# Introduction

Recent decades ASNA increasing, easier, available, refined.

Many ASN recorded, allows for further meta-studies. ASNR one project gathering such ASN.

One aspect of ASNA still lacking in its use or ease of implementation is ASN uncertainty. Early attempts (Snijders and Borgatti, 1999), more modern takes (Farine and Strandburg-Peshkin, 2015), but not much use (REFS that use them), and not much software implementation, usually manually done (but see ANTs (Sosa et al., 2020). Concerns are usually more on hypothesis testing and comparison to null-models, where permutations algorithm now prevail.

ASN uncertainty relevant topic on itself, intertwining notions of sampling, sampling effort, observation, variability, models.

A common way to approach uncertainty, robustness or variability of random variables emerging from more complex phenomenon (like SN metrics and statistics) is through simulating the simpler variables from which the variables of interest emerge. Once again, this is usually done through coding simulations from scratch in a programming language (REFS done in R). These simulations end up being tailored for their task only; and while open-sourced code allows for reusing and adapting a simulation framework to fit another design, this still implies advanced coding reading and writing skills. Indeed, reverse engineering a working algorithm and adapting it to fit another or a more general purpose can be altogether too much for any ASN analyst to tinker with.

Here, we introduce and showcase SimuNet, an ASNA simulation R package that fits three purposes:

1. Present a new approach to weighted SN uncertainty assessment
2. Showcase how this approach can be used to simulate new weighted SN mimicking empirical ASN building based on observation
3. Serve as an evolving and multipurpose simulation framework to try and test the effect fo different experimental designs on network uncertainty

# How to use SimuNet

## Installation

Github install

## Import ASN

From new data : simulated or manual entry (own data, paper)

From asnr

## Simulate new empirical networks

Theoretical network

## Tweak simulations

Introducing empirical sampling methods

## Extract Sn values of interest

Extract from different SimuNet objects

# Underlying network simulation’s paradigm

## Core paradigm: edge weights are assimilable to binomial variables

Most ASN use weighted networks, but calculate edge weights through summing (and standardizing) binary data. In the case of a group-scan sampling, after giving a definition to edges, individuals i and j are either connected (aij = 1) or not in the network (aij = 0). The random variable aij is thus a binary random variable, and a parameter θij can be introduced as the probability of aij being equal to 1 during a given scan. One interested in using ASNA to model and understand animal behaviour can thus think of θij as a biological value depending on i and j themselves, their specific relationship, the context during the scan occurs, their other relations (e.g. i will interact or associate with j only if j is “close” to k) and interactions (e.g. i and j will interact or associate only when individual k associate with h), etc. θij can also be thought to vary overtime, i.e. leading to dynamic SNA, either on its own or once again following other SN (e.g. i will interact or associate with j only if j interacted with k during the previous scan) and non-SN variables (e.g. i will overall interact less on weekends, or as it ages).

Whichever the causes affecting θij, it results from this paradigm that during any scan, one can define such a probability of association/interaction as the underlying driver of binary edges existing or not between two individuals during the scan. This model can be more or less simple or simplified to fit the chosen theoretical assumptions, e.g. θij does not vary across scans (e.g. the weighted SN is considered static), θij is not influenced by anything else than individuals i, j, and their unique relationship. We argue that, whether conscious or not, ASN analysts do choose such specific assumptions and that it is important to be aware of such assumptions underlying the elected SN model.

## What are the benefits of this paradigm?

One benefit of decomposing an edge weight aij into the sum of binary variables with parameter θij is that now aij can be viewed as a binomial variable within a Bayesian framework, from which distributions, credible intervals, variability can be derived from. This also allows an easy step-by-step “forensic” reconstruction of what likely sequences of scans could have led to the observed network, even in the absence of the specific observed scans (which are usually not provided in the ASNA literature, open data often being restricted to the final summed up adjacency matrix of the SN).

## Bayesian framework

θij being a probability, it is bounded on the interval [0,1]. It is very usual in Bayesian statistics to attribute a Beta distribution (Beta(α,β)) to such probability parameter; such a distribution is too bounded on [0,1] and can represent the likelihood of the parameter taking a given value on [0,1], i.e. the “probability” (rather, likelihood) that a probability is equal to X.

Under this framework, if (aij = X | θij), i.e. observing a weight of X knowing the value of θij, observing aij = X and deducing the likely distribution of θij, i.e. (θij | aij = X) is a classic Bayesian situation corresponding to the one ASN analysts face when confronted only to a single observed adjacency matrix. Note that in the first case (aij = X | θij), X is a traditional binomial variable when θij is a constant, with p(aij = X | θij) = (n X) θijX (1 - θij)n - X. In a classic probabilistic/maximum likelihood framework, θij could be calculated from observing aij = X as the number of times i and j associated/interacted divided by the total relevant number of scan n. However, we embrace here the uncertain nature of θij, especially when it is estimated via a single set of observations, by relying on a Bayesian framework where θij is not seen as a constant, but rather as coming from a distribution of likely values, modeled by a Beta distribution. This allows θij to be more or less uncertain/variable according to the sample size (e.g. estimating θij = 0.1 should be more certain after observing 100 associations in 1000 scans rather than only 1 in 10); property that is lost when only relying on θij‘s maximum likelihood estimator X / n.

In this Bayesian framework, p(θij |aij = X), i.e. θij‘s posterior, is also a Beta distribution, relying on the Beta distribution’s conjugate property in the case of a binomial variable (REF). α and β can be view as (pseudo-)counts of association/non-associations, with n = α + β being the total number of observations. That is to say:

1. after adopting an informative or uninformative prior (that can be view as pseudo-counts of association/non-associations) like
   1. α = a and β = b where during a previous experiment, the dyad has been observed associating a times over n = a + b scans as an informative prior
   2. or like α = β = 1 (a uniform distribution of θij) or α = β = 0.5 (Jefrrey’s prior) as an uninformative prior
2. observing aij = X “updates” our knowledge of α and β by summing previous (pseudo-)counts of association/non-associations with the new observations like this: αpost = αprior + X and βpost = βprior + (N – X)

From θij‘s posterior, one can afterward describe θij and its likely values, as well as draw a random value from this posterior and simulate new observations based on this random θij. Such a random draw of θij from a Beta distribution and then drawing new observations leading to a new aij is the process underlying a Beta-Binomial distribution. Note that θij is randomly drawn once and all subsequent observations leading to the new aij uses the same value of θij. Randomly drawing θij before each observation would not be a Beta-Binomial process, but could rather be viewed and analyzed as a simple Binomial process with parameter θij = a / (a + b) (REF). Note that the variance of a Beta-Binomial variable (BetaBinom(n,a,b)) is higher than the one of a similarly parametrized “two-steps” binomial variable (Binom(n,a / (a+b)). This can be viewed as accumulating the uncertainty of drawing a single θij and the uncertainty due to sampling new data, whereas in the second case, the averaged behaviour of θij being drawn at each scan “reduce” this part of uncertainty during the overall process.