

Веб-приложение Shiny для анализа ассоциативности объектов в сетевых структурах

Авторы:

Дмитрий Селезнев

<https://ibiw.ru/index.php?p=staff&id=269>

Владимир Шутиков

<https://stok1946.blogspot.com/>

Настоящее сообщение продолжает обсуждение темы моделирования ассоциативных связей в экологических сообществах с помощью сетей и графов, формируемых на базе пакетов `qgraph` и `igraph` (см. <https://stok1946.blogspot.com/2020/01/blog-post.html> и <https://stok1946.blogspot.com/2022/11/igraph.html>). Научным сотрудником Лаборатории водных беспозвоночных Д.Г. Селезневым разработано Shiny-приложение, которое позволяет проводить анализ видовых ассоциаций на бинарных экологических данных формата «виды-пробы». Приложение развернуто на базе Web-сервера Института биологии внутренних вод РАН (п. Борок) и каждый заинтересованный исследователь может обратиться on-line к его клиентской части по URL <http://apps.ibiw.ru/coobs>, загрузить файл со своими исходными данными и получить табличные и графические файлы с результатами расчетов.

К моменту настоящего сообщения приложение выполняет построение сетей на основе бинарных исходных матриц. Для экологических данных это, например, может быть таблица встречаемости (1) или отсутствия (0) того или иного вида в пробах или местообитаниях. В дальнейшем планируется расширить функциональную часть модуля и осуществлять вычисление мер статистической связи для таблиц произвольного формата данных.

Поскольку адекватность результатов группировки вершин графа напрямую зависит от обоснованности выбора системы весов ребер, то, перед тем, как приступить к описанию web-приложения, остановимся на теоретико-вероятностных аспектах мер связи.

1. Меры ассоциации на основе дискретных распределений вероятности

В экологической литературе описан многочисленный класс мер связи между объектами, представленными в бинарном пространстве хеммингового типа. Таков, в частности, набор коэквивалентных индексов (Сёмкин, 1979), которые, по существу, отражают частоту совместной встречаемости видов i и j по отношению к общей представленности этих видов в сравниваемых выборках, например:

- Жаккара $K_{ij} = a/(a + b + c)$;
- Сьеренсена $K_{xy} = 2a/(2a + b + c)$;
- Охאי (Ochiai) $K_{xy} = a/[(a + b)(a + c)]^{0.5}$,

где a – число совпадающих элементов, $b + c$ – число элементов, уникальных для i и j . Многие авторы делают безнадежные попытки оценить, какие меры "завышают" или "занижают" сходство между выборками и каким коэффициентам следует отдать предпочтение в работе. Однако общей проблемой этих эвристических мер является отсутствие четкого критерия оценки статистической значимости на основе стандартной процедуры проверки гипотез. Исследователь не имеет оснований только по величине рассчитанного индекса утверждать, сходно или различно распределение сравниваемых видов по изучаемой территории, поскольку их кажущая связь может быть обусловлена случайными причинами.

Для корректной оценки взаимной ассоциированности или несовместимости двух произвольных видов целесообразно использовать стандартные статистические тесты. Дж. Вич с соавторами (Veech, 2013, Griffith et al. 2016) показали, что описанная ими вероятностная модель позволяет аналитически оценить вероятность $P(x)$ того, что два произвольно выбранных вида встречаются одновременно с частотой, либо меньшей (P_{lt}), либо большей (P_{gt}), чем теоретически ожидаемая частота их совместного появления. Поскольку для заданного числа местообитаний N существует ограниченное число вариантов,

которыми виды 1 и 2 могут быть распределены между ними, то искомая вероятностная модель сводится, по сути, к задаче комбинаторики.

Д.Г. Селезнев с соавторами (Прокин и др., 2021; Курина и др., 2021) показали, что для оценки вероятности ассоциированности видов можно использовать моделирование эмпирических данных дискретными распределениями X – гипергеометрическим или биномиальным. Если наблюдаемая вероятность $P(x)$ совместного обнаружения двух видов в x из N проб отвечает условию

$$P(x) = P[X \leq x] > (1 - p_{crit}),$$

то виды считаются положительно ассоциированными на уровне статистической значимости p_{crit} . Если справедливо условие

$$P(x) = P[X > x] > (1 - p_{crit}),$$

то принимается предположение об отрицательной ассоциативности видов, т.е. их несовместимости. В противном случае считается, что распределение видов по участкам независимо друг от друга, а частота их совместного появления не отличается от случайной.

