

## Adirondack lakes

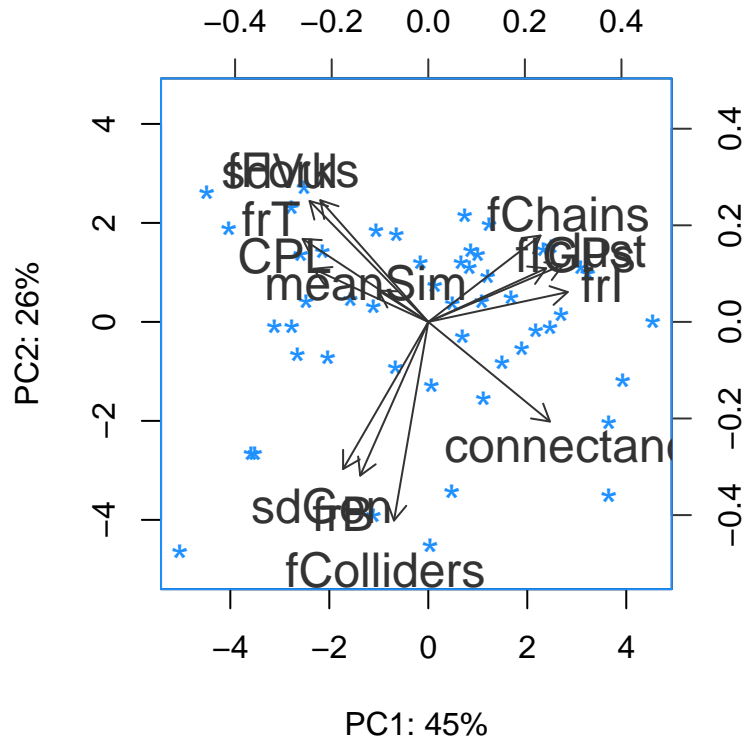
```
library(gatewayr)
#>
#> Attaching package: 'gatewayr'
#> The following object is masked from 'package:base':
#>
#>     subset

d <- subset("study.site", "Adirondack lakes")
#> - 8426 records found.
#> - 1269 records not for adults and removed.
#> - 7157 final records.

jaccard <- species_jaccard(d) # species similarity

net <- lapply(unique(d$foodweb.name), \(fw) { # network properties
  A <- adjacency(d[d$foodweb.name == fw, ])
  return(network_metrics(A))
})
net <- do.call(rbind, net) # concatenate into data.frame

pca <- network_pca(net, axes = 2) # PCA using first 2 axes
#> - Some network properties have NA values and will be omitted:
#>     meanTL
#>     maxTL
#>     sdTL
#>     frOmn
#> - Cumulative variance explained: 0.707
biplot(
  pca,
  scale = FALSE,
  col = c("dodgerblue", "grey20"),
  xlabs = rep("*", nrow(pca$x)),
  cex = 1.5,
  xlab = "PC1: 45%",
  ylab = "PC2: 26%"
)
```



```
sort(abs(pca$rotation[, 1]), decreasing = TRUE) # suggests connectance
#>      frI      clust      frT connectance      sdVul      fIGPs
