

# On the inference of positive and negative interactions and their relation to abundance

*Andrew J. Rominger*

A recent paper inferred species-species interactions from spatially replicated abundance data and further tied those inferred interactions to abundances<sup>1</sup>. Specifically, the authors found that inferred positive interactions were disproportionately associated with rare species, a result they interpreted as rare species being stabilized by facilitative interactions.

```
library(knitr)
opts_chunk$set(echo = TRUE, cache = TRUE,
               fig.width = 4, fig.height = 4, fig.align = 'center')

library(pika)
library(RarePlusComMinus)

## Loading required package: vegan
## Loading required package: permute
## Loading required package: lattice
## This is vegan 2.5-3

library(parallel)
library(MASS)
library(socorro)
library(viridis)

## Loading required package: viridisLite

# threading defaults
nthrd <- detectCores()
nthrd <- ifelse(round(nthrd * 0.8) >= nthrd - 1, nthrd - 1, round(nthrd * 0.8))
if(nthrd < 1) nthrd <- 1

# plotting defaults
parArgs <- list(mar = c(3, 3, 0, 0) + 0.5, mgp = c(1.5, 0.30, 0), tcl = -0.25)
cexDefault <- 1.4
lwdDefault <- 2
```

First we reproduce some of the key results of ref.<sup>1</sup>, namely how abundance relates to positive and negative association networks.

```
# load data from paper
data('abundance_matrices')

# clean it
obsdat <- lapply(abun.mat, function(x) {
  x <- ceiling(x)
  x <- x[rowSums(x) > 0, colSums(x) > 0]

  return(x)
})
```

```

commStats <- mclapply(obsdat, mc.cores = nthrd,
                     FUN = function(x) unlist(plusMinus(x)))

commStats <- as.data.frame(do.call(rbind, commStats))

# remove studies with too few plus or minus links
commStats[is.na(commStats$pos.rho.rho) | is.na(commStats$neg.rho.rho), ] <- NA

```

It should be noted that the authors of the original analysis<sup>1</sup> retained 326 studies after filtering, whereas we retain 300. This may be in part because we removed any dataset for which there were fewer than three links in either the positive or negative networks, whereas the authors of the original analysis removed only those datasets with fewer than two links<sup>1</sup>.

And now for the plot that reproduces Fig. 2 (b and c) from ref.<sup>1</sup>.

```

# ----
# helper function for making fancy histograms
specialHist <- function(x, breaks, col, add = FALSE, ...) {
  h <- hist(x, breaks, plot = FALSE)

  if(!add) plot(range(h$breaks), range(h$counts), type = 'n', ...)

  rect(xleft = h$breaks[-length(h$breaks)], xright = h$breaks[-1],
       ybottom = 0, ytop = h$counts, col = col)
}

# ----
# function to remake Fig 2(b-c)
fig2bc <- function(x, breaksRho, breaksWM) {
  foo <- split.screen(c(1, 2))

  # correlation differences
  screen(1)

  par(parArgs)
  par(mar = parArgs$mar + c(0, 0, 0, -0.5))
  specialHist(x$pos.rho.rho - x$neg.rho.rho, xlim = c(-2, 2),
             breaks = breaksRho, col = 'gray90',
             xlab = '', ylab = 'Number of assemblages')
  mtext('Abundance-species degree differences\n(positive - negative)',
       side = 1, line = 2.5)
  abline(v = 0, col = 'red', lty = 2, lwd = 2)

  # raw correlations
  foo <- split.screen(matrix(c(0.475 + 0.02, 0.975 + 0.02, 0.475, 0.975), nrow = 1),
                     erase = FALSE)
  screen(3, new = FALSE)

  par(parArgs)
  par(cex = 0.75)
  par(mgp = par('mgp') * par('cex'))

  specialHist(x$pos.rho.rho, xlim = c(-1, 1), ylim = c(0, 90),