Смысл использования теоретических дискретных распределений видится нами в следующем. Ставится задача оценить вероятность получить x раз вид А в пробах, уже содержащих вид Б. То есть в выборке без возвращения число белых шаров m (вид А), число черных шаров $N - m$ (число проб, в которых не встретился вид А), а число вынутых из урны шаров, или испытаний – n (вид Б). В этой интерпретации функция вероятности гипергеометрического распределения описывает вероятность получить в выборке из конечной генеральной совокупности определенное число успехов:

$$p(x) = \frac{\binom{m}{x} \cdot \binom{N-m}{n-x}}{\binom{N}{n}},$$

где $\binom{m}{x}$ – число сочетаний из m по x или биномиальный коэффициент, а n и m – частоты встречаемости каждого из видов.

В качестве другого варианта можно использовать схему испытаний Бернулли и выполнить аппроксимацию наблюдаемых данных биномиальным распределением, к которому дискретное гипергеометрическое распределение сходится при бесконечном увеличении объема генеральной совокупности. Если считать генеральной совокупностью весь потенциально возможный (бесконечный) объем проб из изучаемого биотопа, то вероятность совместного обнаружения двух видов точно в x пробах будет определяться функцией вероятности биномиального распределения;

$$p(x) = \binom{N}{x} \cdot p^x \cdot (1-p)^{N-x},$$

где $p = \frac{m}{N} \cdot \frac{n}{N}$ – апостериорная оценка вероятности.

Для апробации предлагаемой методики оценки значимости ассоциативности будем использовать результаты многолетнего изучения донных сообществ малых и средних рек бассейна Средней и Нижней Волги, а именно таблицу ТАХА, содержащую усредненные по количеству сделанных проб значения численности 147 видов донных организмов, обнаруженных в 132 реках (участках) изучаемого региона. Эти данные уже использовались в наших предыдущих сообщениях, например, <https://stok1946.blogspot.com/2022/11/igraph.html>. Для сравнения с ранее построенными графами осуществим отбор подмножества видов с наиболее выраженной ассоциативностью, которые имеют максимальный коэффициент корреляции Пирсона с другими видами $R > 0.55$.

Загружаем скачанные исходные данные

предварительно загрузив с

http://www.ievbras.ru/ecostat/Kiril/R/Blog/InStab_dat.RData

```

load(file = "InStab_dat.RData")
# Оставляем виды с максимальным коэффициентом корреляции не менее 0.55
library(vegan)
# Логарифмируем значения средних численностей
TAXA_Log <- decostand(TAXA, method="log")
CM <- cor(TAXA_Log)
diag(CM) <- 0

# Находим максимальные коэффициенты корреляции для каждого вида
Mcor = apply(abs(CM), 2, max, na.rm = TRUE)
SpList <- names(Mcor[Mcor > 0.55])
length(SpList)
[1] 46
TAXA46 <- TAXA[, colnames(TAXA) %in% SpList]
TAXA46[TAXA46 > 0] <- 1

```

Выполним оценку силы ассоциативной связи двух видов: *Baetis rhodani* (EpBst.r.) и *I. michaelsoni* (OlIso.m.):

```

# Осуществляем построение дискретных распределений
# для совместной встречаемости двух видов
Test <- TAXA46[, c("EpBst.r.", "OlIso.m.")]
d <- nrow(Test)
q <- sum(Test[, 1]*Test[, 2])
a <- colSums(Test)
k <- max(a)
m <- min(a)
n <- d - m
prob <- (m*k)/d^2
print(c(q, m, n, k, prob))
[1] 7 14 118 36 0.0289
p <- 0.05
# Выполняем проверку P(x) для гипергеометрического распределения
distr1 <- phyper(q, m, n, k, lower.tail=TRUE)
distr2 <- phyper(q-1, m, n, k, lower.tail=FALSE)
print(c(distr1, distr2))
[1] 0.98731157 0.04902536

```

Если мы устанавливаем параметр `lower.tail=TRUE` в `phyper`, то вычисляется для наблюдения x вероятности $P[X \leq x]$, равной или большей значения "недопредставленности" (under-representation) при справедливости нулевой гипотезы. Когда мы устанавливаем параметр `lower.tail=FALSE`, то эта вероятность интерпретируется как $P[X > x - 1]$, при этом из x вычитается 1. В настоящем примере вероятность того, что дискретная случайная величина X , распределенная по гипергеометрическому закону, будет больше, чем значение $x = (q - 1) = 6$, ничтожно мала.

