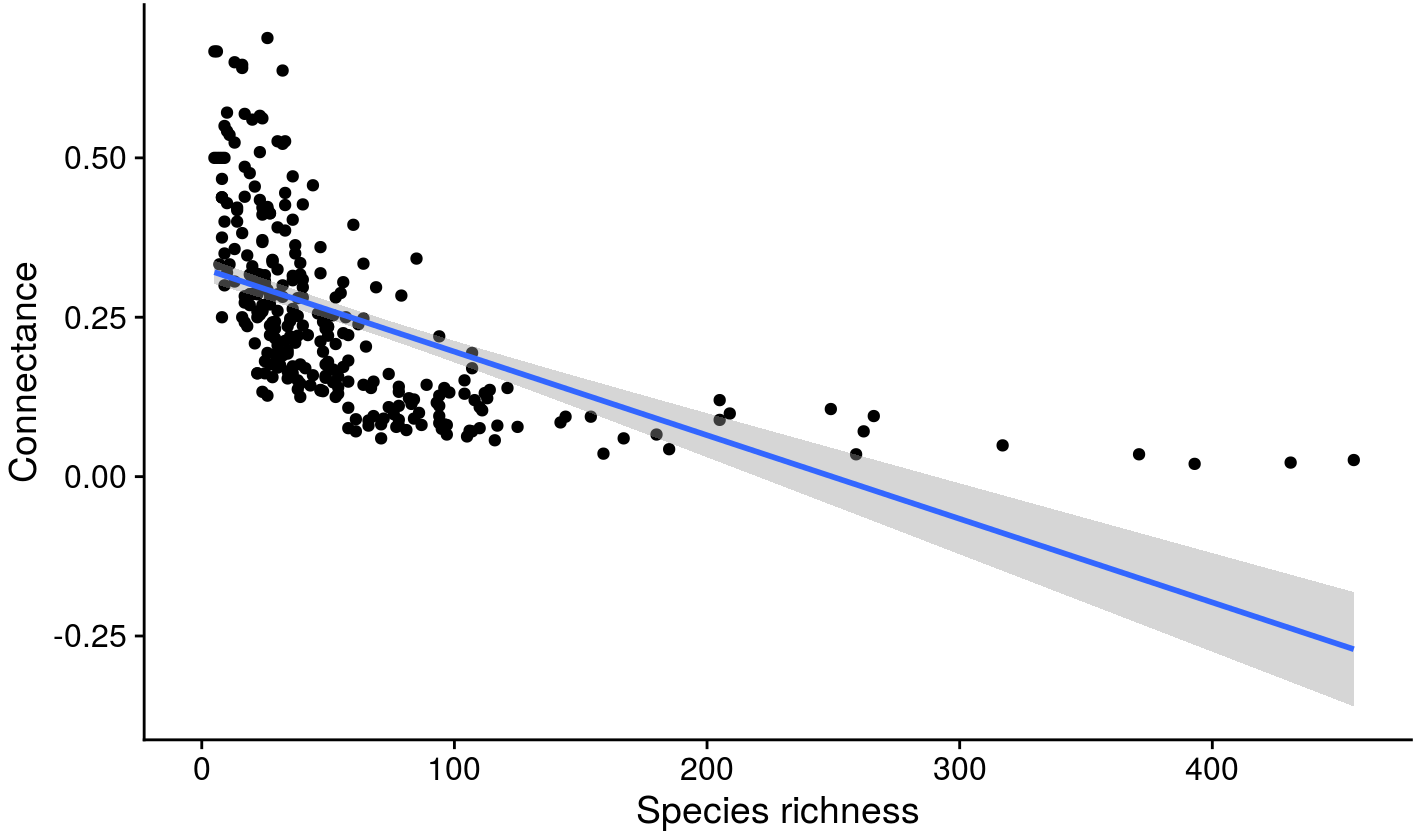
- y = Connectance
- x = Species richness
- Estimate m and b given certain assumptions (e.g. errors are normally distributed)

```
lm(Connectance ~ Species richness + 1) # R notation
```



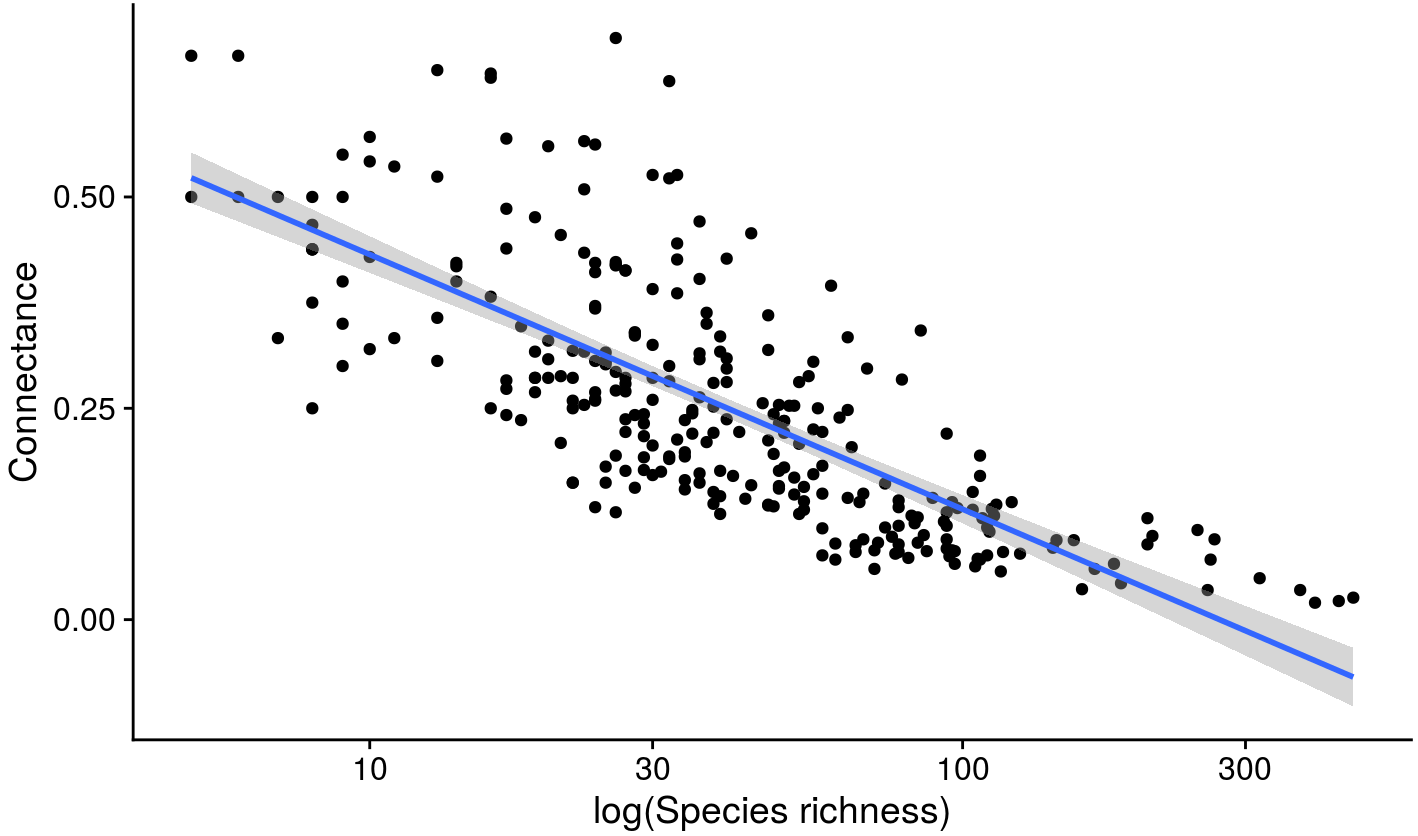
```
lm(Connectance ~ Species richness + 1)
```

- 32% of the variance explained



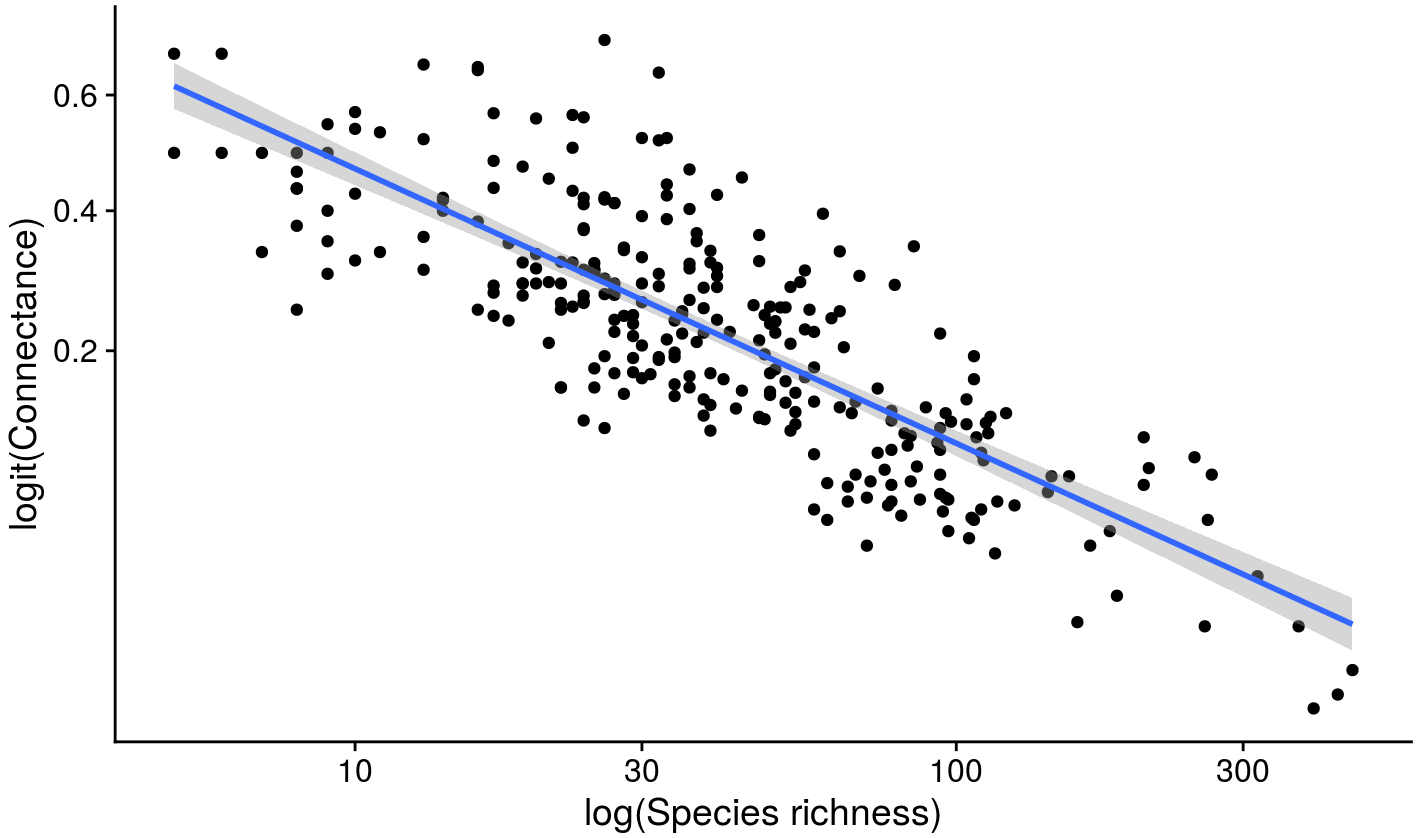
```
lm(Connectance ~ log(Species richness) + 1)
```

- 56% of the variance explained



```
lm(logit(Connectance) ~ log(Species richness) + 1)
```

- 65% of the variance explained



Why use statistical modeling for networks?

