

# Controlling for Network Biases

Data collection biases are a persistent issue in studies of social networks. Two main types of biases can be considered: exposure biases and censoring biases .

To account for exposure biases, we can switch the network link probability model from a *Poisson* distribution to a *Binomial* distribution, as the binomial distribution allows us to account for the number of trials for each data estimation.

To address censoring biases, we need to add an additional equation to account for the probability of missing an interaction during observation when modeling the interaction between individuals  $i$  and  $j$ .

## Considerations



### Example 1

Below is an example code snippet demonstrating a Bayesian network model with a sender-receiver effect, a dyadic effect, and a block model effect while accounting for exposure biases:

```
from BI import bi
# Setup device-----
m = bi(platform='cpu')

m.data_on_model = dict(
    idx = idx,
    Any = Any-1,
    Merica = Merica-1,
    Quantum = Quantum-1,
    result_outcomes = m.net.mat_to_edgl(data['outcomes']),
```

```

kinship = m.net.mat_to_edgl(kinship),
focal_individual_predictors = data['individual_predictors'],
target_individual_predictors = data['individual_predictors'],
exposure_mat = data['exposure']
)

def model(idx, result_outcomes,
          exposure_mat,
          kinship,
          focal_individual_predictors, target_individual_predictors,
          Any, Merica, Quantum):
    # Block -----
    B_any = m.net.block_model(Any,1)
    B_Merica = m.net.block_model(Merica,3)
    B_Quantum = m.net.block_model(Quantum,2)

    ## SR shape = N individuals-----
    sr = m.net.sender_receiver(focal_individual_predictors,target_individual_predictors)

    # Dyadic shape = N dyads-----
    dr = m.net.dyadic_effect(dyadic_predictors)

    m.dist.binomial(total_count = m.net.mat_to_edgl(exposure_mat), logits = jnp.exp(B_any - B_Merica - B_Quantum))

m.fit(model)
summary = m.summary()
summary.loc[['focal_effects[0]', 'target_effects[0]', 'dyad_effects[0]']]

```

## Example 2

Below is an example code snippet demonstrating a Bayesian network model with a sender-receiver effect, a dyadic effect, and a block model effect while accounting for exposure biases and censoring biases:

## Mathematical Details

### Main Formula

$$Y_{[i,j]} \sim \text{Binomial}\left(E_{[i,j]}, Q_{[i,j]}\right)$$

$$Q_{[i,j]} = \phi_{[i,j]} \eta_{[i]} \eta_{[j]}$$

Where:

- $E_{[i,j]}$  is the number of trials for each observation (i.e., the sampling effort).
- $Q_{[i,j]}$  is the indicator of a true tie between  $i$  and  $j$ , defined as:

$$Q_{[i,j]} \sim \begin{cases} 0 & \text{if no interaction occurs or if } i \text{ or } j \text{ is not detectable} \\ 1 & \text{if } i \text{ and } j \text{ are both detectable} \end{cases}$$

- $\phi_{[i,j]}$  is the probability of a true tie between  $i$  and  $j$ .
- $\eta_{[i]}$  is the probability of individual  $i$  being detectable.
- $\eta_{[j]}$  is the probability of individual  $j$  being detectable.

### Defining formula sub-equations and prior distributions

We can let  $\eta_{[i]}$  depend on individual-specific covariates. To model the probability of censoring, we can model  $1 - \eta_{[i]}$ :

$$\text{logit}(1 - \eta_{[i]}) = \mu_\psi + \hat{\psi}_{[i]} \sigma_\psi + \dots$$

Where:

- $\mu_\psi$  is the intercept.
- $\sigma_\psi$  is a scalar for the variance of random effects.
- $\hat{\psi}_{[i]} \sim \text{Normal}(0, 1)$ , and the ellipsis signifies any linear model of coefficients and individual-level covariates. For example, if  $C$  is an animal-specific measure, like a binary variable for cryptic coloration, then the ellipsis may be replaced with  $\kappa_{[5]} C_{[i]}$  to give the effects of coloration on censoring probability.

## Note(s)

### Note

- One major limitation of this model is the necessity of having an estimation of the censoring bias for each individual.