Смысл вычисленных вероятностей легко показать на графике функции гипергеометрического распределения.

```

den <- data.frame(x=seq(0, 12, by = 1), y = phyper(seq(0, 12, by = 1), m, n, k))
den[den$y > 0.85 & den$y < 0.999,]
  x    y
6 5 0.8566500
7 6 0.9509746
8 7 0.9873116
9 8 0.9975566
plot(den, type = "o", pch = 20,
xlab = "Частота совместной встречаемости", ylab = "Накопленная вероятность")
segments(x0 = q, y0 = 0, x1 = q, y1 = distr1,
  col = "cornflowerblue", lwd = 3, lty = "dotted")
segments(x0 = q, y0 = distr1, x1 = 0, y1 = distr1,
  col = "cornflowerblue", lwd = 3, lty = "dotted")

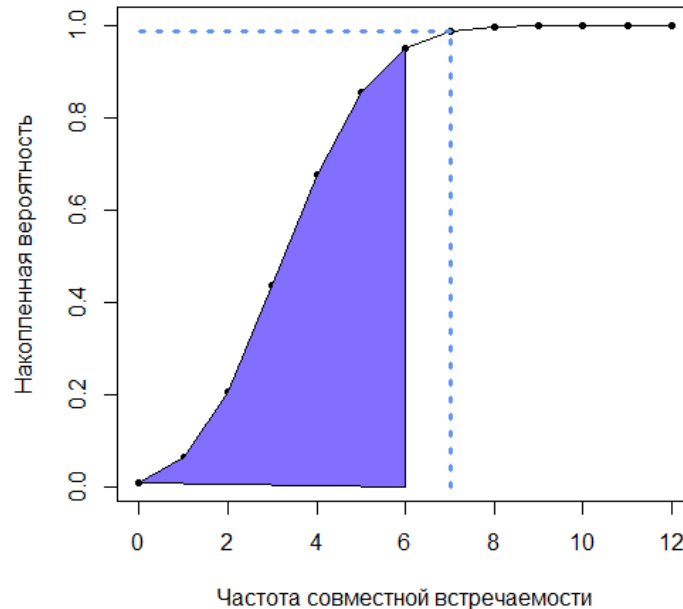
```

```

polygon(c(den$x[den$x <= 6 ], 6),
        c(den$y[den$x <= 6 ], 0),
        col = "slateblue1", border = 1)

```

Из расчетов видно, что верхняя граница доверительного интервала равна 6 частотам и соответствует доверительной вероятности 0.95. На графике цветом залита область значений совместной встречаемости, которая трактуется как случайные совпадения.



Отметим, что к тем же вероятностям приводит точный тест Фишера. В нашем контексте мы рассматриваем частоты "встретился – не встретился" и "в одиночку – совместно", а принятие нулевой гипотезы означает, что вероятность заполнения такой таблицы сопряженности наугад весьма высока. Полученный результат не удивителен, т.к. по замечанию Р. Фишера, подобный тест на независимость частот ведёт к использованию гипергеометрического распределения.

```

mc <- matrix(c(q, m-q, k-q, d- m - k + q), 2, 2)
      [,1] [,2]
[1,]  7  29
[2,]  7  89
fisher.test(mc, alternative='less')$p.value
[1] 0.9873116

```

Использование биномиального распределения приводит к несколько меньшим оценкам вероятности ассоциативности для этих же видов

```

# Выполняем проверку H0 для P(x) биномиального распределения
distr1<-pbinom(q, d, prob, lower.tail=TRUE)
distr2<-pbinom(q-1, d, prob, lower.tail=FALSE)
print(c(distr1, distr2))
[1] 0.96135200 0.08956617

```

Наряду с классическим методом оценки вероятностей по их теоретическим распределениям может быть использован имитационный алгоритм, основанный на байесовском подходе. В качестве оцениваемого параметра примем равномерно распределенную величину одной из наблюдаемых частотных характеристик, а в качестве генеративной функции - формулу вероятности гипергеометрического или биномиального распределения. Полученное апостериорное распределение используем для оценки байесовских доверительных интервалов и вероятностей совместной встречаемости двух