- Predict how interaction networks will change over space and time
- Understand the factors that determine species interactions and network structure
- Account for sampling uncertainty

Two different approaches

Predict change in network properties (e.g. connectance, nestedness, etc.)

- Straightforward with standard regression techniques
- Need to standardize sampling effort first

Predict change in species interactions and network properties

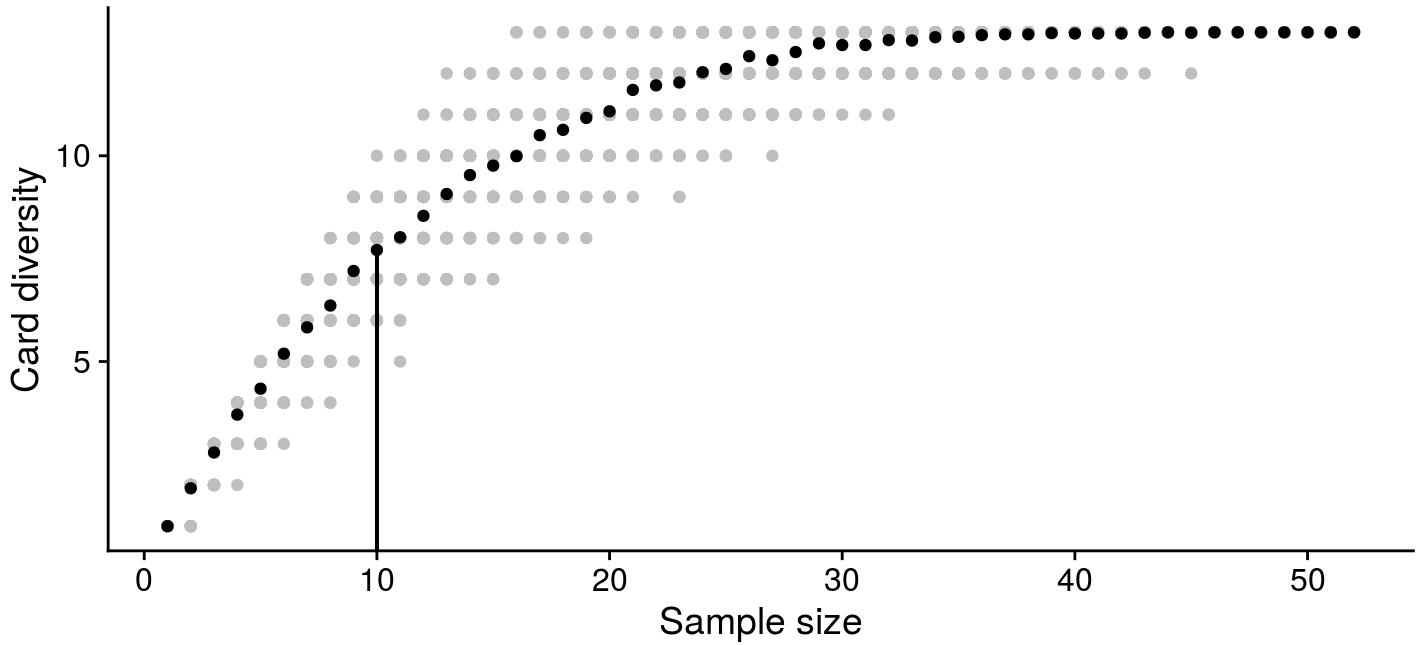
- Require more and better data
- Often require advanced approaches (e.g. mixed-effects, Bayesian)
- Better insight to **why** network properties change

Standardizing sampling effort

Time for a card trick!

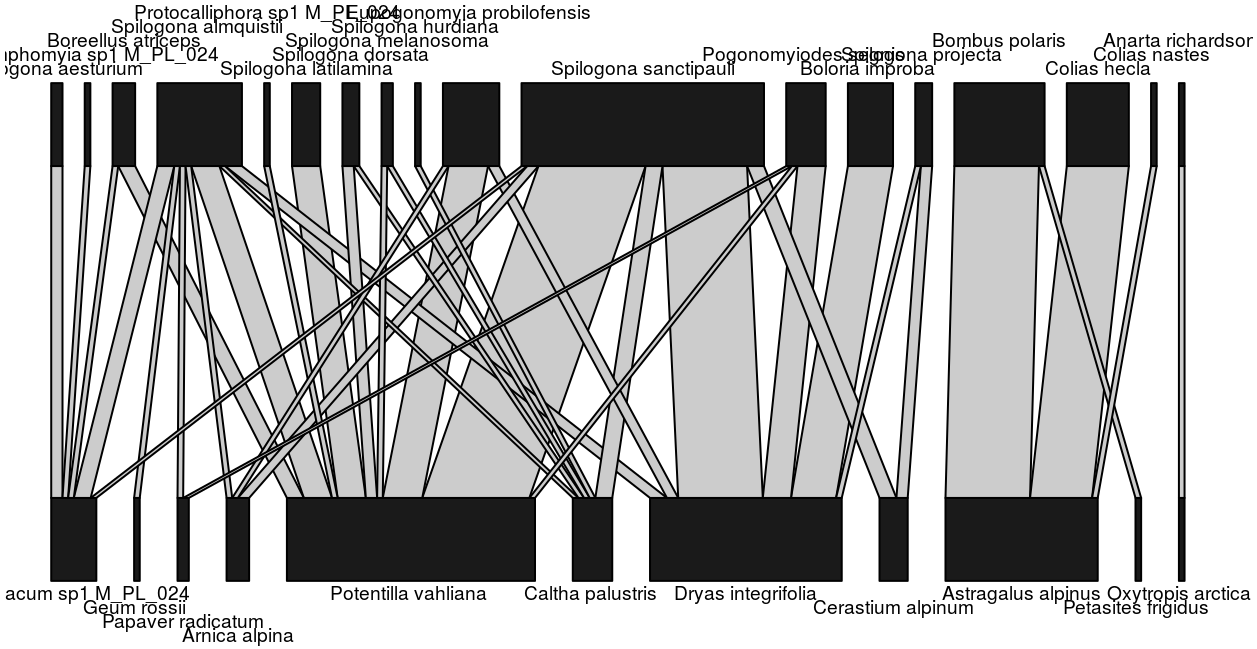
Rarefaction

- Randomly sample **x** items from a population and compute the aggregate property **y**
- Repeat many times to estimate average and uncertainty (e.g. standard deviation)

