Missing and Spurious Interactions in Complex Networks

Joan Verguizas I Moliner

Introduction

Network reliability is a concern where complex networks are studied. In this work, we present a model to identify missing and spurious interactions in noisy network observations.

Given an observed network A_{ij}^{O}, the reliability of the link expresses the probability of the link connecting nodes i and j truly existing. This is given by:

$$R_{ij}^{L} = \frac{1}{Z} \sum_{P \in \mathcal{P}} \left(\frac{l_{\sigma_i \sigma_j}^{O} + 1}{r_{\sigma_i \sigma_j} + 2} \right) \exp\left[-H(P)\right]$$

where $l^{O}_{\sigma_i\sigma_j}$ represents the number of links between the nodes belonging to the same groups as nodes i and j. Meanwhile, $r_{\sigma_i\sigma_j}$ represents the maximum number of these connections.

Introduction

$$R_{ij}^{L} = \frac{1}{Z} \sum_{P \in \mathcal{P}} \left(\frac{l_{\sigma_i \sigma_j}^{O} + 1}{r_{\sigma_i \sigma_j} + 2} \right) \exp\left[-H(P)\right]$$

The formula sums over all the partitions of the network. Additionally, H(P) and Z are defined as:

$$H(P) = \sum_{P \in \mathcal{P}} \left[\ln\left(r_{\alpha\beta} + 1\right) + \ln \binom{r_{\alpha\beta}}{l_{\alpha\beta}^{O}} \right] \qquad Z = \sum_{P \in \mathcal{P}} \exp\left[-H(P)\right]$$

Where Z is used as a normalization parameter of the reliability and H(P) represents the cost (energy) of a particular partition configuration in terms of how well it models the network structure. The lower the value of H(P), the more likely that partition P fits the network data.

Metropolis Algorithm

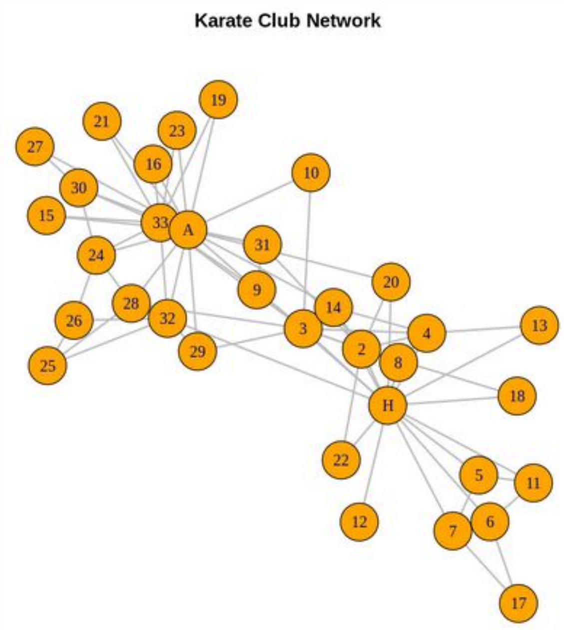
In practice, it is not possible to sum over all partitions. Instead, we can use the Metropolis algorithm to correctly sample relevant partitions and obtain estimates for the link reliability.

At every interaction we change the community of one node at random.
If $\Delta H = H_f - H_i < 0$ we accept the change. Otherwise the change is accepted with a probability equal to $\exp(-\Delta H)$. At the end of the procedure we will have a set of samplings equal to the number of epochs of the metropolis algorithm (10,000).