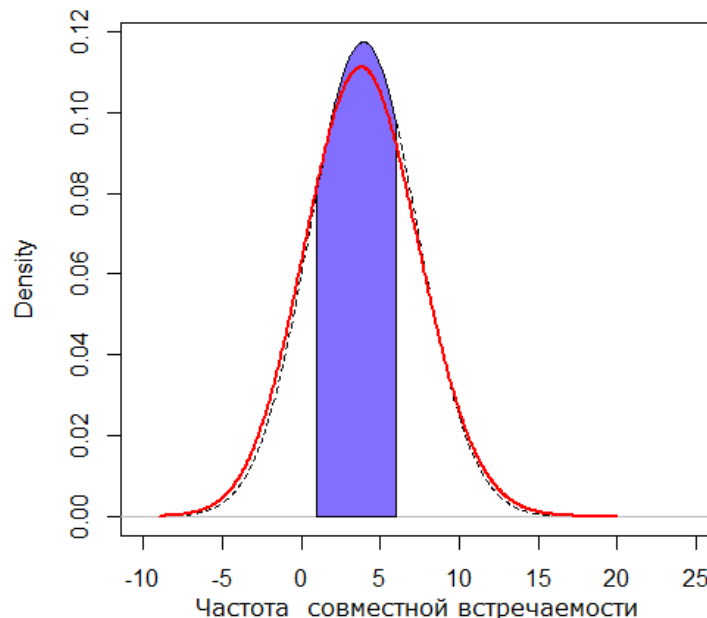
видов, позволяющих принять один из трех выводов: виды встречаются совместно случайно, виды ассоциированы или виды несовместны.

```
Prior <- as.integer(runif(100000, 0, d+1))
pr <- prior[prior>=k]
posterior <- rhyper(length(pr), m, (pr-m), k)
dh1 <- quantile(posterior[which(pr==d)], p)
dh2 <- quantile(posterior[which(pr==d)], (1-p))
print(c(dh1, dh2))
5% 95%
1 6
Fn<-ecdf(posterior[which(pr==d)])
distr1 <- Fn(q)
distr2 <- (1-Fn(q))
print(c(distr1, distr2))
[1] 0.98991173 0.01008827
```

Расчеты показывают, что итоговые выводы, выполненные по обоим алгоритмам, почти всегда совпадают.

Построим для сравнения графики распределения плотности вероятности гипергеометрического и биномиального распределения (цветом залита область байесовских доверительных интервалов совместной встречаемости).

```
plot(den <- density(rhyper(100, m, n, k), 3),
     lty = 2, xlim=c(-10, 25))
l <- min(which(den$x >= dh1))
h <- max(which(den$x < dh2))
polygon(c(den$x[c(l, l:h, h)]),
        c(0, den$y[l:h], 0), col = "slateblue1")
lines(density(rbinom(1000, d, prob), 3), col="red",
      lwd=2, xlim=c(-10, 25))
```



Можно предположить, что гипергеометрическое распределение склонно к гипердиагностике, однако точный вывод о соотношении ошибок 1-го и 2-го рода можно сделать только после детальных имитационных просчетов.

Для оценки вероятностей ассоциативности всех возможных пар из 46 видов и построения сети в приложении Shiny выгрузим матрицу данных в файл csv:

```
write.csv(t(TAXA46), file = "TAXA46.csv")
```

2. Система меню приложения Shiny и порядок выполнения расчетов

Клиентская часть Web-приложения SAA (Species association analysis), к которому можно подключиться по URL <http://apps.ibiw.ru/coobs/>, представляет собой стандартную страницу Интернет-браузера, состоящую из панели управления слева и совокупности вкладок для размещения результатов. На панели управления располагаются различные диалоговые компоненты, позволяющие задать необходимые параметры, и кнопки, запускающие процедуры расчетов.

Начать работу с приложением целесообразно с нажатия кнопки Browse, что позволит загрузить с любого ресурса клиентского компьютера файл в формате CSV с таблицей исходных данных, где виды по умолчанию должны располагаться в строках. Установка флажка Transpose позволяет получить анализ ассоциированности уже не видов, а проб по их видовому составу, что тоже потенциально может быть очень интересно.