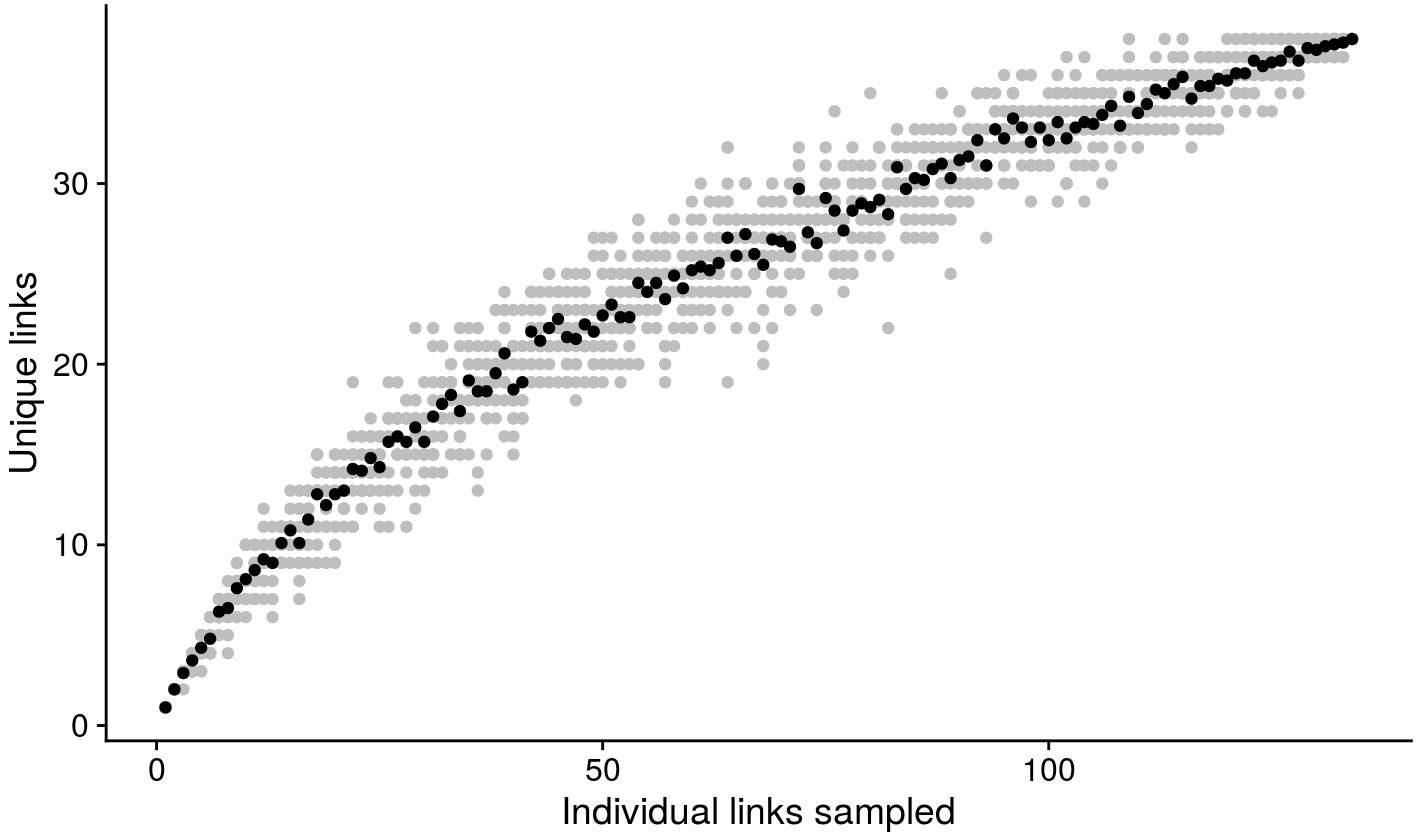


Example with a real network

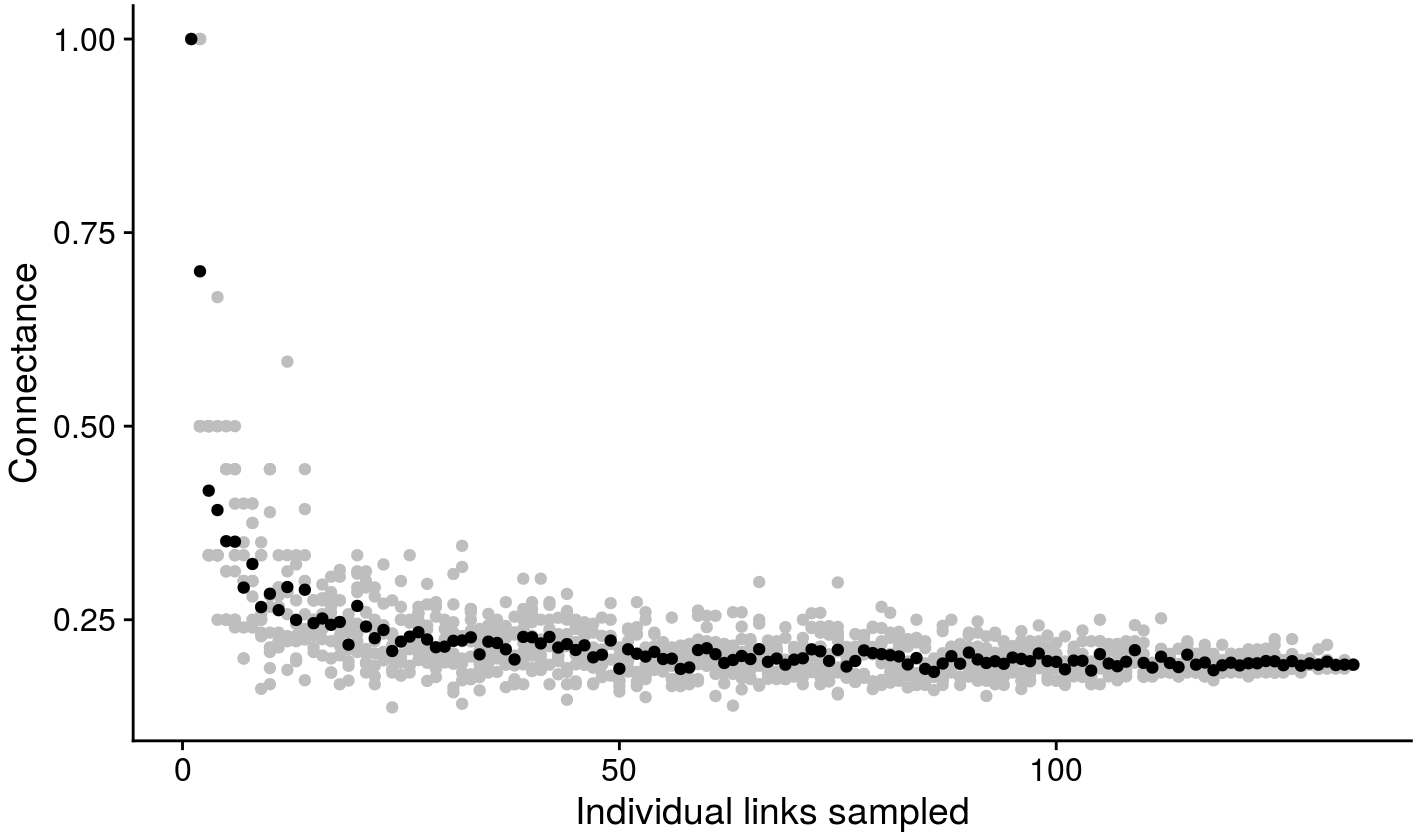
- Number of unique links = 38
- Number of individual links = 134