#> 0.36095929 0.34943119 0.32520554 0.31486046 0.30729031 0.30517846
#>      CPL      fChains      fForks      sdGen      frB      meanSim
#> 0.29741379 0.29052200 0.27942039 0.22050925 0.17621176 0.12909359
#> fColliders
#> 0.08874363

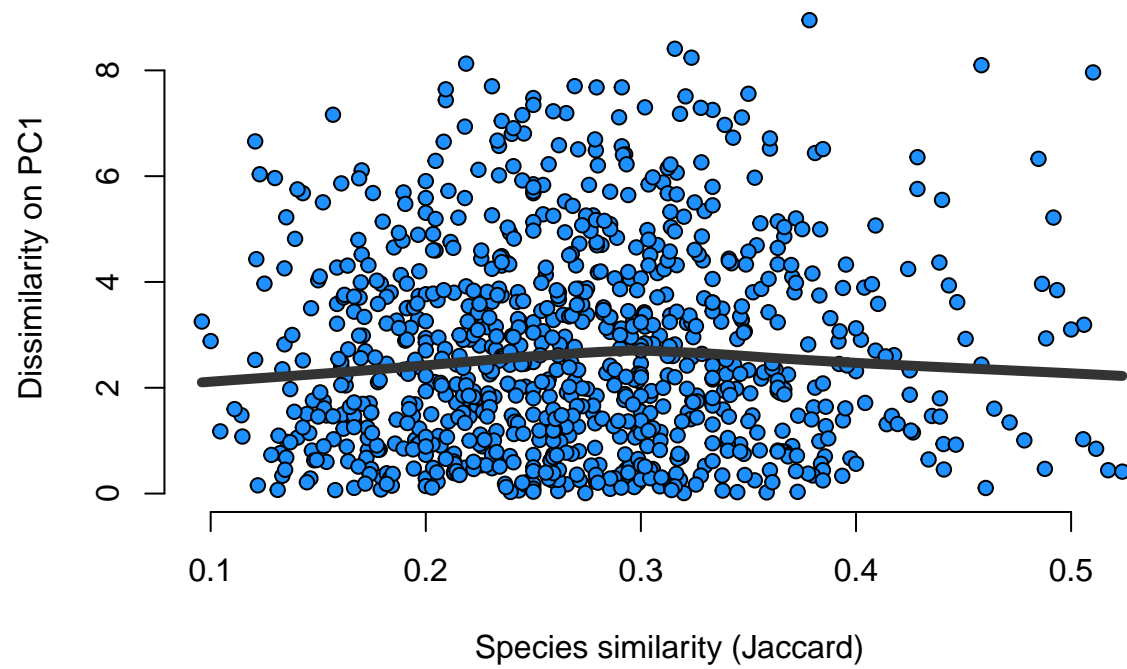
sort(abs(pca$rotation[, 2]), decreasing = TRUE) # suggests frB
#> fColliders      frB      sdGen      fForks      sdVul connectance
#> 0.51419153 0.39775000 0.38078478 0.31503717 0.31292966 0.25792814
#>      fChains      frT      clust      fIGPs      CPL      meanSim
#> 0.22350679 0.21478316 0.14810429 0.13867744 0.13746439 0.08109875
#>      frI
#> 0.07731608

pc1 <- pca$x[, 1]
pc2 <- pca$x[, 2]

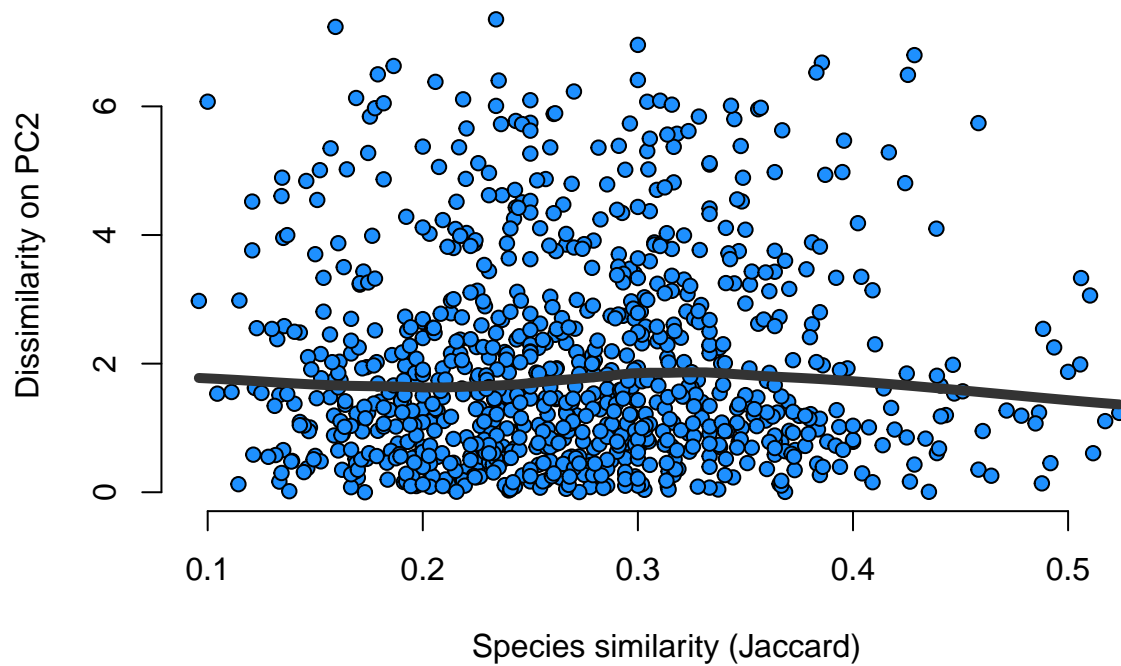
dissimilarity_pc1 <- dist(pc1) # distance in PC1
dissimilarity_pc2 <- dist(pc2) # distance in PC2

scatter.smooth(
  jaccard[lower.tri(jaccard)],
  dissimilarity_pc1[lower.tri(dissimilarity_pc1)],
  pch = 21,
  cex = 1,
  bg = "dodgerblue",
  frame = FALSE,
  xlab = "Species similarity (Jaccard)",
  ylab = "Dissimilarity on PC1",
  lpars = list(col = "grey20", lwd = 5, lty = 1),
```