	V1.1	V2	V3	V4	V5	V6	V7
BiEug.a.	0	0	0	0	0	0	0
BiHen.h.	0	0	0	0	0	0	0
BiPis.i.	0	0	0	0	0	0	0
BiRiv.r.	0	0	0	0	0	0	0
CeSph.sp	0	0	0	1	0	1	0
ChChi.sr	0	0	0	0	0	0	0
ChCri.sf	0	0	0	0	0	0	0
ChEuk.gr	0	0	0	0	0	0	0
ChHet.ma	0	0	0	0	0	0	0
ChMir.p.	1	1	0	0	0	0	0
ChMnd.ba	1	0	1	0	0	0	0
ChOdo.f.	1	0	1	0	0	0	0

Фиг. 1


Перед началом расчета на панели слева устанавливаются необходимые параметры: функцию используемого распределения (Hypergeometric distribution или Binomial distribution), способ оценки значимости ассоциаций (Confidence intervals of distribution или Credible intervals of Bayesian inference) и критический уровень значимости (Set up significance level). При нажатии кнопки Calculate выполняется отбор ассоциированных пар видов и переход к вкладке **Result**.

Show:

All 407 pairs

316 positively associated pairs

91 negatively associated pairs

 Download

Show entries Search:

SP1	SP2	N1	N2	N1N2	p.lt	p.gt	res
BiEug.a.	BiHen.h.	14	14	9	1	0	1
BiRiv.r.	UnUni.p.	10	23	9	1	0	1
CeSph.sp	ChPse.s.	26	42	21	1	0	1
CeSph.sp	ChPtt.sp	26	46	20	1	0	1
CeSph.sp	EpCsn.b	22	26	15	1	0	1
CeSph.sp	OIOph.s.	26	32	16	1	0	1

Фиг. 2

На вкладке располагается таблица со списком положительно и отрицательно ассоциированных пар (Positive/Negative associated pairs); каждый из таких блоков может отображаться в отдельности. Таблица выводится на экран заданными порциями (entries). Этот список может быть выгружен и сохранен на клиентском компьютере в формате CSV-файла (кнопка Download).

В таблице результатов анализа SP1 и SP2 – названия видов в паре; N1 и N2 – встречаемость первого и второго видов, N1N2 - их совместное наблюдаемое появление; p.lt – вероятность того, что, согласно заданного закона распределения, виды будут одновременно обнаружены с меньшей частотой, чем наблюдаемое значение N1N2; p.gt – вероятность того, что виды будут одновременно обнаружены в том же или большем, чем наблюдаемая частота (или значения кумулятивной функции распределения). При нажатии на заголовок столбца выполняется сортировка по убыванию или возрастанию выбранного столбца.

Каждому столбцу внизу таблицы соответствует поле для фильтрации списка по заданному значению. Например, если мы в поле столбца SP1 введем строку "EpVst.r.", то мы выберем пары, связанные с *Baetis rhodani* (см. Фиг. 3). В этом списке присутствует пара связи этого вида с OIIsom., расчет вероятностей для которой был подробно рассмотрен в разделе 1. После построения графа в этой таблице появляются два новых столбца с номерами кластеров для SP1 и SP2 и по нижним поисковым полям удобно выделять для каждой группы список всех как внутрикластерных, так и межкластерных связей. Эти поля также экспортируются в CSV-файл. Т.к. вид может быть как первым в паре (SP1), так и вторым (SP2), то для фильтрации по всем связям одного вида удобно пользоваться поиском сверху таблицы (Search).

Если переключить компонент типа "радиобокс", то можно получить отдельно список отрицательно ассоциированных видов (см. Фиг. 4). Если на панели управления выбрать для оценки значимости связи функцию биномиального распределения, то число ассоциированных пар уменьшится с 407 до 301 (см. Фиг. 5).