- Number of observed links changes with sampling effort

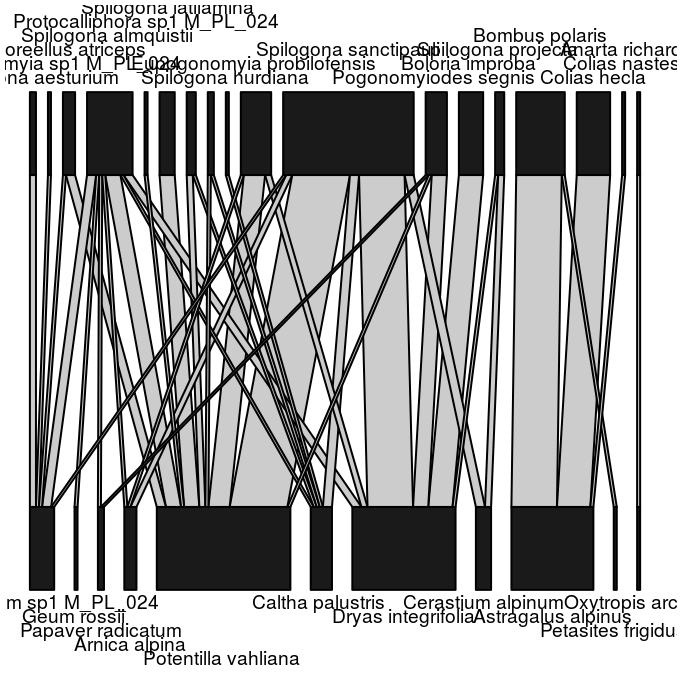


Rarefaction for any network property

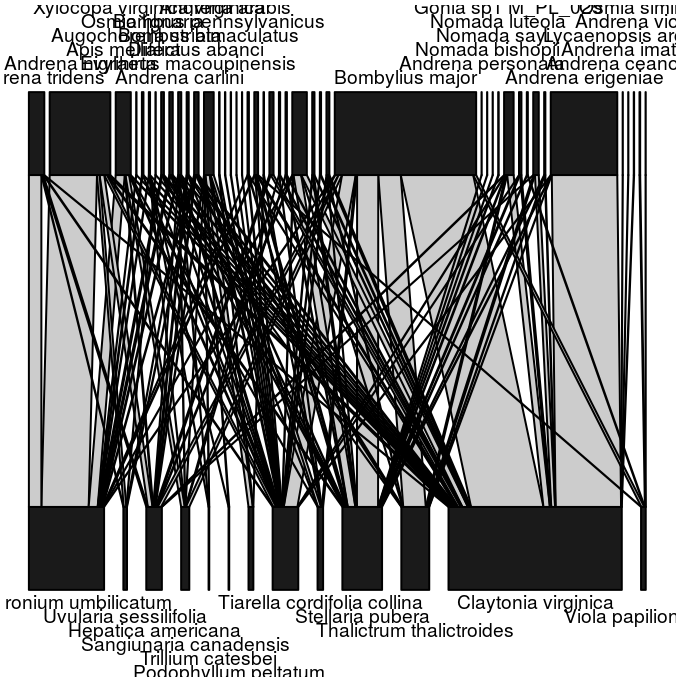


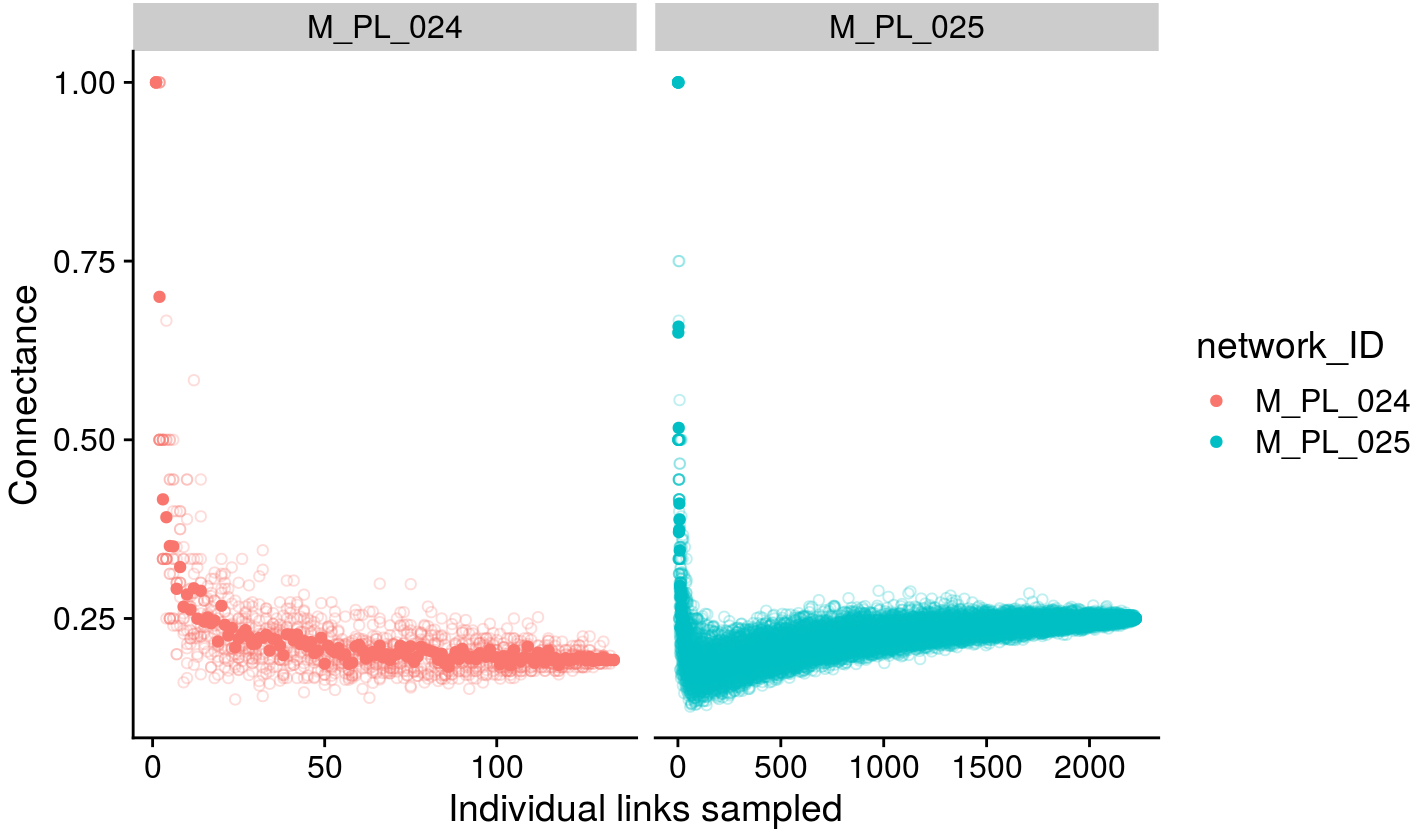
Comparing two networks

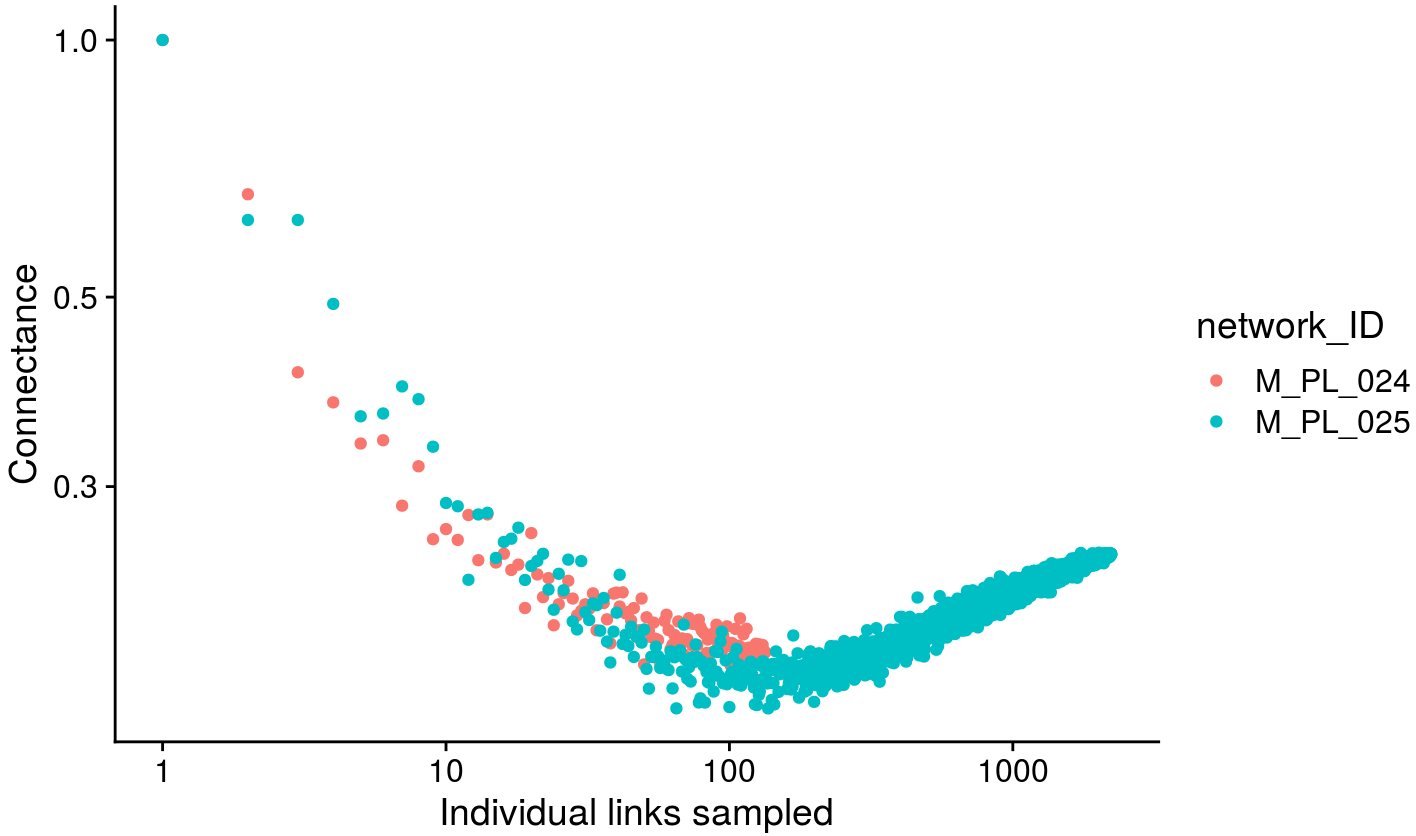
C = 0.19



C = 0.25




M_PL_024
M_PL_025
1.00
0.75
0.50
0.25
Connectance
0
50
100
0
500
1000
1500
2000
Individual links sampled
network_ID
M_PL_024
M_PL_025


1.0
0.5
0.3
Connectance
1
10
100
1000
Individual links sampled
network_ID
M_PL_024
M_PL_025

Rarefaction

- Important first step before building a statistical model
- Solid approach for reliably comparing network properties
- Only possible for weighted networks
- Limited ability to address other interesting questions

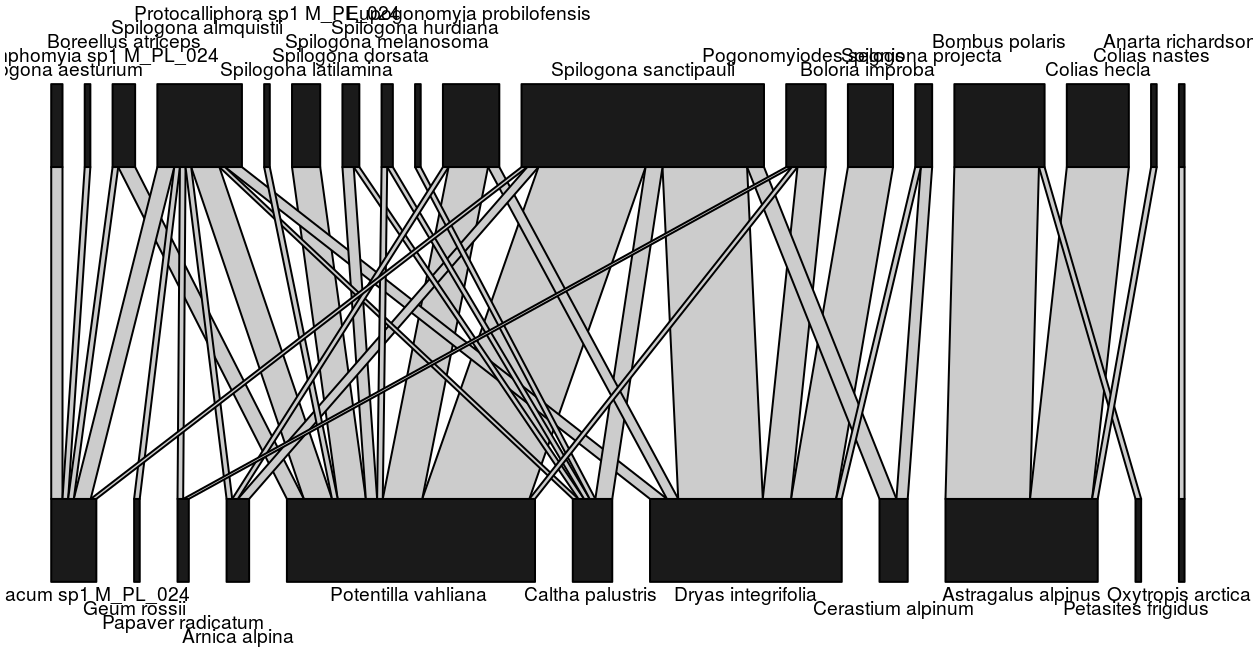
Modeling species interactions

Goal: specify mathematical relationships that predict the occurrence and/or frequency of species interactions

What are some factors that might influence species interactions?

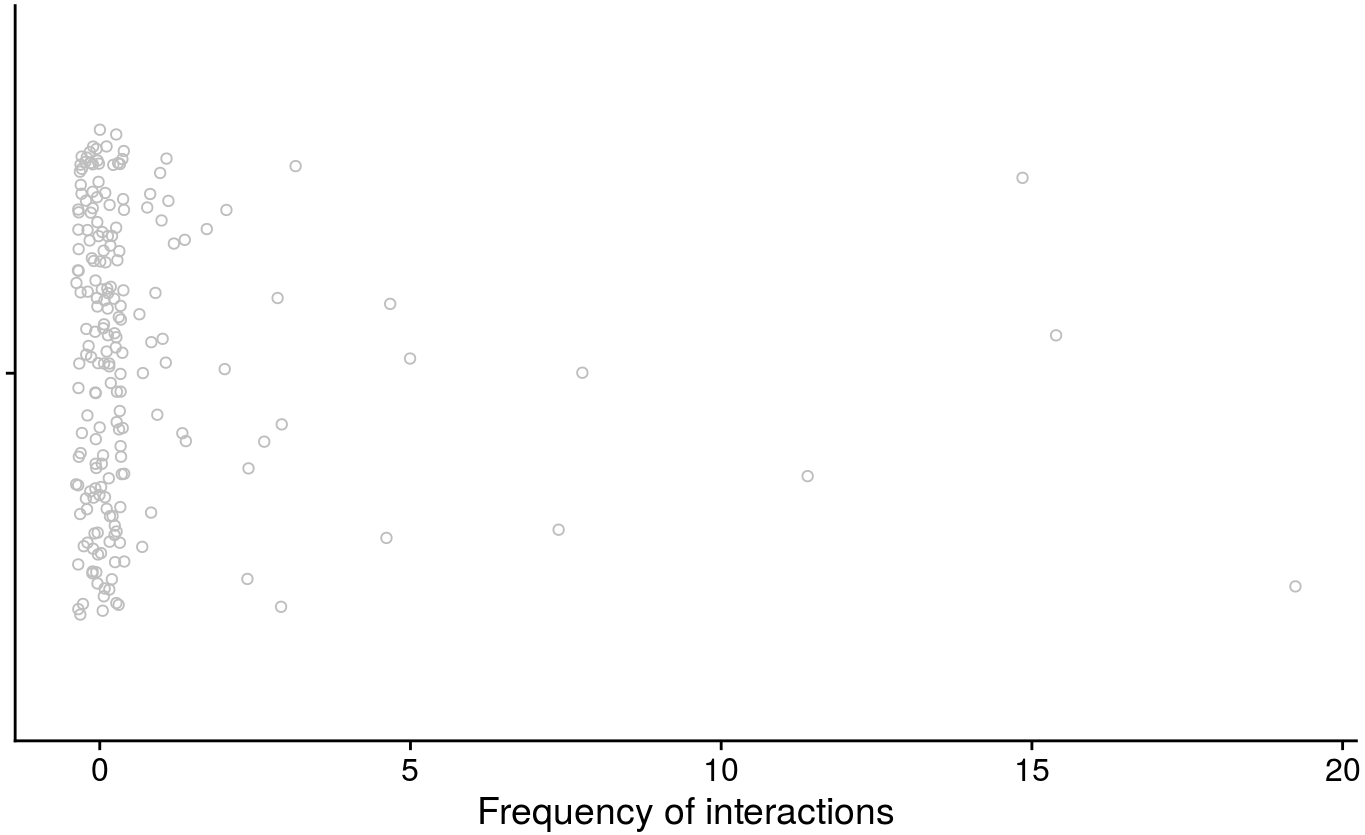
- Abundance
- Traits
- Phylogenetic relationships

