Biodiversity Soil Germany

```r
library(gatewayr)

d <- subset("study.site", "Biodiversity Exploratory soil food webs")
#>  - 93266 records found.
#>  - 44688 records not for adults and removed.
#>  - 48578 final records.

jaccard <- species_jaccard(d)  # species similarity
jaccard[jaccard > .8] <- NA

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 properies have NA values and will be omitted:
#>      meanTL
#>      maxTL
#>      sdTL
#>      fr0mn
#>  - Cumulative variance explained: 0.773
biplot(
  pca,
  scale = FALSE,
  col = c("orange3", "grey20"),
  xlabs = rep("*", nrow(pca$x)),
  cex = 1.5,
  xlab = "PC1: 49%",
  ylab = "PC2: 29%"
)
```
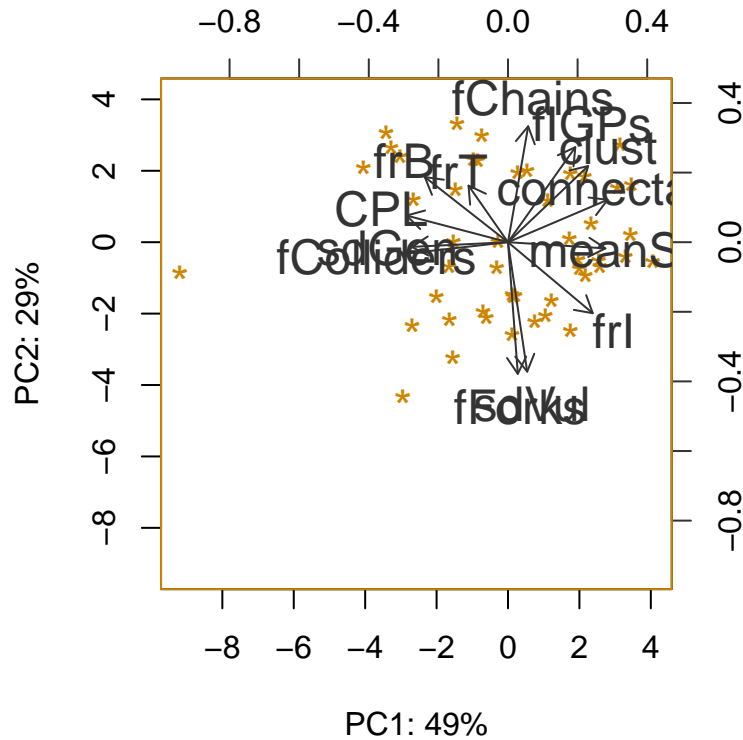
```r
sort(abs(pca$rotation[, 1]), decreasing = TRUE)  # suggests connectance
#>  fColliders          CPL connectance      meanSim        sdGen          frI
#>  0.37781390   0.36419817  0.35758813   0.34982972   0.34607774   0.30440461