SP1	SP2	N1	N2	N1N2	p.lt	p.gt	res
EpBst.r.	LiDir.b.	36	37	23	1.00000	0.00000	1
EpBst.r.	SiSim.sp	28	36	21	1.00000	0.00000	1
EpBst.r.	TrHyd.p	30	36	26	1.00000	0.00000	1
EpBst.r.	EpEpl.i.	15	36	12	1.00000	0.00001	1
EpBst.r.	OIPrt.h.	25	36	13	0.99931	0.00315	1
EpBst.r.	OISty.la	32	36	14	0.99482	0.01658	1
EpBst.r.	MeSia.sp	12	36	7	0.99670	0.01792	1
EpBst.r.	Ollso.m.	14	36	7	0.98731	0.04903	1

Showing 1 to 8 of 8 entries (filtered from 316 total entries) Previous 1 Next

Фиг. 3

Show:

- All 407 pairs
 316 positively associated pairs
 91 negatively associated pairs

 Download

Show entries

Search:

SP1	SP2	N1	N2	N1N2	p.lt	p.gt	res
ChPse.s.	OIPrt.h.	25	42	4	0.04541	0.98695	-1
ChEuk.gr	ChPse.s.	21	42	3	0.04686	0.98818	-1
ChPol.n.	EpBst.r.	36	98	22	0.03154	0.98888	-1
ChPol.n.	ChPro.o.	39	98	24	0.02777	0.99022	-1

Фиг. 4

Show:

- All 301 pairs
 248 positively associated pairs
 53 negatively associated pairs

 Download

Фиг 5

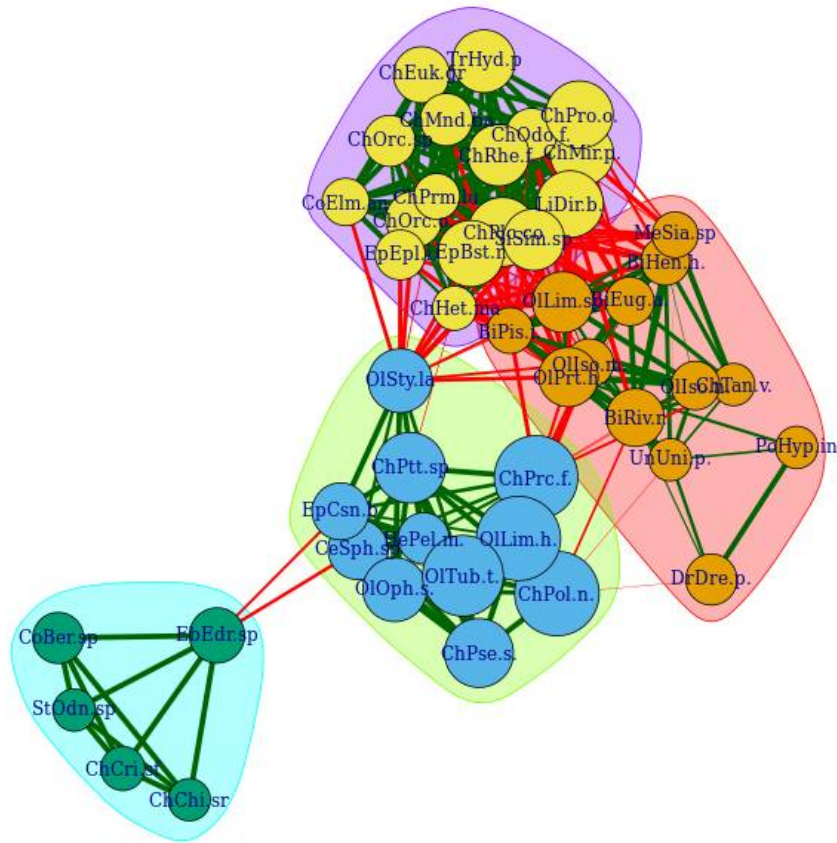
<p>Graph parameters</p> <p>Node size factor <input type="text" value="5"/></p> <p>Edge thickness factor <input type="text" value="5"/></p> <p><input checked="" type="checkbox"/> Weighted edges clustering</p> <p>Edges weight factor: <input type="range" value="10"/></p> <p><input checked="" type="checkbox"/> Species names abbreviations</p> <p>Nodes position method <input type="text" value="Kamada-Kawai (fixed)"/></p> <p>Nodes clustering method <input type="text" value="Multi-level optimization of modularity"/></p> <p><input type="button" value="Draw and analyze graph"/></p>	<p>Сеть строится с использованием пар видов, статистически значимо связанных между собой (положительно, отрицательно или обоих). Для настройки конфигурации графа в нижней части панели слева располагается набор компонентов управления параметрами его построения. Диаметр кружков для вершин пропорционален логарифму частоты встречаемости данного вида, а толщина линий для ребер пропорциональна вероятности $P(x)$. Обе эти величины умножаются на постоянный множитель (factor), который можно изменить. Названия видов в кружках можно заменить их сокращенными уникальными аббревиатурами.</p> <p>Для поиска взаимного расположения вершин графа, обеспечивающего минимум пересечений ребер, могут применяться "силовые" алгоритмы размещения Камада-Кавай (Kamada, Kawai) и Фрухтермана-Рейнгольда (Fruchterman, Reingold). Еще одна возможность – использовать стандартную процедуру многомерного шкалирования (MDS).</p>
--	--