Example network



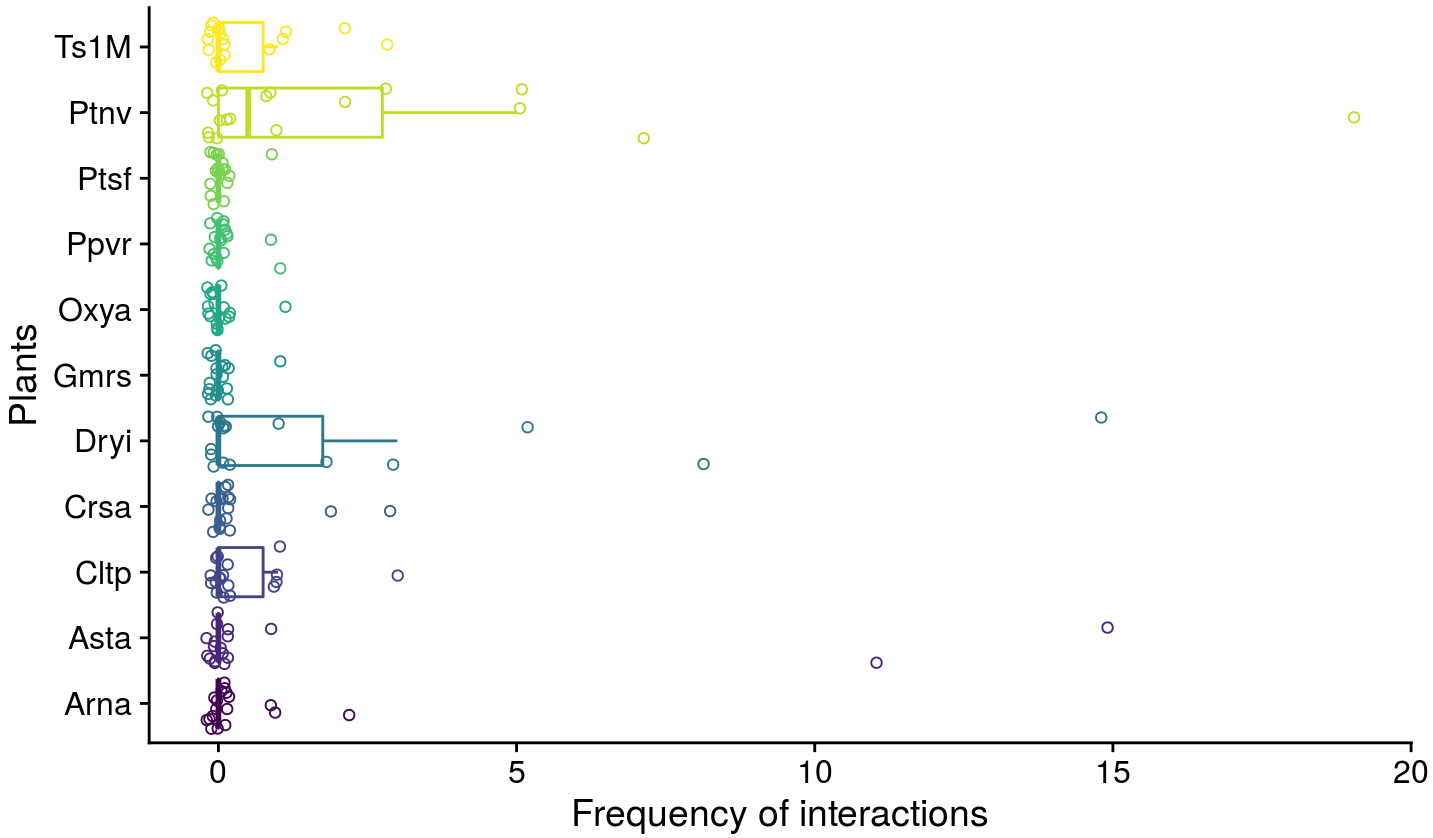
Model variability in interactions

- Start with a linear model



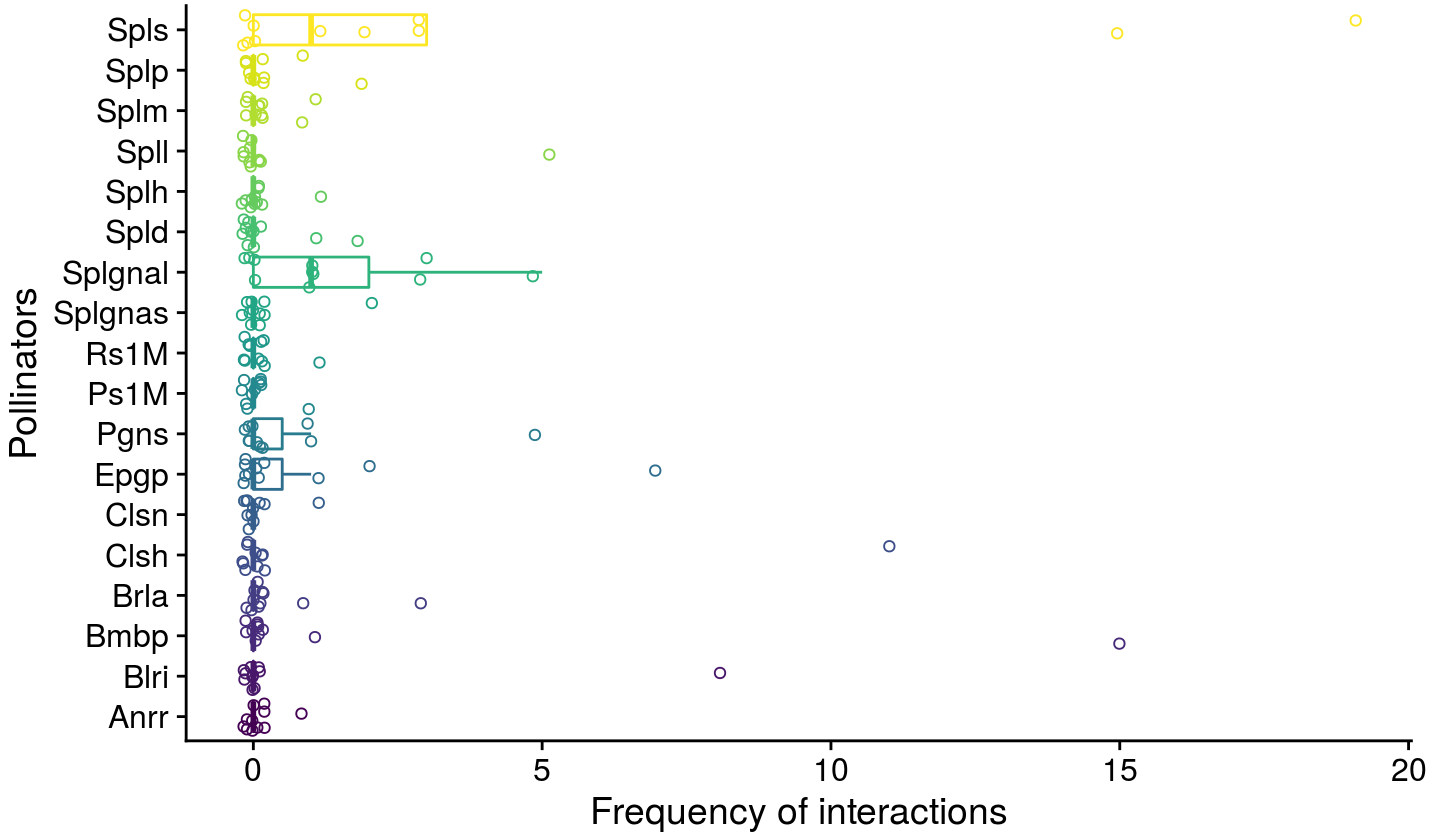
Effect of plant species

- Plant species explains 12% of the variance



Effect of pollinator species

- Pollinator species explains 14% of the variance



Linear model fit

```
lm(frequency ~ plants + pollinators + 1)
```

- Model explains 26% of the variance

```
## Analysis of Variance Table
##
## Response: frequency
##              Df Sum Sq Mean Sq F value   Pr(>F)
## plants       10 130.54 13.0535  2.6907 0.004386 **
## pollinators  17 158.04  9.2965  1.9163 0.019486 *
## Residuals   170 824.74  4.8514
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- What about the interaction between plants and pollinators?

Perfect fit?

```
lm(frequency ~ plants + pollinators + plants:pollinators + 1)
```

- Model explains 100% of the variance…

```
## Analysis of Variance Table
##
## Response: frequency
##                     Df Sum Sq Mean Sq F value Pr(>F)
## plants              10 130.54 13.0535
## pollinators         17 158.04  9.2965
## plants:pollinators 170 824.74  4.8514
## Residuals            0   0.00
```

- Need replicate sampling to estimate the *statistical interaction* between species

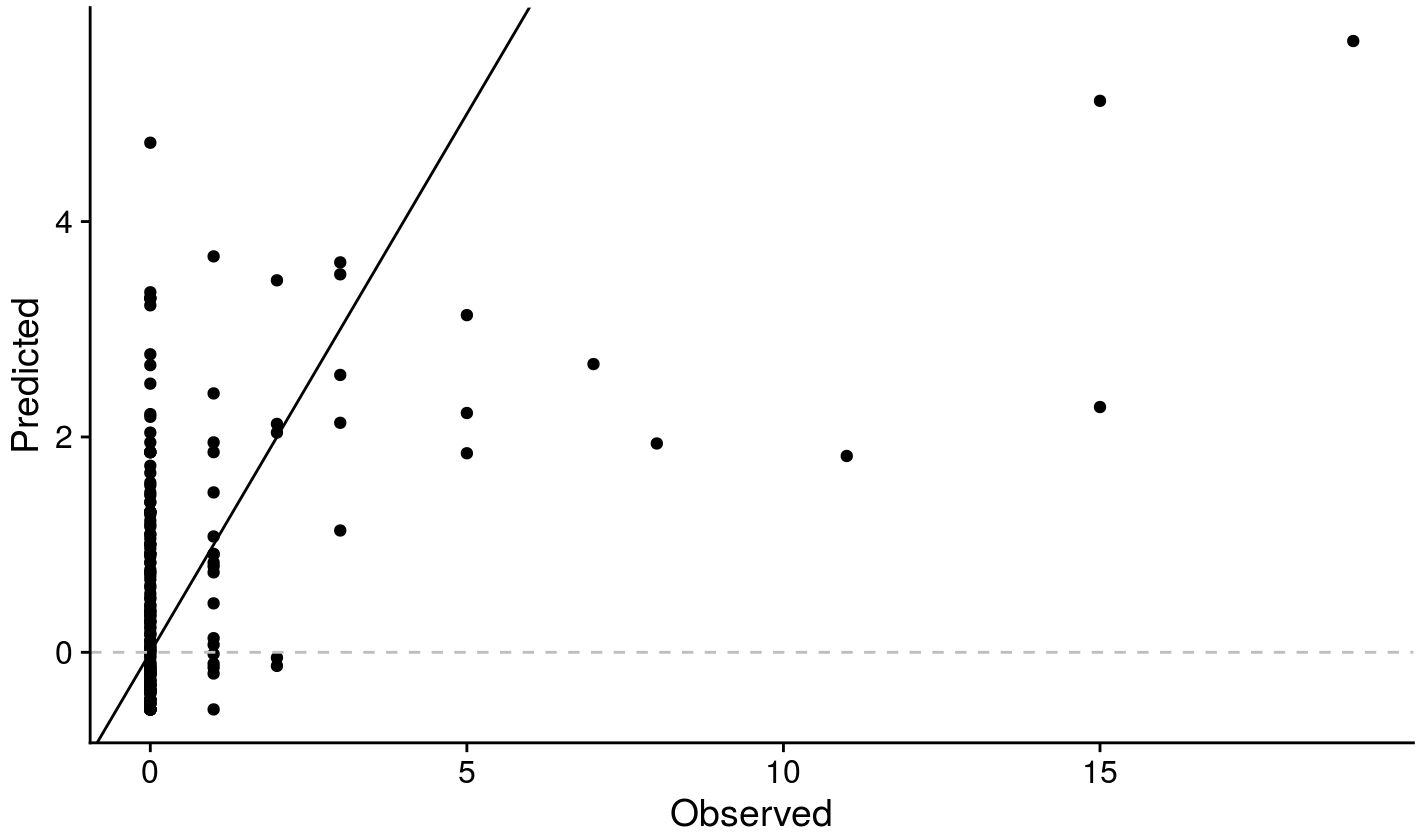
Linear model fit