```

```

        breaks = breaksRho / 2, col = hsv(0.65, alpha = 0.5),
        xlab = expression("Spearman's"~rho),
        ylab = '')
specialHist(x$neg.rho.rho,
            breaks = breaksRho / 2, col = hsv(0, alpha = 0.5),
            add = TRUE)

# mean differences
screen(2)
par(parArgs)
par(mar = parArgs$mar + c(0, -1, 0, +0.5))

specialHist(x$pos.wm - x$neg.wm, xlim = c(-0.3, 0.3),
            breaks = breaksWM, col = 'gray90',
            xlab = '', ylab = '')
mtext('Abundance differences\n(positive - negative)',
      side = 1, line = 2.5)
abline(v = 0, col = 'red', lty = 2, lwd = 2)

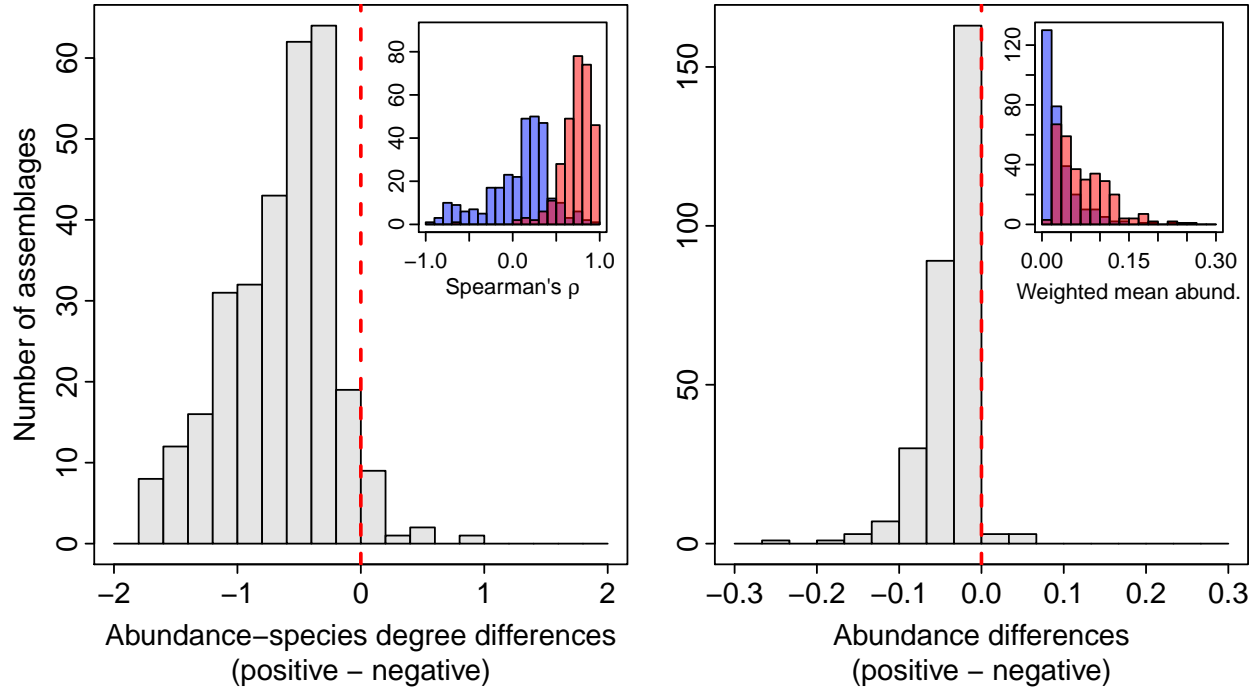
# raw means
foo <- split.screen(matrix(c(0.475 - 0.04, 0.975 - 0.04, 0.475, 0.975), nrow = 1),
                    erase = FALSE)
screen(4, new = FALSE)

par(parArgs)
par(cex = 0.75)
par(mgp = par('mgp') * par('cex'))

specialHist(x$pos.wm, xlim = c(0, 0.3),
            breaks = (breaksWM + 0.3) / 2, col = hsv(0.65, alpha = 0.5),
            xlab = 'Weighted mean abund.',
            ylab = '')
specialHist(x$neg.wm,
            breaks = (breaksWM + 0.3) / 2, col = hsv(0, alpha = 0.5),
            add = TRUE)
}

# ----
# plotting
fig2bc(commStats, breaksRho = seq(-2, 2, by = 1/5),
        breaksWM = seq(-0.3, 0.3, by = 0.1/3))

```



*# estimate number of assemblages suggested by Fig 2c; number refer to arbitrary  
# pixel coordinates from the figure rescaled to the y-axis units*

```
y0 <- 517
ys <- c(173, 345)

r <- 220 / mean((y0 - ys) * 1:2)

yblu <- c(234, 399, 478, 498, 507, 507, 509, rep(512, 5))
yred <- c(498, 375, 403, 419, 467, 482, 505)

fig2est <- r * (sum((y0 - yblu)) + sum((y0 - yred)))
```

It should be noted that our inset figure of mean weighted abundances differs from the original paper<sup>1</sup> in that their y-axis ranges from 0 to 220, while ours ranges from 0 to 120; we suspected the original authors mislabeled their axis, because the scale as presented would suggest over 600 assemblages, while the number should be 326. Rescaling their axis to range from 0 to 120 brings this estimate more in line with the reported number of assemblages.