Для оптимальной визуализации и интерпретации графа выполняется группировка его вершин, исходя из условий максимальной компактности и связности (в графоаналитической терминологии – "выделение сообществ"). Здесь проблема заключается в выборе подходящего алгоритма кластеризации вершин графа из некоторого числа рекомендуемых к использованию (Brandes et al., 2008; Коломейченко и др., 2016). В приложении SAA предоставлена возможность выбрать один из 7 возможных методов "Optimal community structure", "Affinity propagation", "Multi-level optimization of modularity", "Greedy optimization of modularity", "Propagating labels", "Short random walks", "Edge betweenness". Все они, в той или иной степени, основаны на анализе модулярности – количественном параметре, оценивающим, насколько плотность внутригрупповых связей больше по сравнению с некой плотностью при справедливости нулевой гипотезы, т.е. при случайном распределении по вершинам такого же числа ребер. Нажатие кнопки Draw and analyze graph вызывает построение и анализ графа.

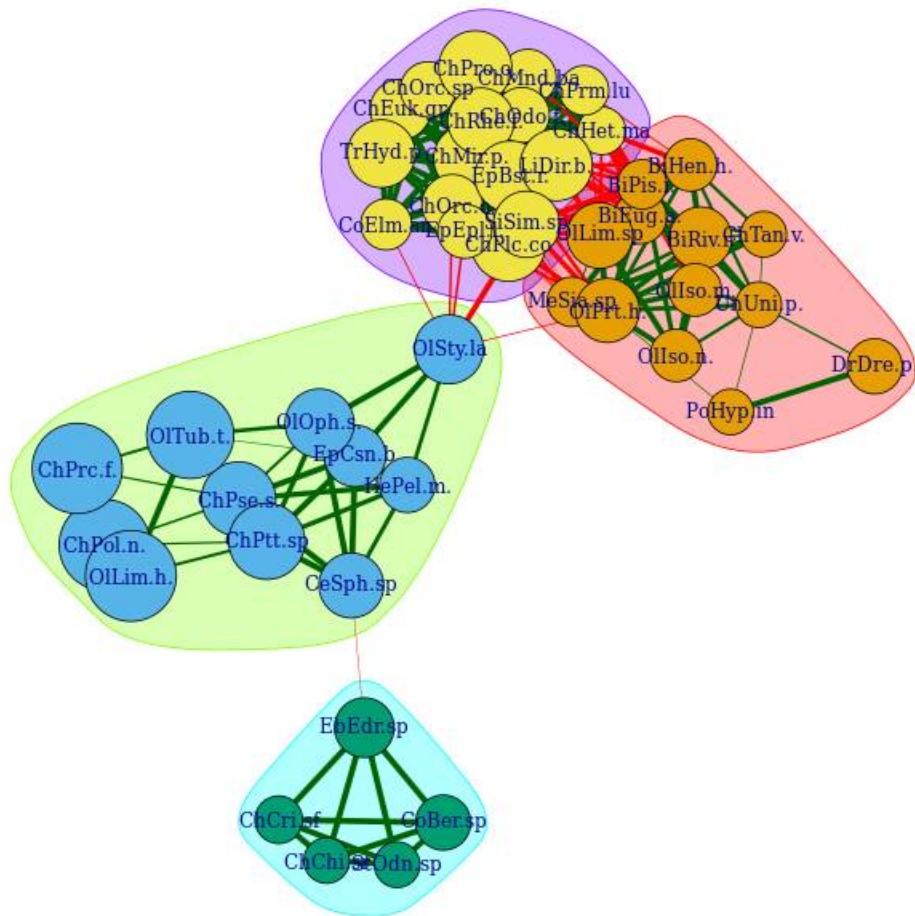
На фиг. 6 представлен граф, построенный на основе пар видов, сформированных функцией гипергеометрического распределения, и использующий алгоритм Affinity propagation – единственный умеющий учитывать негативные связи в качестве отрицательных весов ребер. Для сравнения фиг. 7 показывает граф, использующий функцию биномиального распределения и алгоритм кластеризации Optimal community structure, который добивается максимальной модулярности, но время его работы экспоненциально зависит от числа ребер.