```
lm(frequency ~ plants + pollinators + 1)
```

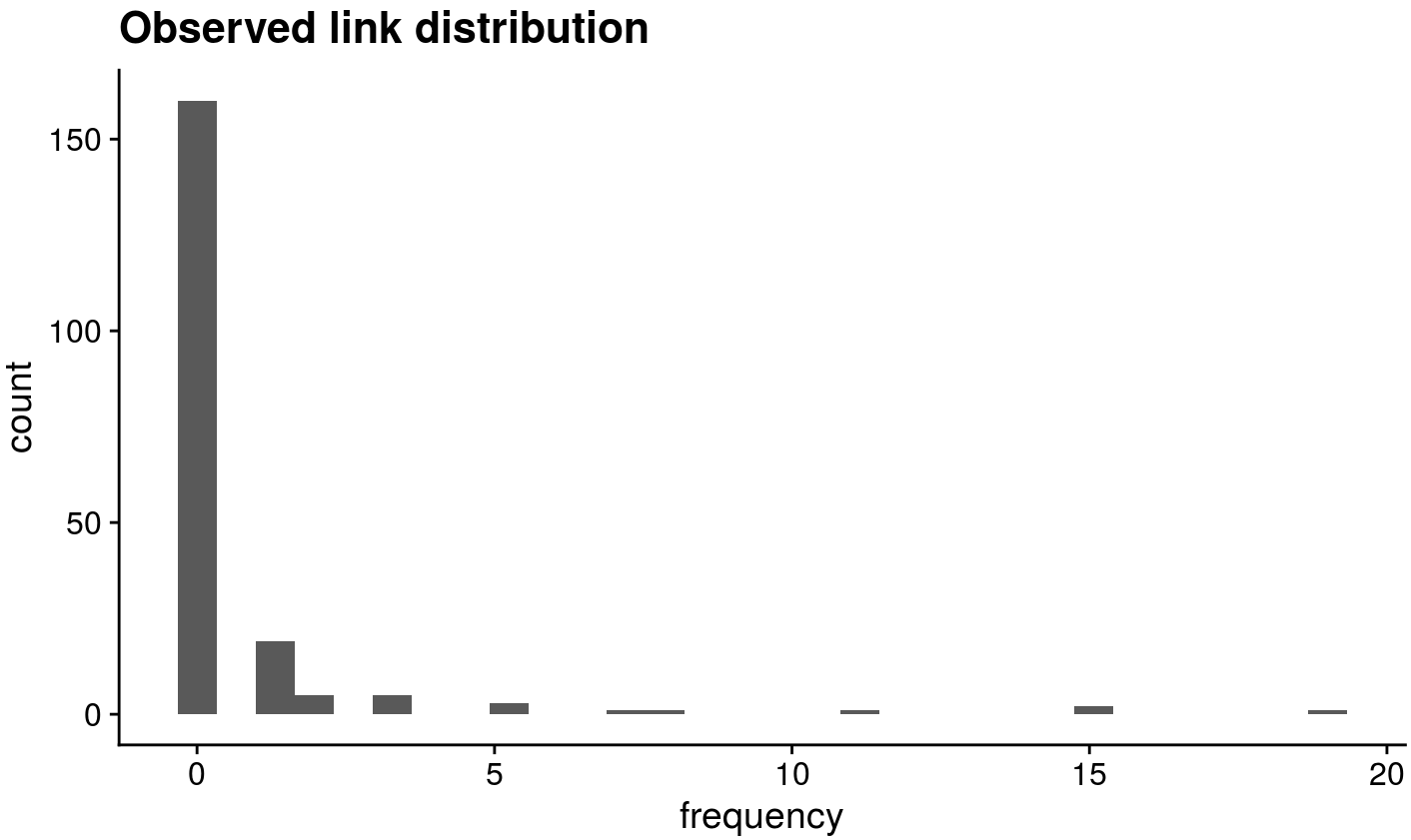
- Model explains 26% of the variance

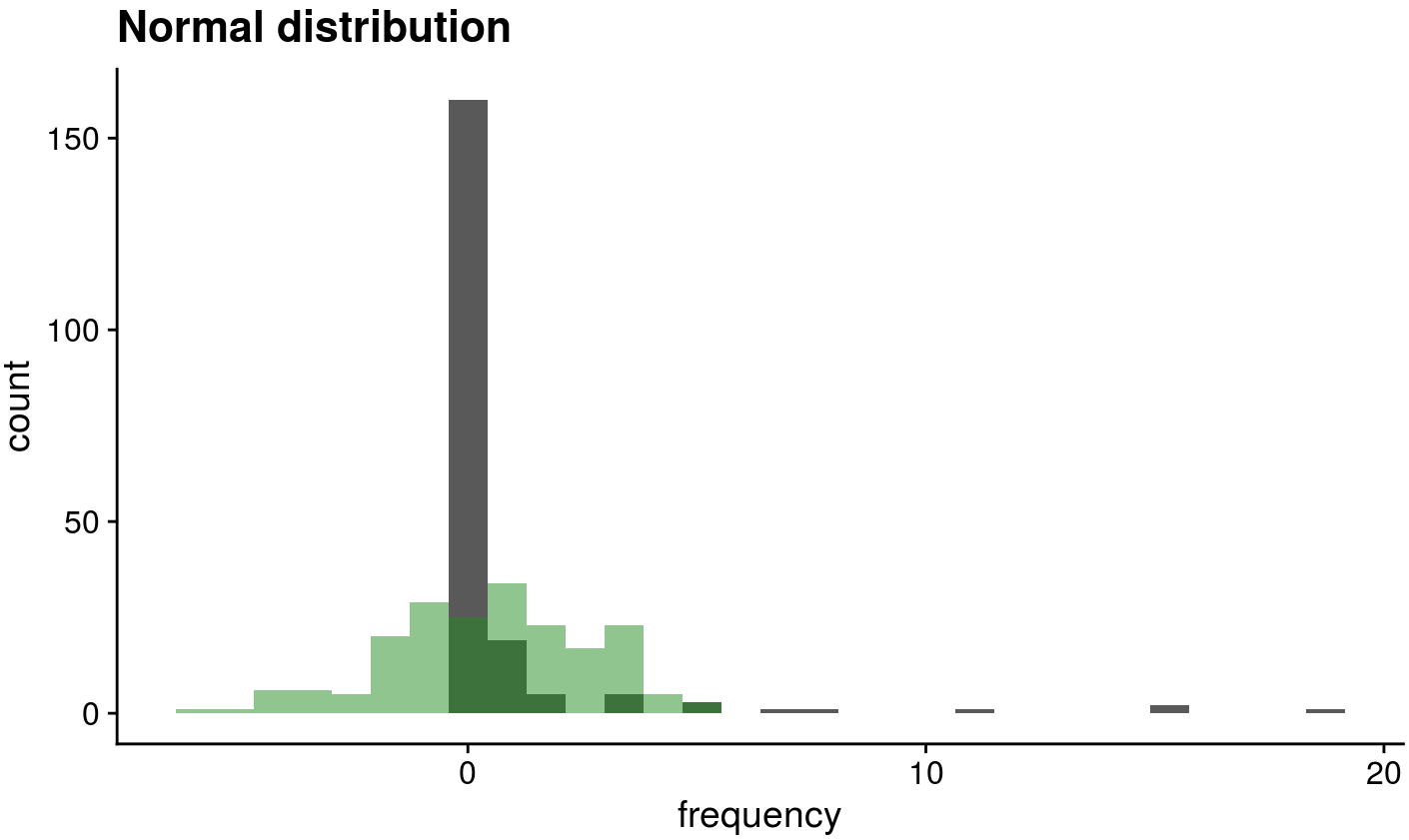
How accurate was the model?

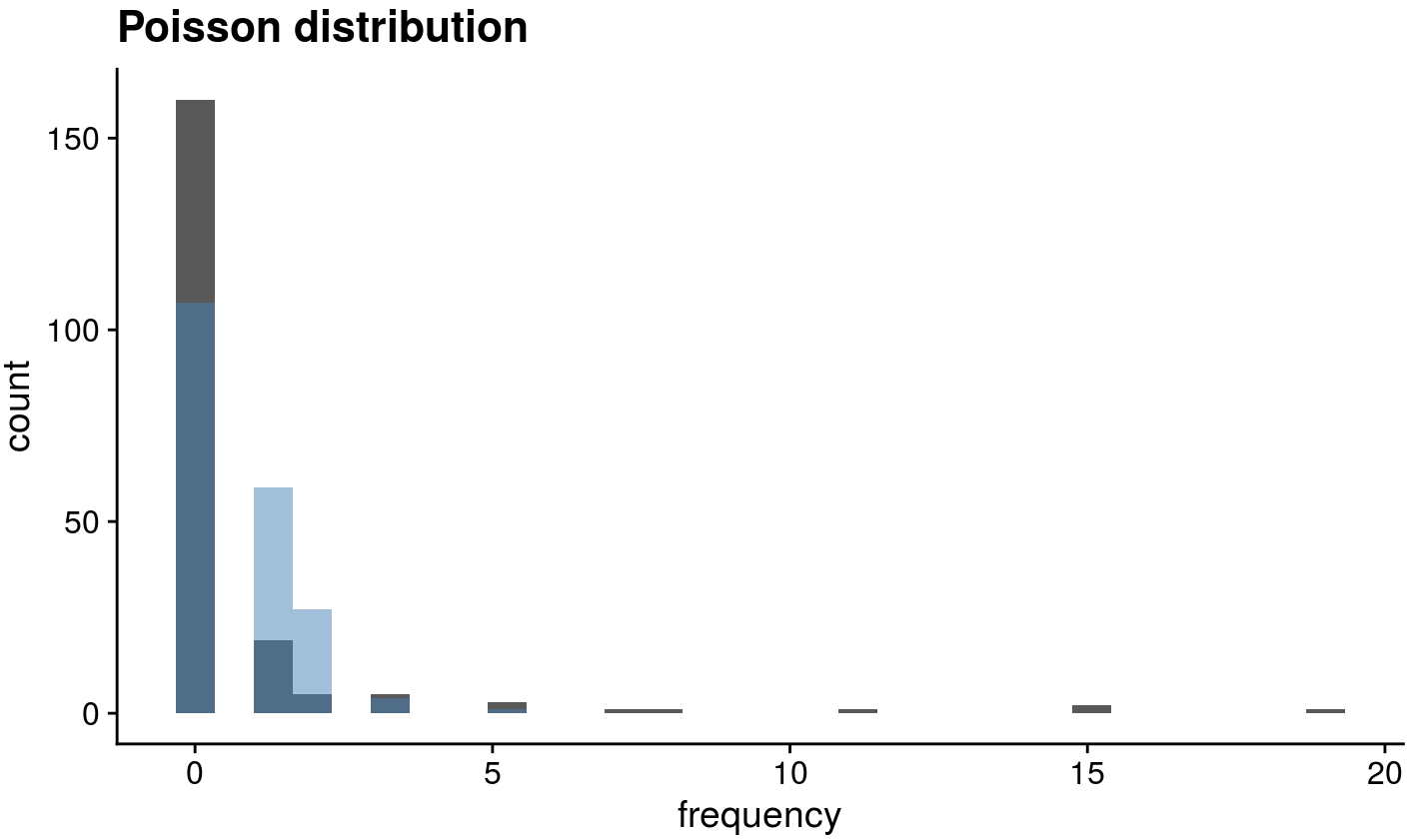
Not great…

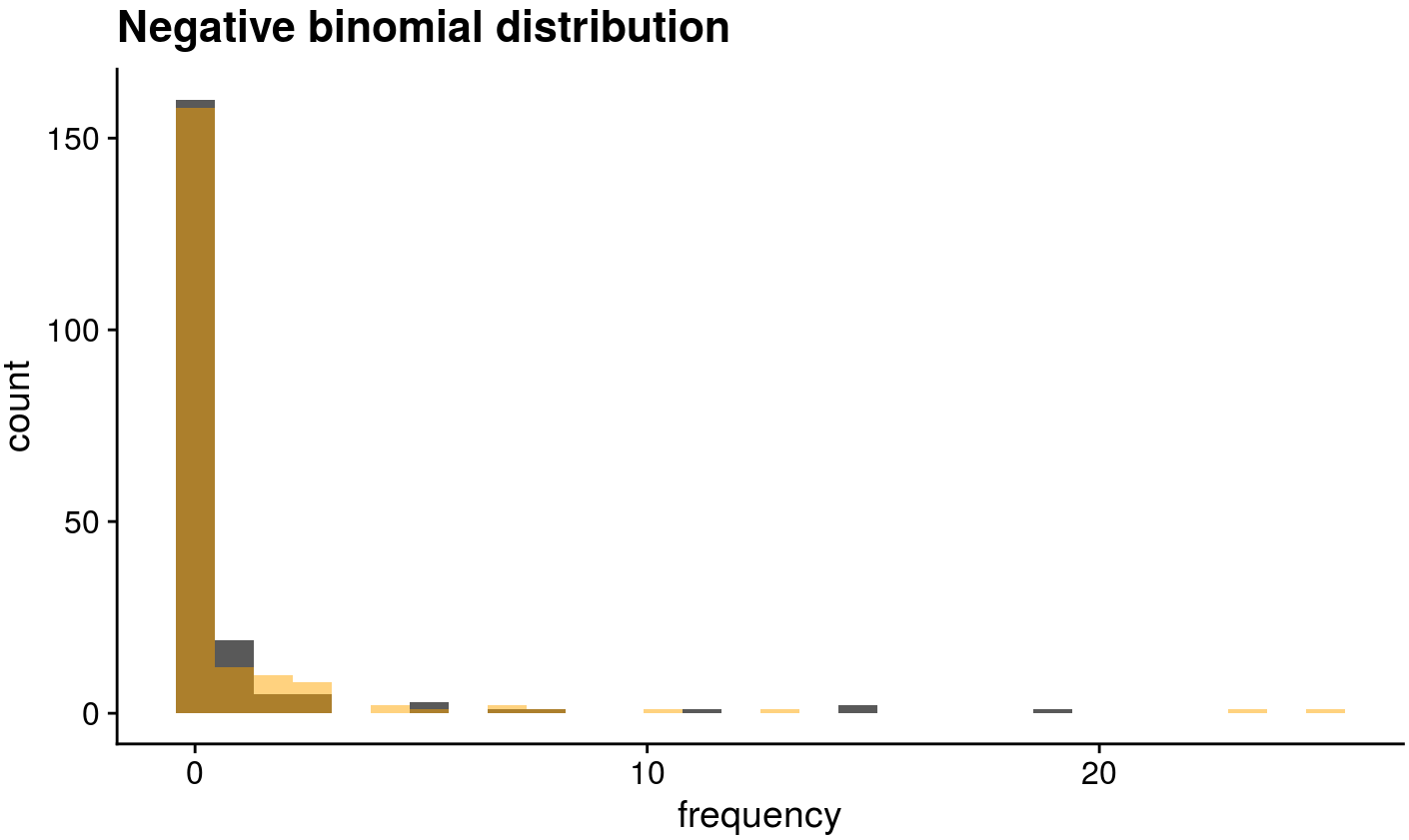


Pick a better error distribution




Normal distribution
150
100
50
0
count
0
10
20
frequency


Poisson distribution
150
100
50
0
count
0
5
10
15