#>         frB        clust        fIGPs          frT       fChains        sdVul
#>  0.29870531   0.28786991  0.24176543   0.14187493   0.07253128   0.06896773
#>      fForks
#>  0.03526988
sort(abs(pca$rotation[, 2]), decreasing = TRUE)  # suggests sdVul
#>      fForks        sdVul      fChains        fIGPs        clust          frI
#>  0.47274633   0.46489246  0.41617999   0.34102863   0.27361369   0.25503371
#>         frB          frT  connectance          CPL    fColliders      meanSim
#>  0.23395354   0.20325829  0.15228607   0.09568807   0.04243848   0.02221542
#>       sdGen
#>  0.01682133
pc1 <- pca$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 = "orange3",
  frame = FALSE,
  xlab = "Species similarity (Jaccard)",
  ylab = "Dissimilarity on PC1",
  lpars = list(col = "grey20", lwd = 5, lty = 1),
```
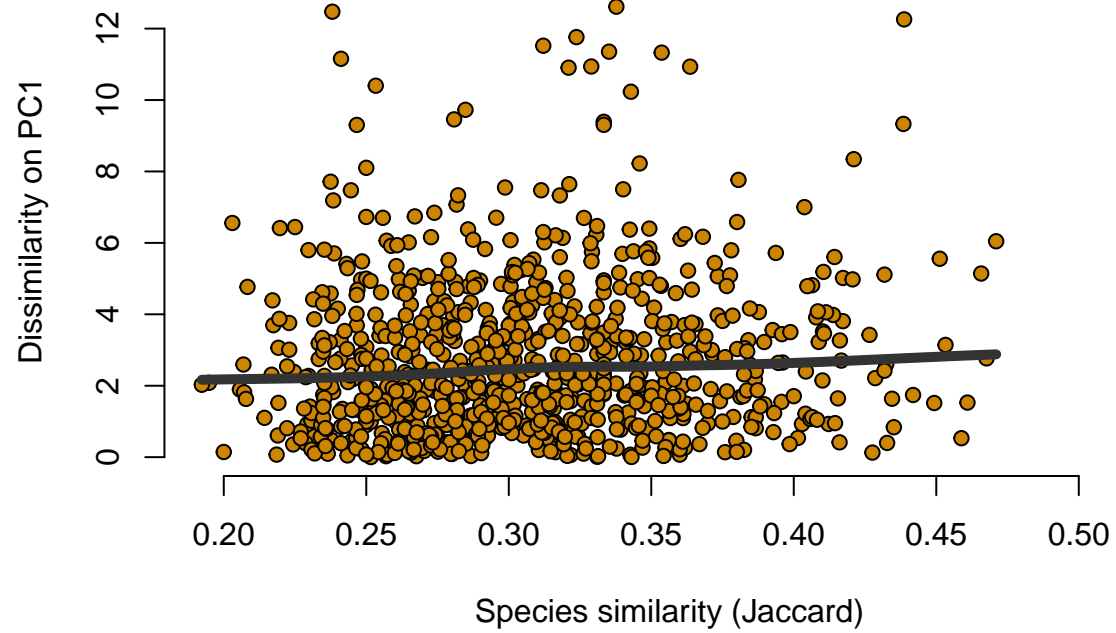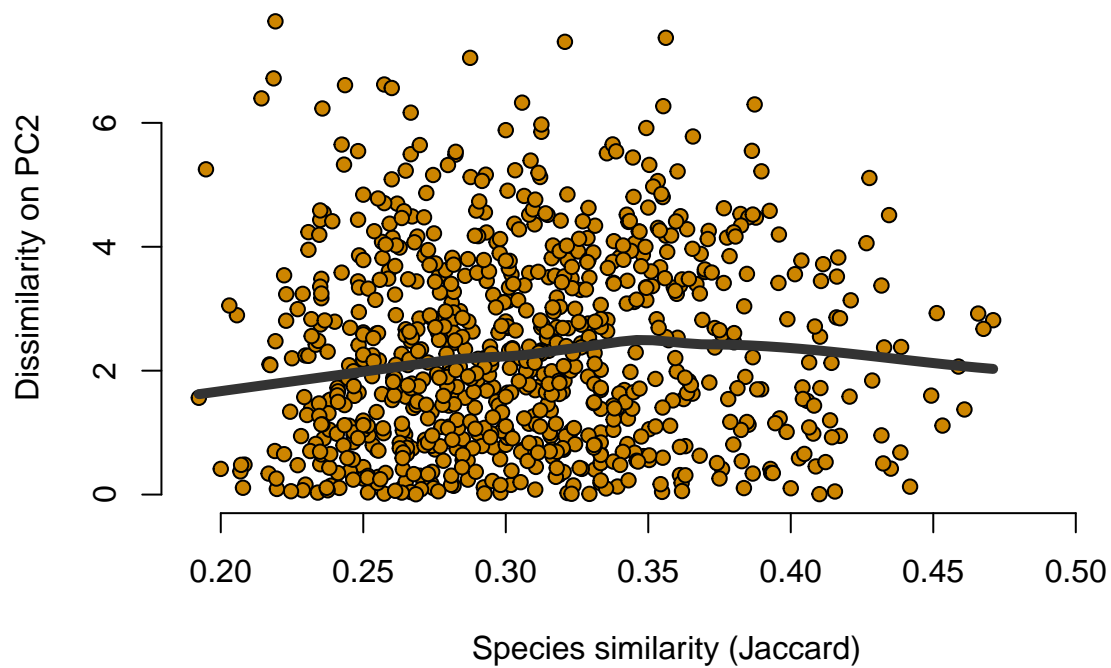
```
)
```



```
scatter.smooth(
  jaccard[lower.tri(jaccard)],
  dissimilarity_pc2[lower.tri(dissimilarity_pc2)],