Once we obtain the partitions, we can calculate the reliability of the links,

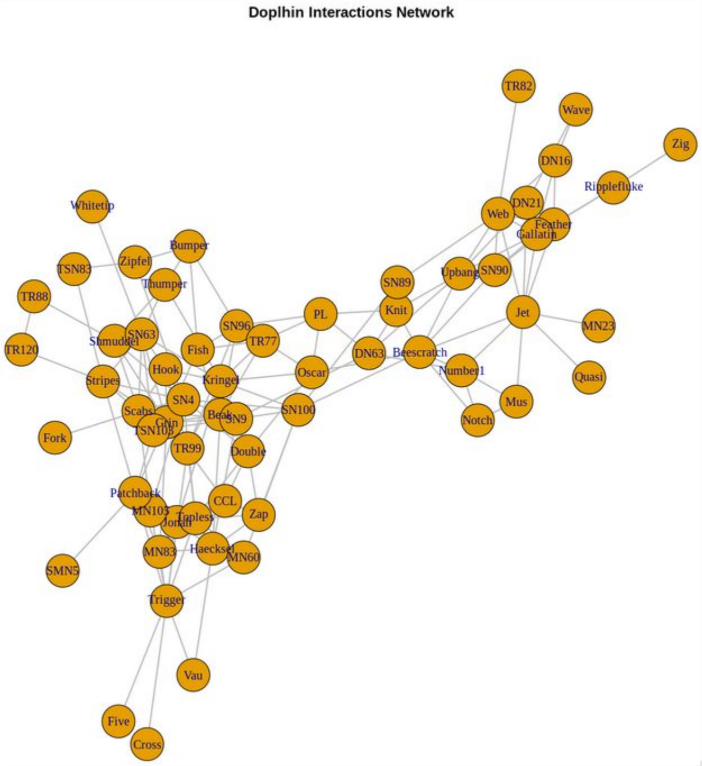
Networks: Karate Club

Represents the social interactions between 34 members of a karate club at a U.S. university during a period of conflict.
Nodes: 34
Vertices: 78
Each vertex represents students that interacted between them outside the club.
Network divides into communities between groups of students that have a social life outside the club.



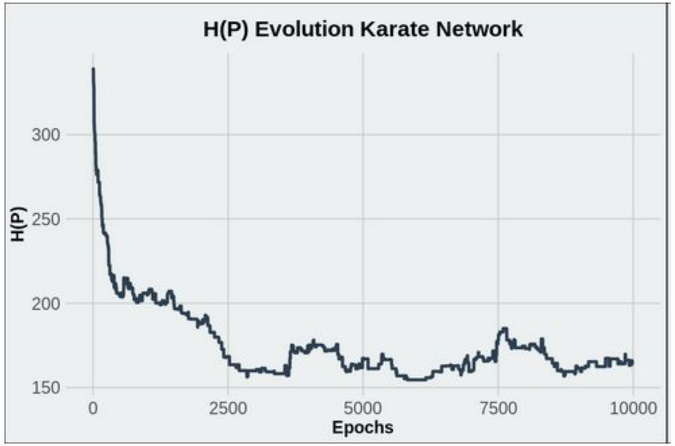
Networks: Dolphins

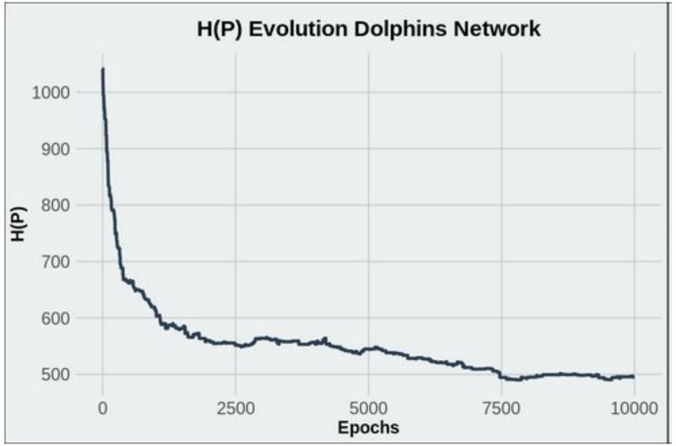
Represents the social interactions among a population of bottlenose dolphins living in Doubtful Sound, New Zealand.
Nodes: 62
Vertices: 159
The interactions are based on actions observed such as swimming together or being in close proximity.
Network divides into communities corresponding to groups of dolphins that associate frequently with each other.