20
frequency

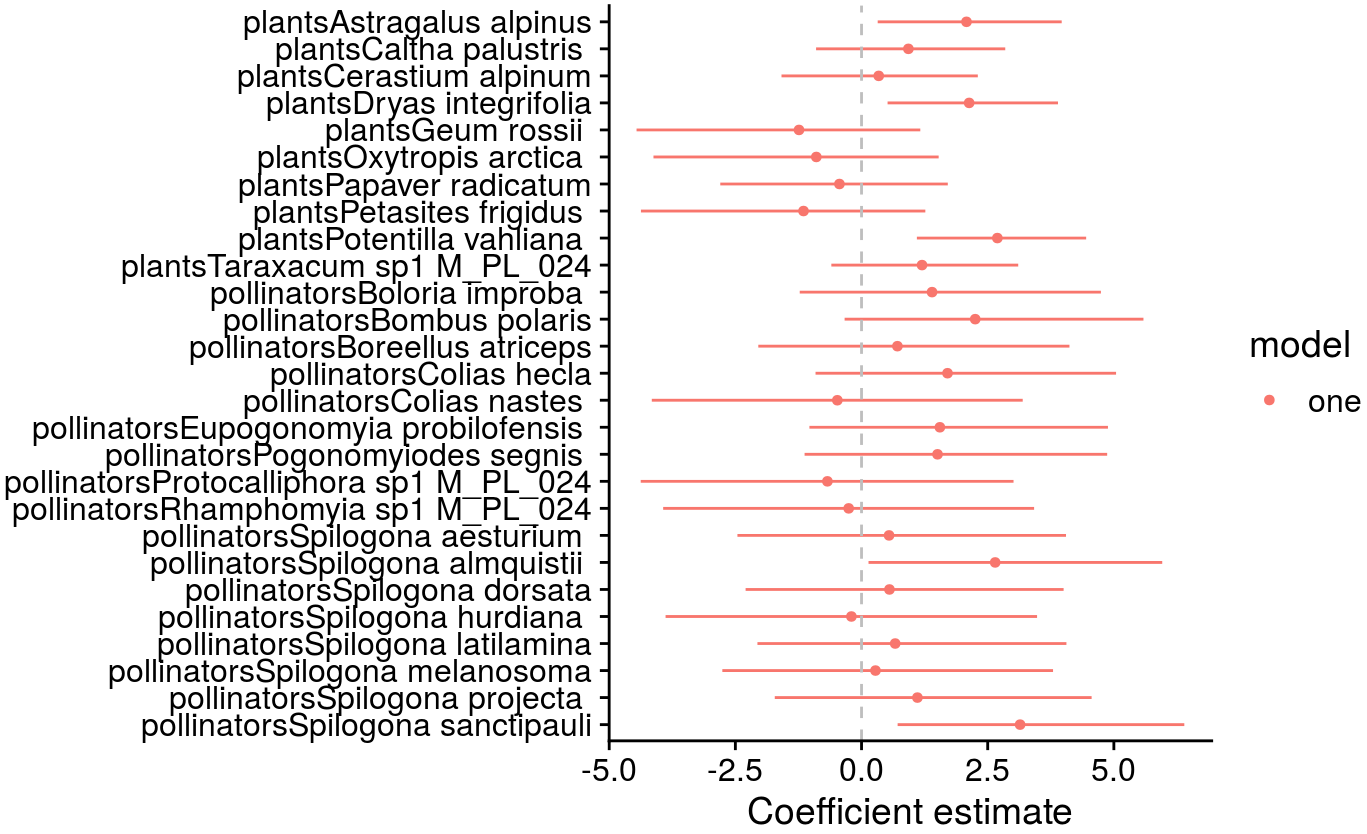

Negative binomial distribution
150
100
50
0
count
0
10
20
frequency

Negative binomial model fit

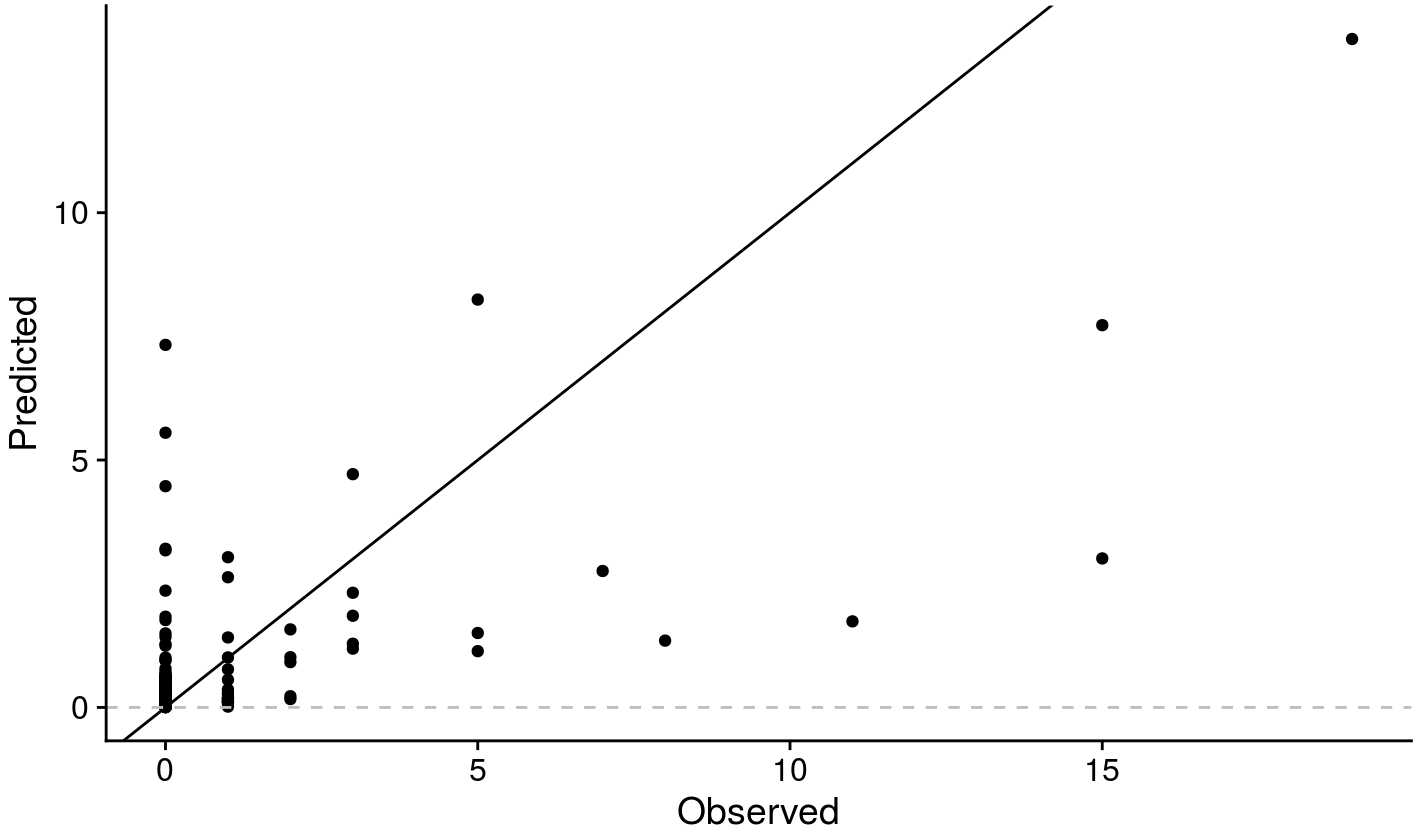
```
glm.nb(frequency ~ plants + pollinators + 1)
```

- Nagelkerke $R^2 = 0.66$

Identify key players

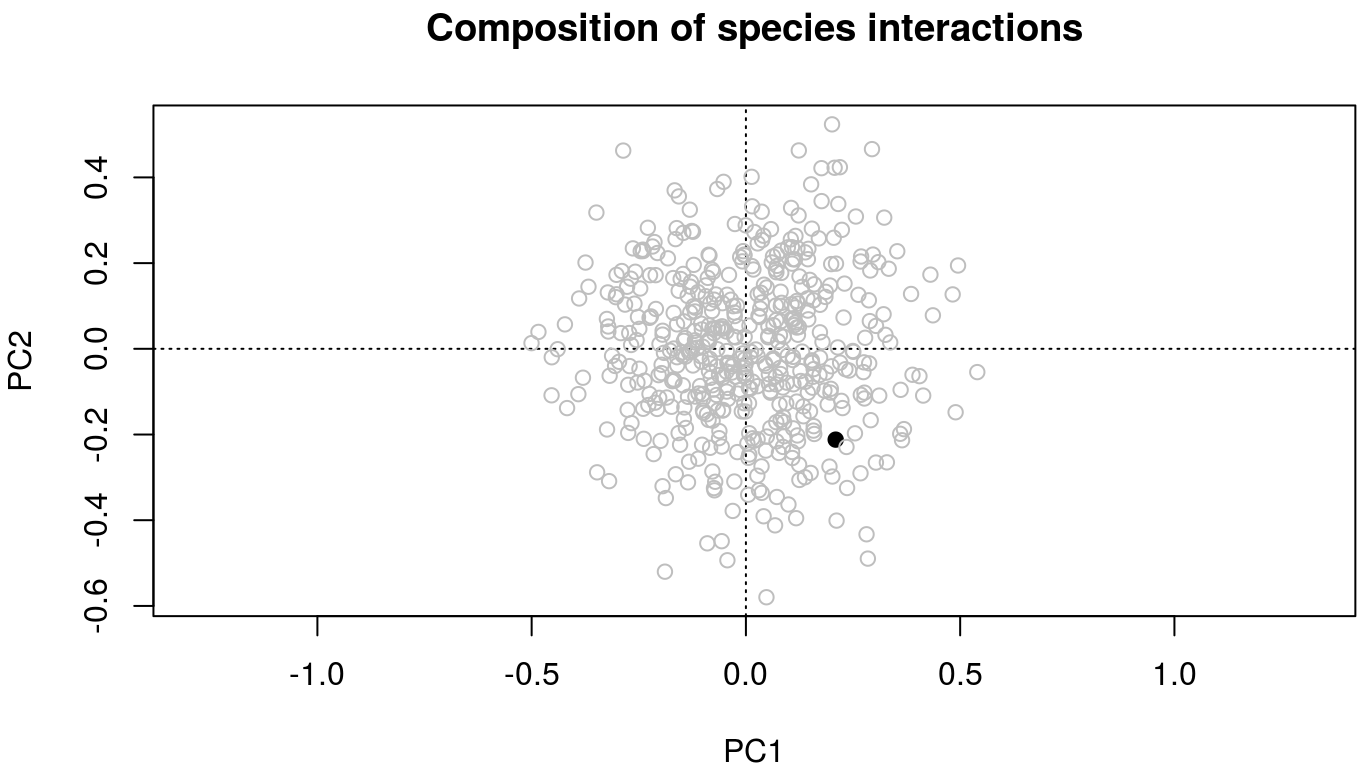


How accurate was the model?

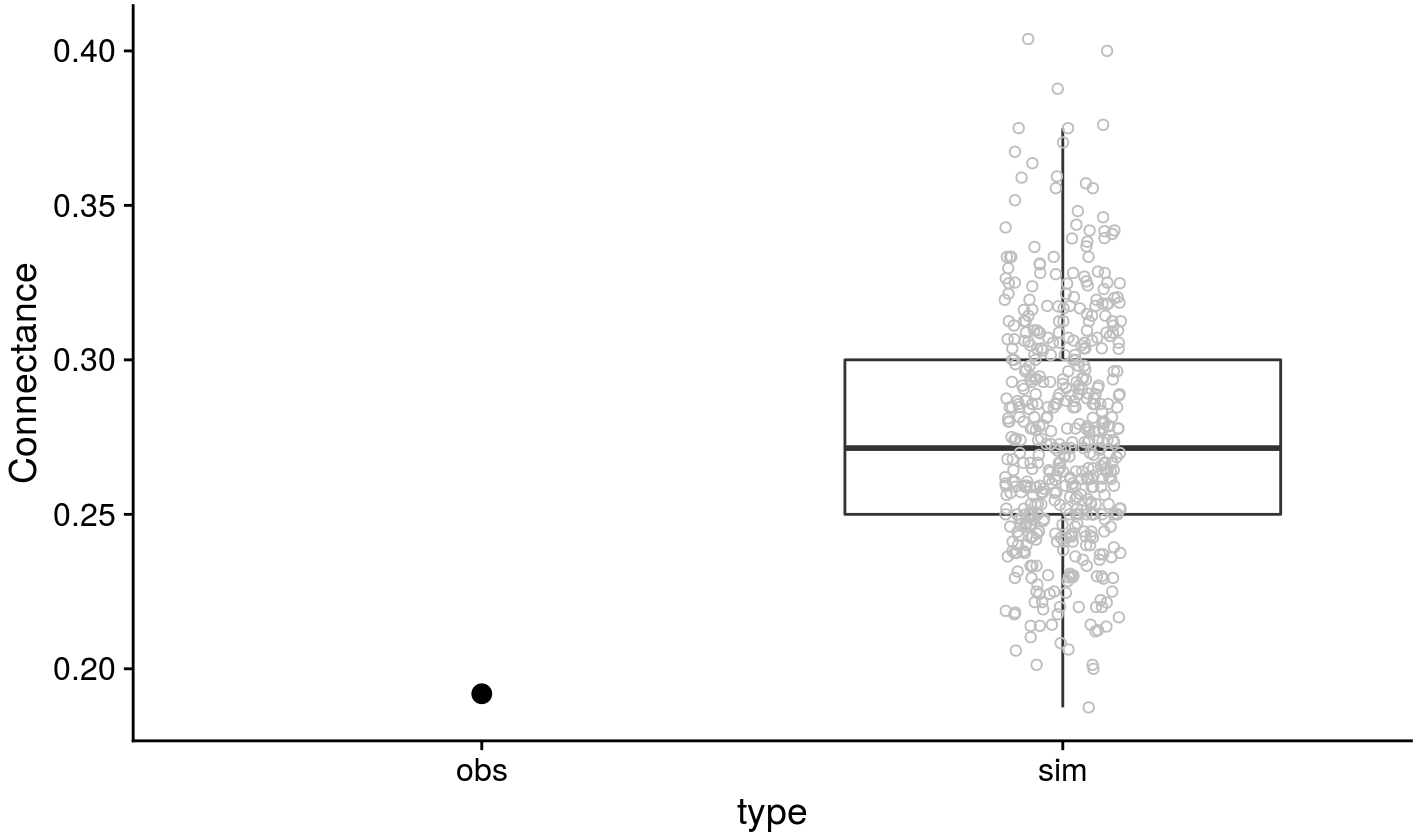


Simulating species interactions

- Aiming for a bullseye



Simulating network properties



Modeling ecological networks

Two approaches:

1. Focus on network properties
   - Advantage: fewer assumptions
   - Disadvantage: no insight to how network properties change
2. Focus on species interactions
   - Advantage: insight to how network properties change
   - Disadvantage: require more data and assumptions, important to check that model can reproduce observations