Изображение построенного графа может быть сохранено в векторном формате в файле PDF или в стандартном растровом формате PNG. Таблица, включающая координаты вершин графа, номер кластера, к которому отнесен каждый узел и число его связей (т.е. степень ассоциативности вида), также может быть выгружена в CSV-файл для последующего использования иными методами анализа.

Отметим, что уменьшение уровня значимости уже после выполнения расчетов, автоматически приводит к фильтрации результатов по новому p_{crit} и перерисовку графа. Это удобно для определения, какие связи "отваливаются" при более жестком критерии их отбора.



Фиг. 6.



Фиг. 7

Очевидно, что выделенные кластеры будут неодинаковы как по средней плотности связей (Connectivity), так и по средней силе связи, приходящейся на каждый кластер (Mean connection strength). Плотность связей в кластере или связность можно вычислять как отношение числа реализованных связей (Inner connections) к общему числу возможных связей в группе $C_n = 2k/(n \cdot (n-1))$, где k – число реализованных связей в группе, n – число видов в группе (Cluster members). Общую силу связности (Total connectivity strength) в группе можно оценивать как среднее от уровня значимости ребер, приведенного к единице ($s_{i-} = 1 - P(x) / p_{crit}$ для отрицательно ассоциированных пар и $s_{i+} = 1 - (1 - P(x)) / p_{crit}$ для положительно ассоциированных), дополненного нулями на каждую незначимую связь. Предположим, в кластере из трех видов при пороговом уровне значимости 0.05 выделены две пары положительно ассоциированных видов и значения s_{i+} для этих связей составили 0.04 и 0.01. Тогда общая сила связности группы будет $(0.2+0.8+0)/3=0.33$.

Все эти расчеты представлены на вкладке **Graph analysis**:

Method	Data	Result	Graph			
			Graph analysis			
			CI_1	CI_2	CI_3	CI_4
Cluster members			13.00	11.00	5.00	17.00
Inner connections			44.00	29.00	10.00	120.00
Cross connections			39.00	7.00	1.00	43.00
Connectivity			0.56	0.53	1.00	0.88
Mean connection strength			0.72	0.73	0.99	0.85
Total connectivity strength			0.41	0.39	0.99	0.75

Список цитируемой литературы

Сёмкин Б.И. Эквивалентность мер близости и иерархическая классификация многомерных данных // Иерархические классификационные построения в географической экологии и систематике. Владивосток: ДВНЦ АН СССР, 1979. С. 97-112.

Wolda H. Similarity indices, sample size and diversity. // Oecologia. 1981. V. 50. P. 296-302. <http://dx.doi.org/10.1007/BF00344966>

Veech J.A.. A probabilistic model for analysing species co - occurrence // Global Ecology and Biogeography. 2013. V. 22, N 2. P. 252-260.

Griffith D.M., Veech J.A., Marsh C.J. cooccur: Probabilistic Species Co-Occurrence Analysis in R. // Journal of Statistical Software. 2016. V. 69. Code Snippet 2.

Курина Е.М., Селезнев Д.Г., Шерышева Н.Г. Распространение чужеродных видов макрозообентоса и их ценоотические комплексы в камских водохранилищах // Российский журнал биологических инвазий. 2021. Т. 14. № 4. С. 85-96. <https://doi.org/10.35885/1996-1499-2021-14-4-85-96>

Прокин А.А., Селезнев Д.Г., Цветков А.И. Влияние факторов среды на межгодовые изменения макрозообентоса пойменных озер. Трансформация экосистем. 2021. Т. 4. № 2 (12). С. 65-77. <https://doi.org/10.23859/estr-210306>

Brandes U., Dellling D., Gaertler M., Gorke R., Hoefler M., Nikoloski Z., Wagner D. On Modularity Clustering // IEEE Transactions on Knowledge and Data Engineering. 2008. V. 20(2). P. 172-188,.

Коломейченко М.И., Поляков И.В., Чеповский А.А., Чеповский А.М. Выделение сообществ в графе взаимодействующих объектов // *Фундамент. и прикл. матем.* 2016, 21, №3. С. 131–139