)

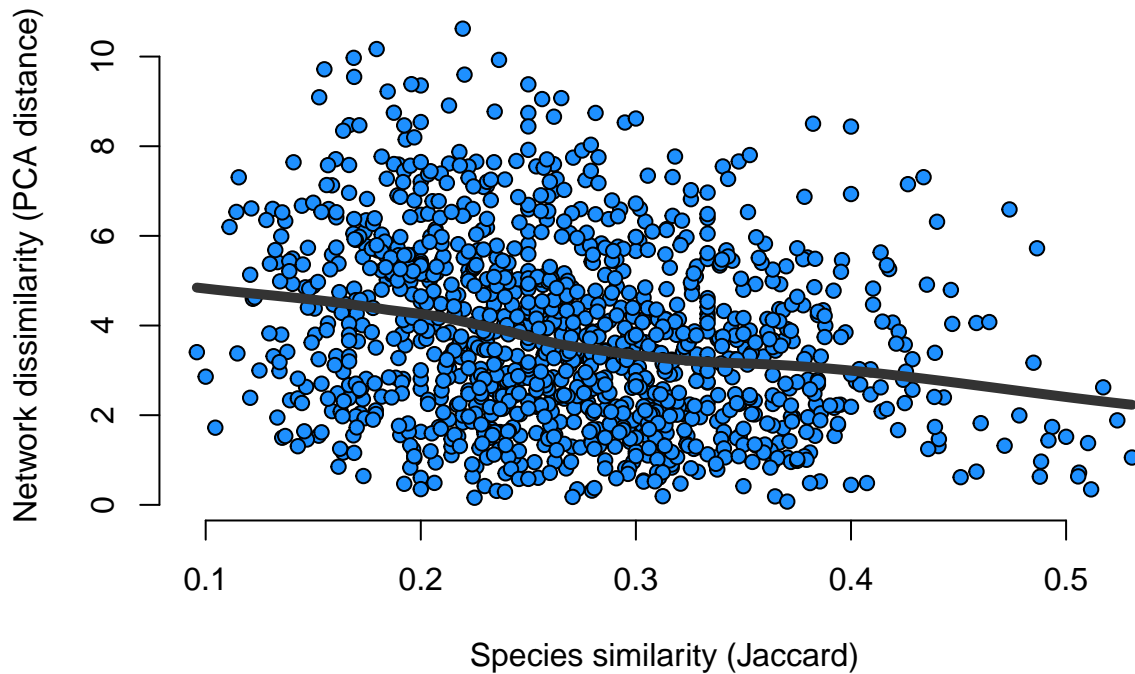


```
scatter.smooth(  
  jaccard[lower.tri(jaccard)],  
  dissimilarity_pc2[lower.tri(dissimilarity_pc2)],  
  pch = 21,  
  cex = 1,  
  bg = "dodgerblue",  
  frame = FALSE,  
  xlab = "Species similarity (Jaccard)",  
  ylab = "Dissimilarity on PC2",  
  lpars = list(col = "grey20", lwd = 5, lty = 1)  
)
```



```
network_distance <- network_dissimilarity(d, axes = 2)
#> - Some network properties have NA values and will be omitted:
#>   meanTL
#>   maxTL
#>   sdTL
#>   frOmn
#> - Cumulative variance explained: 0.707

scatter.smooth(
  jaccard[lower.tri(jaccard)],
  network_distance[lower.tri(network_distance)],
  pch = 21,
  cex = 1,
  bg = "dodgerblue",
  frame = FALSE,
  xlab = "Species similarity (Jaccard)",
  ylab = "Network dissimilarity (PCA distance)",
  lpars = list(col = "grey20", lwd = 5, lty = 1)
)
```



```
logit <- function(x) {
  return( log(x / (1 - x)) )
}

m <- lm(
  scale(as.vector(network_distance[lower.tri(jaccard)])) ~
    -1 + scale(as.vector(jaccard[lower.tri(jaccard)]))
)
summary(m)$adj.r.squared
#> [1] 0.06447864
anova(m)$`Pr(>F)`[1]
#> [1] 1.031743e-19
coef(m)[[1]]
#> [1] -0.2554258
```