  pch = 21,
  cex = 1,
  bg = "orange3",
  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 properies have NA values and will be omitted:
#>      meanTL
#>      maxTL
#>      sdTL
#>      frOmn
#>  – Cumulative variance explained: 0.773

scatter.smooth(
  jaccard[lower.tri(jaccard)],
  network_distance[lower.tri(network_distance)],
  pch = 21,
  cex = 1,
  bg = "orange3",
  frame = FALSE,
  xlab = "Species similarity (Jaccard)",
  ylab = "Network dissimilarity (PCA distance)",
  lpars = list(col = "grey20", lwd = 5, lty = 1)
)
```
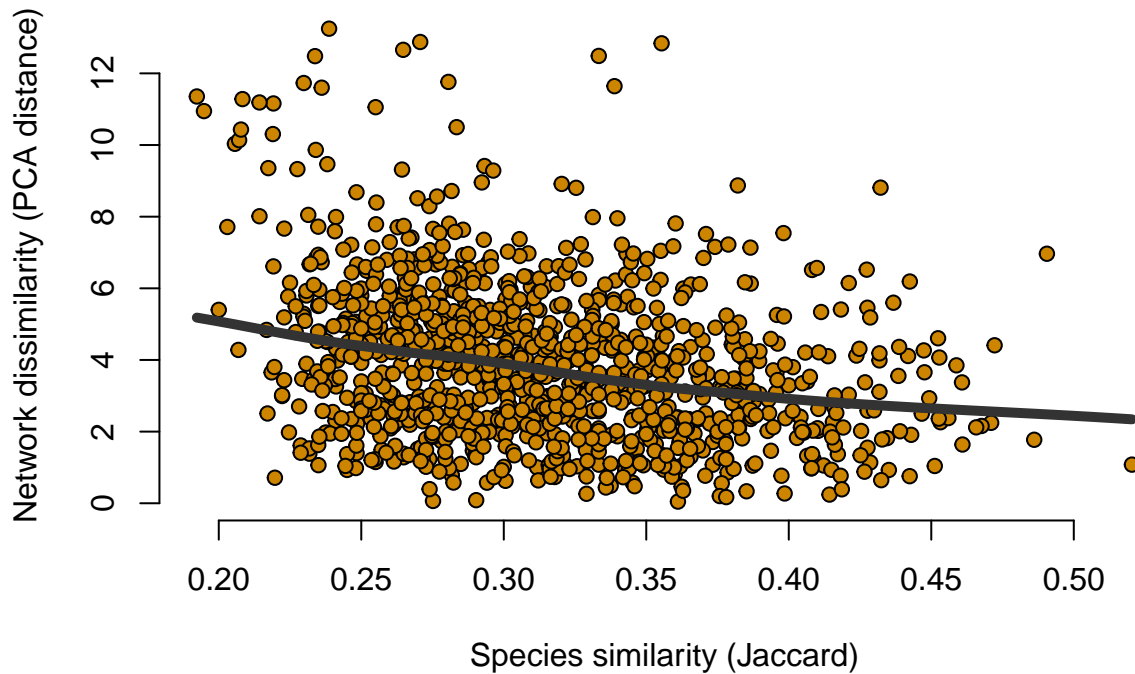
```r
logit <- function(x) {
  return( log(x / (1 - x)) )
}


m <- lm(
  scale(as.vector(network_distance[lower.tri(jaccard)])) ~
    -1 + scale(as.vector(jaccard[lower.tri(jaccard)]))
)
summary(m)$adj.r.squared
#> [1] 0.0852229
anova(m)$`Pr(>F)`[1]
#> [1] 8.094565e-24
coef(m)[[1]]
#> [1] -0.2930426
```

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.