Now we explore the shape of the spatial abundance patterns in the real data provided by ref. 1.

```
summStats <- mclapply(1:length(obsdat), mc.cores = nthrd, FUN = function(i) {
  x <- obsdat[[i]]
  o <- sapply(1:ncol(x), function(j) {
    # sometimes optimization fails, so catch that and return NA
    params <- try(fitdistr(x[, j], 'negative binomial'))
    if(class(params) == 'try-error') {
      params <- c(size = NA, mu = NA, loglik = NA)
    } else {
      params <- c(params$estimate, loglik = params$loglik)
    }
  })

  nocc <- sum(x[, j] > 0)
```

```

    abund <- sum(x[, j])

    return(c(nsite = nrow(x), nspp = ncol(x), nocc = nocc, abund = abund, J = sum(x),
            params))
  })
})

summStats <- data.frame(study = rep(names(obsdat), sapply(summStats, ncol)),
                       t(do.call(cbind, summStats)))

# remove failed optims
summStats <- summStats[!is.na(summStats$size), ]

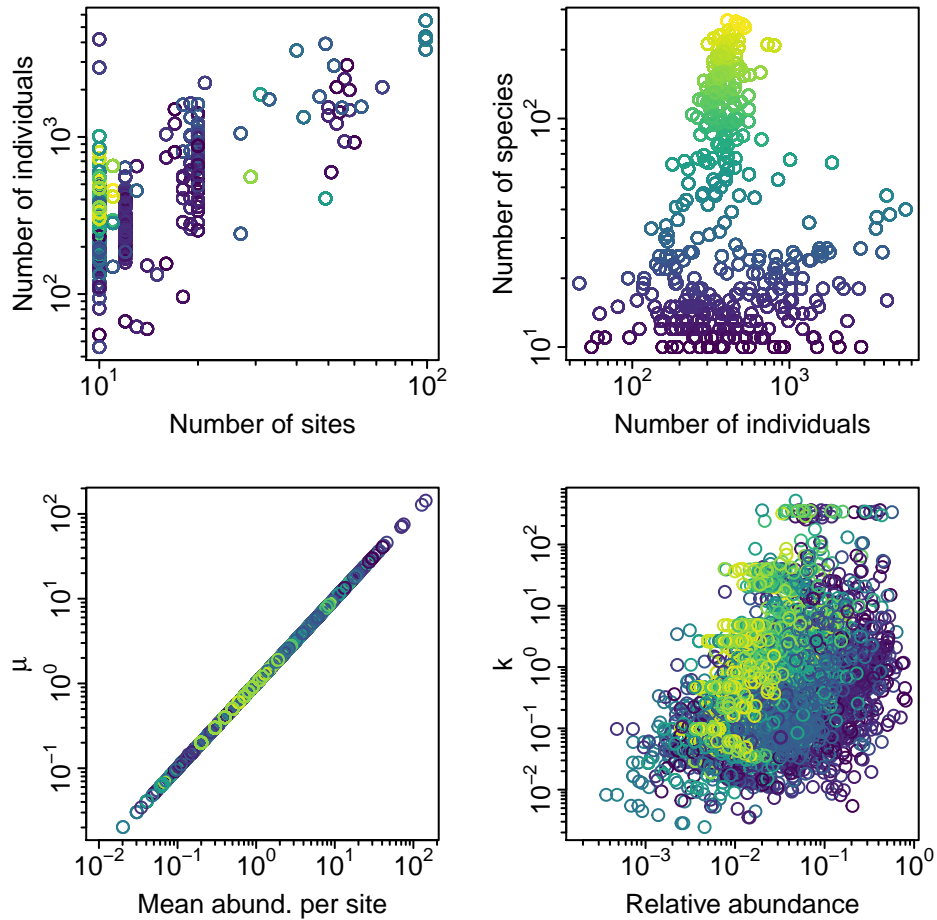
# looks like an optim bug around size = 100, remove it
summStats[round(summStats$size) == 100, c('size', 'mu', 'loglik')] <- NA

# helper function for easy log-log ploits
llplot <- function(x, y, explab = TRUE, ...) {
  plot(x, y, xaxt = 'n', yaxt = 'n', ...)
  logAxis(1:2, explab = explab)
}

layout(matrix(1:4, ncol = 2, byrow = TRUE))
par(parArgs)

llplot(summStats$nsite, summStats$J, log = 'xy',
       xlab = 'Number of sites', ylab = 'Number of individuals',
       col = quantCol(summStats$nspp, viridis(40), 'log'))
llplot(summStats$J, summStats$nspp, log = 'xy',
       xlab = 'Number of individuals', ylab = 'Number of species',
       col = quantCol(summStats$nspp, viridis(40), 'log'))
llplot(summStats$abund / summStats$nsite, summStats$mu, log = 'xy',
       xlab = 'Mean abund. per site', ylab = expression(mu),
       col = quantCol(summStats$nspp, viridis(40), 'log'))
llplot(summStats$abund / summStats$J, summStats$size, log = 'xy',
       xlab = 'Relative abundance', ylab = expression(k),
       col = quantCol(summStats$nspp, viridis(40), 'log'))

