

# КРАТКИЕ СООБЩЕНИЯ

Самарская Лука: проблемы региональной и глобальной экологии.  
2023. – Т. 32. – № 3. – С. 50-53.

УДК 574.9

DOI 10.24412/2073-1035-2023-10498

## ПРОГРАММА ОПРЕДЕЛЕНИЯ И АНАЛИЗА АССОЦИИРОВАННОСТИ ВИДОВ В ЭКОЛОГИЧЕСКИХ СООБЩЕСТВАХ

© 2023 Д.Г. Селезнев

Институт биологии внутренних вод им. И.Д. Папанина РАН, пос. Борок, Некоузский район, Ярославская область (Россия)

Поступила 17.10.2023

*Аннотация.* Описана разработанная автором методика выделения ассоциаций видов, которая отличается простотой реализации, устойчивостью и воспроизводимостью результатов. Недостатком методики является использование только качественных данных о наличии или отсутствии вида в пробе, а не его количественных характеристик.

*Ключевые слова:* видовые ассоциации, статистическая методика, общая сила связности.

Важной частью анализа качественных экологических данных является выявление видовых ассоциаций, обусловленных прямыми (трофические, топические, форические и др.) и косвенными (общность биотопических условий, характер питания, защита от хищников) межвидовыми отношениями (Odum, Barrett, 1971). С этой целью было разработано несколько специализированных матричных индексов, для определения уровня значимости которых предлагается использовать одну из девяти рандомизационных процедур построения теоретического распределения индекса (Gotelli, 2000). Однако в экологических исследованиях задача определения видовых ассоциаций решается сравнительно редко из-за неочевидности математического аппарата и сложности его реализации.

Автором была разработана статистическая методика выделения положительных и отрицательных ассоциаций видов, позволяющая избежать использования матричных индексов и рандомизационной процедуры определения их уровня значимости. Для этого на бинарной матрице встречаемости видов в экологических пробах ис-

пользуются дискретные гипергеометрическое или биномиальное распределения, дающие оценку вероятности наблюдаемой совместной встречаемости пар видов. Критическая область функции распределения (confidence intervals of distribution) позволяет принять одно из трех решений: пара видов отрицательно либо положительно ассоциирована или их совместная встречаемость в пробах носит случайный характер. В качестве альтернативы критической области возможно использование байесовского вывода (Bayesian inference), где в качестве оцениваемого параметра может быть взята любая равномерно распределенная наблюдаемая характеристика, например, общее число проб, а в качестве генеративной функции – функция вероятности дискретного распределения.

Предложенная методика реализована в виде web-приложения, размещенного на сервере Института биологии внутренних вод по адресу [http://apps.ibiw.ru/coobs\\_](http://apps.ibiw.ru/coobs_) (рис. 1). Приложение написано на языке R в среде Shiny (Chang et al., 2023) и позволяет проанализировать бинарную пользовательскую матрицу формата «виды – пробы» с использованием двух дискретных распределений с заданной критической областью

---

Селезнев Дмитрий Геннадьевич, науч. сотр.,  
dmitriy@seleznev.name

или с помощью байесовских доверительных интервалов (credible intervals of Bayesian inference). Исходная матрица может быть транспонирована для определения ассоциаций между пробами по видовому со-

ставу или, другими словами, определения фаунистического сходства. Полученные результаты в виде списка ассоциированных видов или проб могут быть сохранены в формате CSV для дальнейшего анализа.

Species association analysis

Method Data Result Graph Graph analysis

Choose CSV File

Browse... TAXA46.csv **1**

Upload complete

Analysis parameters

Transpose matrix

Choose method **2**

Binomial distribution

Choose Interval **3**

Confidence intervals of distribution

Set up significance level **4**

0,05

Calculate

Graph parameters

Node size factor

5

Edge thickness factor

5

Weighted edges clustering

Edges weight factor:

0 10 20

Species names abbreviations

Nodes position method **5**

Kamada-Kawai (fixed)

Nodes clustering method **6**

Multi-level optimization of modularity

Draw and analyze graph

Show: **Show:**

All 301 pairs

248 positively associated pairs

53 negatively associated pairs

Download

Show 25 entries Search:

SP1	SP2	N1	N2	NIN2	p.lt	p.gt	res
ChChi.sr	ChCri.sf	10	11	10	1.00000	0e+00	1
ChChi.sr	CoBer.sp	10	16	9	1.00000	0e+00	1
ChCri.sf	CoBer.sp	11	16	10	1.00000	0e+00	1
ChEuk.gr	EpBst.r.	21	36	19	1.00000	0e+00	1
ChRhe.f.	EpBst.r.	28	36	23	1.00000	0e+00	1
ChRhe.f.	TrHyd.p	28	30	21	1.00000	0e+00	1
TrHyd.p	EpBst.r.	30	36	26	1.00000	0e+00	1
ChCri.sf	EbEdr.sp	11	20	10	1.00000	1e-05	1
ChRhe.f.	SiSim.sp	28	28	19	1.00000	1e-05	1
SiSim.sp	EpBst.r.	28	36	21	0.99999	2e-05	1
BiEug.a.	BiHen.h.	14	14	9	1.00000	2e-05	1
ChChi.sr	EbEdr.sp	10	20	9	1.00000	2e-05	1
CoElm.an	EpEpl.i.	13	15	9	1.00000	2e-05	1
StOdn.sp	EbEdr.sp	10	20	9	1.00000	2e-05	1
Ollso.m.	Ollso.n.	14	14	9	1.00000	2e-05	1
OIPrt.h.	OILim.sp	25	27	17	1.00000	2e-05	1
EpCsn.b	CeSph.sp	22	26	15	0.99999	3e-05	1
ChEuk.gr	TrHyd.p	21	30	16	0.99999	3e-05	1
OIOph.s.	OISly.la	32	32	21	0.99999	3e-05	1
PoHyp.in	DrDre.p.	10	16	8	1.00000	3e-05	1

SP1 SP2 N1 N2 NIN2 p.lt p.gt res

**Рис. 1. Интерфейс web-приложения Species association analysis. Цифрами обозначены: 1 – загрузка пользовательского файла, 2 – выбор статистического распределения, 3 – выбор метода определения видовых ассоциаций, 4 – пороговый уровень значимости, 5 – выбор алгоритма расположения (layout) узлов графа, 6 – выбор алгоритма кластеризации узлов графа**  
**Fig. 1. Interface of the Species association analysis web application. The numbers indicate: 1 – loading a user file, 2 – choosing a statistical distribution, 3 – choosing a method for determining species associations, 4 – threshold level of significance, 5 – choosing an algorithm for the layout of graph nodes, 6 – choosing an algorithm for clustering graph nodes**

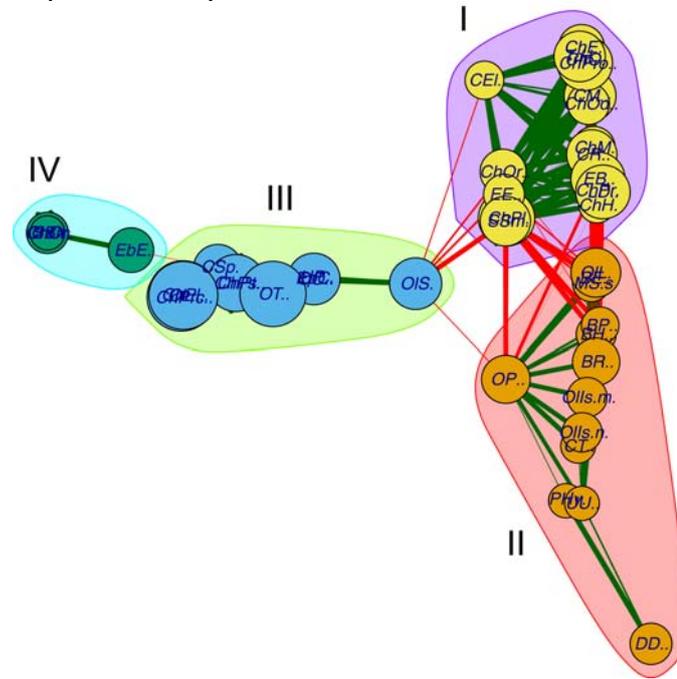
Для визуализации результатов анализа с помощью пакета igraph (Csardi, Nepusz, 2006) приложение строит неориентированный граф, где в качестве узлов представлены виды из списка ассоциированных, а в качестве ребер – выявленные ассоциации между ними. Размер маркера узла пропорционален частоте встречаемости вида, а толщина ребра пропорциональна силе свя-

зи через настраиваемые коэффициенты (рис. 2).

Взаимное расположение узлов на графе может быть определено с помощью трех алгоритмов (выпадающий список 5 на рис. 1), одним из которых является метод многомерного шкалирования (MDS). Этот метод часто строит не оптимальный с визуальной точки зрения граф, но расположе-

ние вершин в нем несет дополнительную информацию о представленных фаунистических комплексах. На рис. 2 в кластер I входят реофильные пресноводные виды, в кластер II – лимнофильные пресноводные, в кластер III – лимнофильные эвригалин-

ные, в кластер IV – лимнофильные галофильные виды. Таким образом, на графе по оси абсцисс расположен градиент солености воды, а по оси ординат – градиент скорости течения.