## Bootstrapping

The models above have many points, however we only have 50 food webs, which is our true sample size. In other words, we have pseudo-replication. Bootstrapping is a crude way to solve this problem.

```
plot(
  c(0, 0.55), c(0, 11),
  cex = 0,
  frame = FALSE,
  xlab = "Species similarity (Jaccard)",
  ylab = "Network dissimilarity (PCA distance)"
)

n <- 1000 # bootstrap samples
slopes <- rep(NA, n)
p_values <- rep(NA, n)
for (i in seq_len(n)) {
```

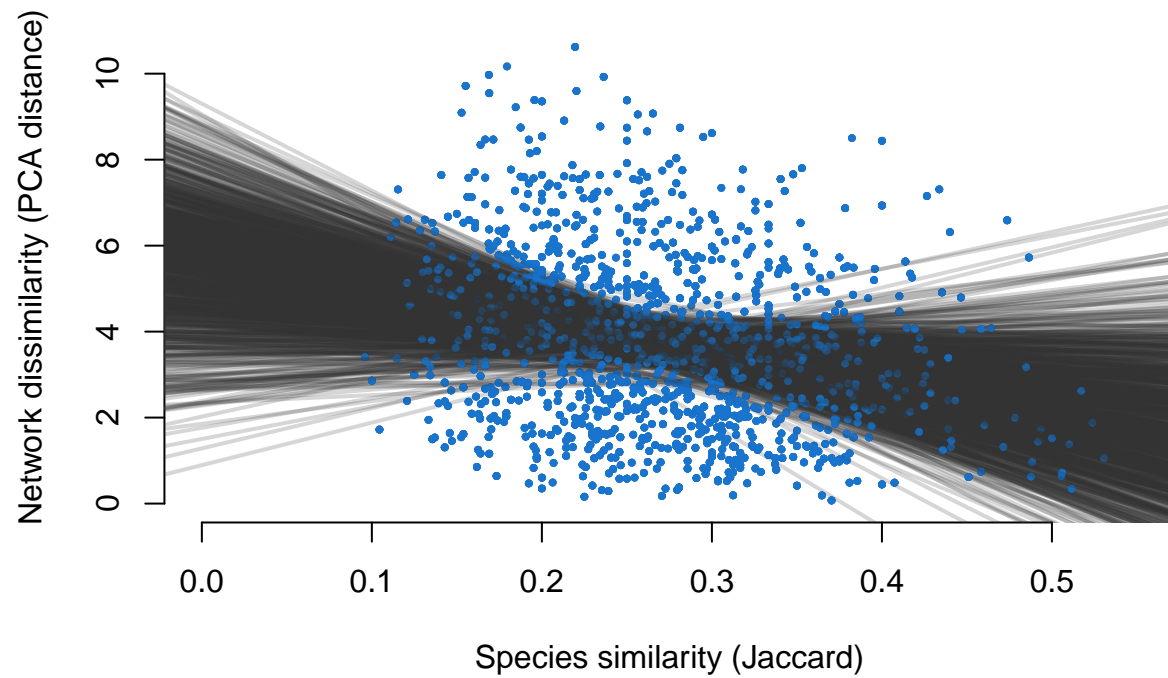
```

# first set of food webs
first_set <- sample(
  rownames(jaccard),
  floor(nrow(jaccard) / 2)
)
# second set of food webs
second_set <- sample(
  setdiff(rownames(jaccard), first_set),
  floor(nrow(jaccard) / 2)
)
comparisons <- cbind(first_set, second_set)
# subset the whole dataframe to keep only 25 pairwise comparisons
boot <- data.frame()
for (j in seq_len(nrow(comparisons))) {
  boot <- rbind(
    boot,
    data.frame(
      jaccard = jaccard[comparisons[j, 1], comparisons[j, 2]],
      network = network_distance[comparisons[j, 1], comparisons[j, 2]]
    )
  )
}

with(boot, points(
  jaccard, network,
  pch = 20,
  cex = .6,
  col = adjustcolor("dodgerblue3", alpha.f = .3)
))

lm <- with(boot, lm(network ~ jaccard))
abline(lm, lw = 2, col = adjustcolor("grey20", alpha.f = .2))
lm <- with(boot, lm(scale(network) ~ -1 + scale(jaccard)))
slopes[i] <- coef(lm)
p_values[i] <- anova(lm)$`Pr(>F)`[1]
}

```



```
summary(slopes)
#>   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#> -0.6546 -0.3747 -0.2645 -0.2474 -0.1234  0.3803
hist(
  slopes,
  breaks = seq(
    floor(min(slopes) * 10) / 10,
    ceiling(max(slopes) * 10) / 10,
    by = .1
  ),
  main = "", xlab = "Slopes",
  col = "dodgerblue3"
)
```