```



We can also get a sense for the SADs of each assemblage

```
mods <- c('fish', 'plnorm', 'tnegb')
sadStats <- mclapply(obsdat, mc.cores = nthrd, FUN = function(x) {
  x <- colSums(x)
  s <- fitSAD(x, mods)

  i <- which.min(sapply(s, AIC))
  o <- s[[i]]$MLE
  if(i == 1) o <- c(o, NA)

  o <- c(i, o)
  names(o) <- NULL

  return(o)
})

sadStats <- as.data.frame(do.call(rbind, sadStats))
names(sadStats) <- c('mod', 'par1', 'par2')
sadStats$mod <- mods[sadStats$mod]

# limit to only those sites that produced good networks
sadStats <- sadStats[rownames(sadStats) %in%
  rownames(commStats[!is.na(commStats$pos.n), ]), ]
```

We'll make linear models of the SSAD relationships to sample from in addition to sampling from the SAD

```
# indBySiteMod <- lm(log(summStats$J) ~ log(summStats$nsite))
# data for linear model of k, excluding sites that didn't produce a good network
dat4k <- with(summStats[summStats$study %in%
                rownames(commStats[!is.na(commStats$pos.n), ]), ],
              data.frame(logk = log(size),
                        loga = log(abund / J),
                        logS = log(nspp))
)

kByRelSppMod <- lm(logk ~ loga + logS, data = dat4k)

# function to sample random params and generate random data
makeData <- function(nsite, nspp) {
  # make SAD
  iSAD <- sample(nrow(sadStats), 1)
  rfun <- get(paste0('r', sadStats$mod[iSAD]))
  pars <- as.numeric(sadStats[iSAD, 2:3])
  pars <- pars[!is.na(pars)]
  abund <- do.call(rfun, c(list(nspp), as.list(pars)))

  # calculate known quantities from abund
  J <- sum(abund)
  mu <- abund / nsite

  # calculate k (size param)
  k <- exp(predict(kByRelSppMod, newdata = data.frame(loga = log(abund / J),
                                                       logS = log(nspp))) +
            rnorm(nspp, sd = summary(kByRelSppMod)$sigma))

  # loop over abundances and generate ssad
  sapply(1:length(abund), function(i) {
    return(rnbinom(nsite, k[i], mu = mu[i]))
  })
}
```