**Рис. 2. Граф положительно ассоциированных видов макрозообентоса средних и малых рек бассейна Нижней Волги (Golovatyuk et al., in press). Выделение ассоциаций: биномиальное распределение с 5% верхней критической областью. Расположение вершин методом MDS**  
**Fig. 2. Graph of positively associated macrozoobenthos species of medium and small rivers of the Lower Volga basin [Golovatyuk et al. In press]. Identification of associations: binomial distribution with 5% upper critical region. Location of vertices using MDS method**

Группировка узлов в кластеры также осуществляется с помощью различных алгоритмов, использующих принцип максимизации модулярности графа (выпадающий список 6 на рис. 1). Отдельно можно выделить алгоритм Optimal community structure, добивающийся максимально возможной модулярности средствами целочисленного программирования, но имеющий экспоненциальную сложность (Brandes et al., 2008). Если в визуализацию включены отрицательно ассоциированные виды, задачей алгоритма кластеризации является максимизация отрицательных и минимизация положительных связей между кластерами. Эту задачу успешно решает доработанный алгоритм Affinity propagation, корректно обрабатывающий отрицательные веса ребер (Трааg, Bruggeman, 2008), тогда как остальные алгоритмы считают их нулевыми.

В приведенном рис. 2 кластеризация по-

ложительно ассоциированных видов осуществлялась методом многоуровневой оптимизации модулярности (Multi-level optimization of modularity) (Blondel et al., 2008). Кроме указанных выше солености и скорости течения, распределенных по ортогональным осям графа, алгоритм кластеризации учел температуру воды и тип грунта, не имея о них явных данных. Так от кластера I к кластеру IV температура воды в летний период последовательно повышается с 16,9 °C до 27°C, а грунт меняется от гравийно-галечного через заиленный песок и глину до черного ила в кластере IV.

В web-приложении на вкладке «Graph analysis» выделенные кластеры анализируются по плотности связей и их средней силе в кластере. Интегральная характеристика, названная «общей силой связности» (total connectivity strength) объединяет относительное число и среднюю силу связей

в кластере в единую нормированную характеристику, позволяющую сравнивать кластеры между собой.

Предложенная методика отличается простотой реализации (для проверки степени ассоциированности видов возможно использование MS Excel 2010+), а также устойчивостью и воспроизводимостью ре-

зультатов. Недостатком методики является использование только качественных данных о наличии или отсутствии вида в пробе, а не его количественных характеристик: численности, биомасса и др.

Работа выполнена в рамках государственного задания № 121051100109-1.

## СПИСОК ЛИТЕРАТУРЫ

### Reference List

**Blondel V.D., Guillaume J.-L., Lambiotte R., Lefebvre E.** Fast unfolding of communities in large networks // *Journal of Statistical Mechanics: Theory and Experiment*. 2008. P10008. P. 1-12. DOI: 10.1088/1742-5468/2008/10/P10008.

**Brandes U., Delling D., Gaertler M., Gorke R., Hofer M., Nikoloski Z., Wagner D.** On Modularity Clustering // *IEEE Transactions on Knowledge and Data Engineering*. 2008. Vol. 20, iss. 2. P. 172-188. DOI: 10.1109/TKDE.2007.190689.

**Chang W., Cheng J., Allaire J., Sievert C., Schloerke B., Xie Y., Allen J., McPherson J., Dipert A., Borges B.** shiny: Web Application Framework for R. R package version 1.7.4.1. 2023. [https://CRAN.R-project.org/package=shiny].

**Csardi G., Nepusz T.** The igraph software package for complex network research // *Internation-*

*Journal of complex systems*. 2006. Vol. 1695 (5). P. 1-9. [https://igraph.org].

**Golovatyuk L.V., Seleznev D.G., Kurina E.M.** Analysis of macrozoobenthos species associations in the eastern european plain under conditions of changing climatic zones // *Hydrobiologia*. (In press).

**Gotell N.J.** Null model analysis of species co-occurrence patterns // *Ecology*. 2000. Vol. 81, iss. 9. P. 2606-2621. DOI: 10.1890/0012-9658(2000)081[2606:NMAOSC]2.0.CO;2.

**Odum E.P., Barrett G.W.** *Fundamentals of Ecology*. Third edition. Philadelphia: Saunders, 1971. 574 p.

**Traag V.A., Bruggeman J.** Community detection in networks with positive and negative links // *Physical Review E*. 2008. Vol. 80, iss. 3. P036115. DOI: 10.1103/PhysRevE.80.036115.

## DEFINITION AND ANALYSIS PROGRAM ASSOCIATIONS SPECIES IN ECOLOGICAL COMMUNITIES

© 2023 D.G. Seleznev

Institute of Biology of Inland Waters named I.D. Papanin RAS, Borok village, Nekouzsky district, Yaroslavl region (Russia)

*Annotation.* The method developed by the author for identifying species associations is described, which is characterized by ease of implementation, stability and reproducibility of results. The disadvantage of the technique is the use of only qualitative data on the presence or absence of a species in a sample, and not its quantitative characteristics.

*Key words:* species associations, statistical methodology, total connectivity strength.