Partitions

Initial configuration Number communities = Number of Nodes
10,000 epochs for Metropolis Algorithm





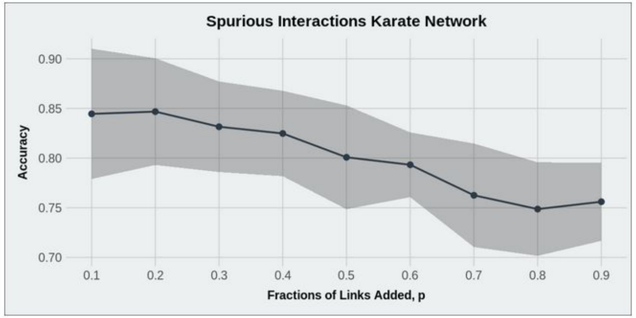
Spurious Interactions

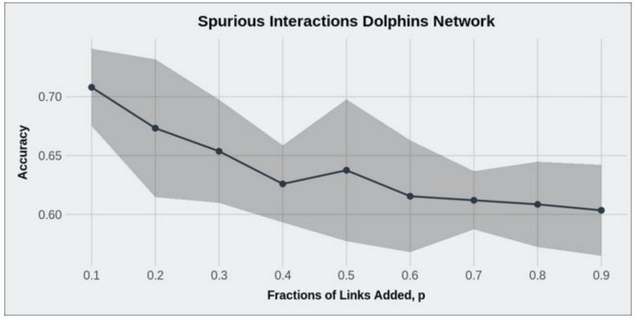
We create fake observations with spurious interactions by adding a fraction of the already existent number links to the true network.

Accuracy of the algorithm of detecting spurious interactions is given by the fraction of true positives ($A^O_{ij} = 1$ and $A^T_{ij} = 1$) with a higher link reliability than false positives ($A^O_{ij} = 1$ and $A^T_{ij} = 0$).

$$\text{Accuracy} = P(R^L(\text{True Positives}) > R^L(\text{False Positives}))$$

We average the results over 25 iterations for different values of links added into the observed network





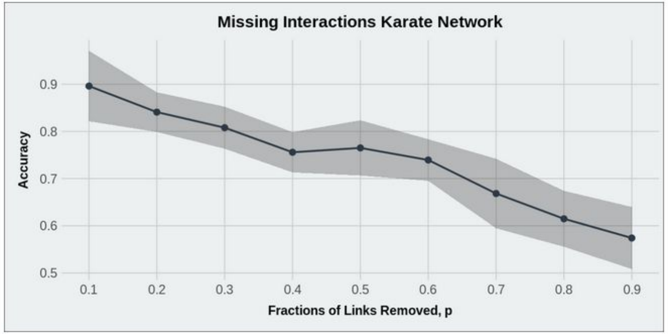
Missing Interactions

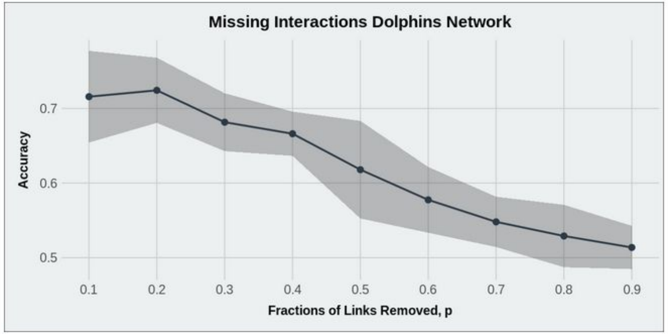
We create fake observations with missing interactions by removing a fraction of the links of the true network.

Accuracy of the algorithm of detecting missing interactions is given by the fraction of true negatives ($A^O_{ij} = 0$ and $A^T_{ij} = 0$) with a lower link reliability than false negatives ($A^O_{ij} = 0$ and $A^T_{ij} = 1$).

$$\text{Accuracy} = P(R^L(\text{True Negatives}) < R^L(\text{False Negatives}))$$

We average the results over 10 iterations for different values of links removed.





Conclusions

- The algorithm overall is able to detect with good accuracy both the spurious and missing links from the observed networks.
- From missing interactions, accuracy drops as the number of links removed p increases.
- The drop is not that heavy for spurious interactions.
- The accuracy is higher for the karate class network than for the dolphin network.

Thank You for Your Attention