Now we can do a simulation where we make slightly more than 300 simulated communities (slightly more because some will be rejected) and see how their positive and negative associate networks look.

```
# number of simulations to run
nsim <- round(1.25 * sum(!is.na(commStats$pos.wm)))

# numbers of sites and species to sample from
# note: only using those sites that produced good networks
nsitenspp <- unique(summStats[
  summStats$study %in% rownames(commStats[!is.na(commStats$pos.n), ]),
  c('study', 'nsite', 'nspp')])

# loop over simulation replicates
simPlusMinus <- mclapply(1:nsim, mc.cores = nthrd,
                        FUN = function(i) {
  j <- sample(nrow(nsitenspp), 1)
  simDat <- makeData(nsite = nsitenspp$nsite[j],
                    nspp = nsitenspp$nspp[j])
})
```

```

simDat <- simDat[rowSums(simDat) > 0,
                 colSums(simDat) > 0]

if(any(dim(simDat) < 10)) {
  o <- rep(NA, ncol(commStats))
  names(o) <- names(commStats)
} else {
  o <- unlist(plusMinus(simDat))
}

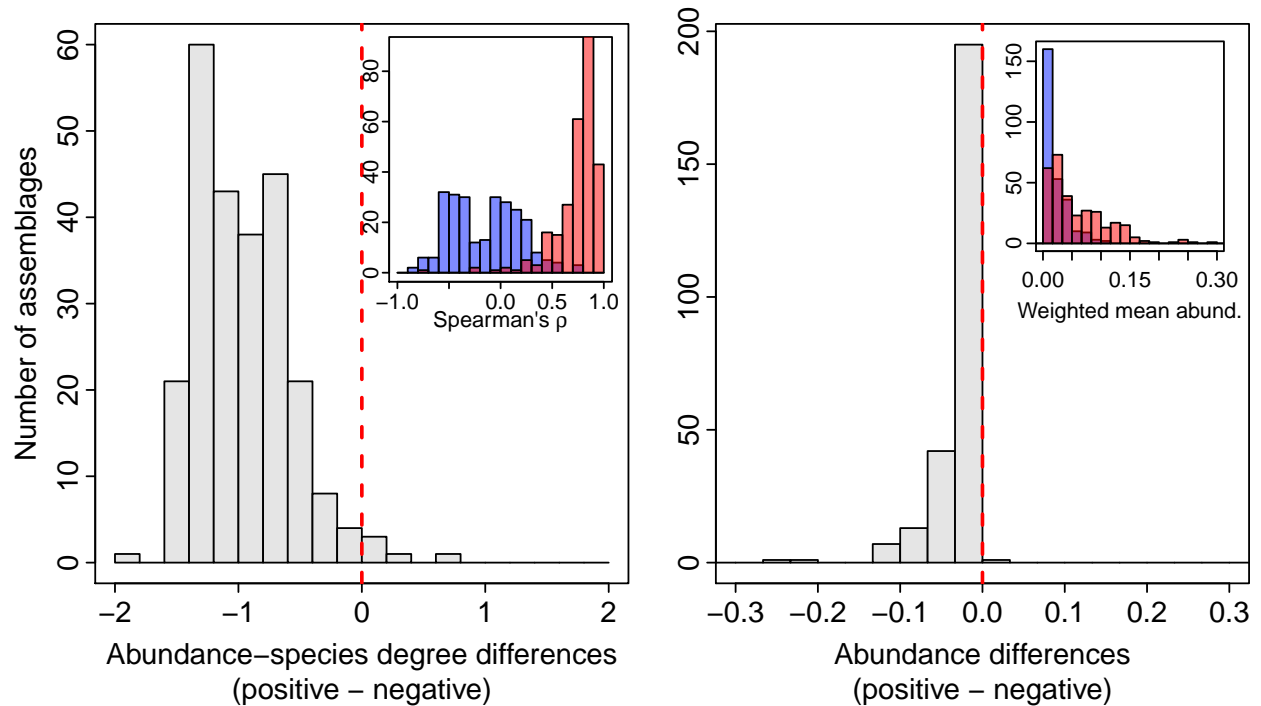
return(o)
})

simPlusMinus <- as.data.frame(do.call(rbind, simPlusMinus))
simPlusMinus <- simPlusMinus[!is.na(simPlusMinus$pos.n), ]

wmmax <- ceiling(max(simPlusMinus$pos.wm, simPlusMinus$neg.wm, na.rm = TRUE) * 9) / 3
fig2bc(simPlusMinus, breaksRho = seq(-2, 2, by = 1/5),
        breaksWM = seq(-wmmax, wmmax, by = 0.1/3))

## Warning in par(new = TRUE): calling par(new=TRUE) with no plot

```



## References

1. Calatayud, J. *et al.* Positive associations among rare species and their persistence in ecological assemblages. *Nat Ecol Evol* (2019).