```r
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)) {
```

```r
# 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 = 21,
  cex = .6,
  bg = adjustcolor("orange3", 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]
}
```
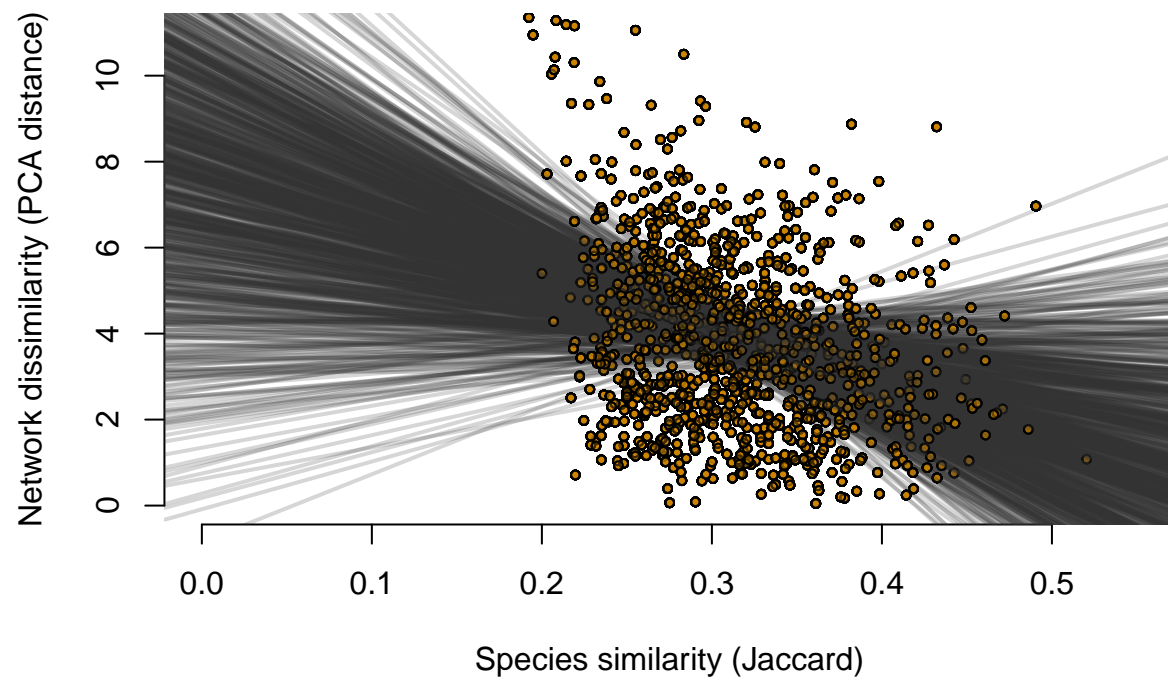
```
summary(slopes)
#>    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#> -0.7762 -0.4405 -0.3165 -0.2964 -0.1721  0.4115
hist(
  slopes,
  breaks = seq(
    floor(min(slopes) * 10) / 10,
    ceiling(max(slopes) * 10) / 10,
    by = .1
  ),
  main = "", xlab = "Slopes",
  col = "orange3"
)
```