

E. MARCZEWSKI i H. STEINHAUS (Wrocław)

O ODLEGŁOŚCI SYSTEMATYCZNEJ BIOTOPÓW

1. Podobieństwo zespołów roślinnych. Przyrodnicy spotykają się w różnych badaniach z problematem podobieństwa systematycznego biotopów. Np. botanicy uważają za podobne takie zespoły roślinne, które mało się różnią co do gatunków w nich występujących. Tak więc dwa lasy takie, że każdy gatunek spotykany w jednym rośnie także i w drugim lesie i, na odwrót, każdy rosnący w drugim występuje w pierwszym, uważają za podobne w najwyższym stopniu, natomiast takie dwa lasy, które nie wykazują ani jednego gatunku wspólnego — za zupełnie niepodobne. Usiłowali oni ująć to pojęcie podobieństwa ilościowo. Nazwijmy a liczbę gatunków występujących w zespole roślinnym \mathcal{A} , nazwijmy b liczbę gatunków występujących w zespole \mathcal{B} , nazwijmy zaś literą w liczbę gatunków wspólnych obu zespołom. Wtedy wzór

$$(1) \quad s = \frac{1}{2}(w/a + w/b)$$

określa podobieństwo systematyczne s zespołów \mathcal{A} i \mathcal{B} . W pierwszym z dwóch opisanych skrajnych przypadków mamy wtedy $s = 1$, a w drugim $s = 0$.

Jeden z autorów niniejszej pracy zapoznał się z tą definicją z końcem wojny i skrytykował ją jak następuje: Koło Nicei rosną na obszarze 10-morgowym same fiołki uprawiane ze względu na olejek do perfum, a koło Jasła jest 10-morgowy las, w którym wśród 1000 gatunków roślinnych można też znaleźć fiołki. Nazwijmy \mathcal{A} plantację nicejską, \mathcal{B} las podkarpacki i zastosujmy wzór (1). Będzie $a = 1$, $b = 1000$, $w = 1$, $s = \frac{1}{2}(1 + 0,001) = 50,05\%$, a więc podobieństwo obu obszarów jest niemal dokładnie 50%. Ale podobieństwo s obliczone według (1) nigdy nie jest większe niż 100%, bo w/a nigdy nie przekracza 1. Dlatego 50% należy uważać za niemałe podobieństwo — tymczasem jest widoczne, że podobieństwo roślinne obu zespołów jest znikome. Jest jednak jeszcze inna wada wzoru (1). Intuicyjne pojęcie podobieństwa chce, żeby silne podobieństwo \mathcal{A} do \mathcal{B} i \mathcal{B} do \mathcal{C} pozwalało wnosić o jakimś niezbyt małym podobieństwie \mathcal{A} do \mathcal{C} . Weźmy jednak zespół \mathcal{B} i podzielmy go na dwie części \mathcal{A} i \mathcal{C} , tak żeby w \mathcal{A} i \mathcal{C} było po 50 gatunków, ale żeby nie było gatunków wspólnych. Wtedy podobieństwo \mathcal{A} do \mathcal{C} będzie według

(1) zerem, podobieństwo \mathcal{A} do \mathcal{B} będzie 75% i \mathcal{C} do \mathcal{B} 75%. Tu więc występują silne podobieństwa \mathcal{A} do \mathcal{B} i \mathcal{B} do \mathcal{C} , a towarzyszy im zupełny brak podobieństwa \mathcal{A} do \mathcal{C} . Ten drugi paradoks wzmacnia jeszcze argumentację przeciw definicji (1).

2. Inna definicja podobieństwa. Odległość. Podamy teraz inną definicję, wolną od powyższych zarzutów:

$$(2) \quad s = \frac{w}{a+b-w}.$$

Mianownikiem jest tu liczba wszystkich gatunków dostrzeżonych w \mathcal{A} i \mathcal{B} , tj. takich, że każdy występuje w \mathcal{A} lub w \mathcal{B} . Definicja (2) jest rachunkowo wygodniejsza, bo wymaga tylko jednego dzielenia — to nie jest obojętne, gdyż botanicy obliczają podobieństwa setek par obszarów. Po wtóre, podobieństwo plantacji fiołkowej z lasem wypada według (2) 1 promille, zgodnie z intuicją. W przykładzie zespołu \mathcal{C} podzielonego na \mathcal{A} i \mathcal{B} będzie teraz podobieństwo \mathcal{A} do \mathcal{C} 50%, \mathcal{C} do \mathcal{B} 50%, \mathcal{A} do \mathcal{B} 0%; to już nie przeczy intuicji, bo zupełnie niepodobni rodzice \mathcal{A} , \mathcal{B} mogą mieć dziecko \mathcal{C} podobne w 50% do matki i w 50% do ojca i wyda się to każdemu naturalne, natomiast trudno byłoby zrozumieć, żeby dziecko było podobne do każdego z rodziców w 75%, a oni zupełnie niepodobni do siebie.

Przeciw wzorowi (2) można by wysunąć pewien nowy zarzut: gdy w \mathcal{A} rośnie np. jeden tylko gatunek roślin, a \mathcal{B} jest pustynią, gdzie nie rośnie, wówczas według (2) mamy $s = 0$. Oba te zespoły są skrajnie ubogie, czy słusznie więc uznać, że nie ma między nimi żadnego podobieństwa? Chyba jednak słusznie: można przecież powiedzieć, że nie ma podobieństwa między życiem a śmiercią.

Wprowadźmy odległość systematyczną r zespołów: im zespoły są podobniejsze, tym są bliższe, im mniej podobne, tym dalsze (oczywiście, ta odległość nie ma nic wspólnego z wzajemnym położeniem badanych zespołów roślinnych, np. z ich odległością geometryczną). Najprościej będzie można to uczynić pisząc $r = 1 - s$. Tak więc odległość będzie 0, gdy podobieństwo będzie stuprocentowe, będzie 1, gdy podobieństwo będzie zerowe.

Podobieństwu s określoneму wzorem (1) odpowiadać będzie odległość

$$(3) \quad r = 1 - \frac{1}{2} \left(\frac{w}{a} + \frac{w}{b} \right),$$

a podobieństwu zdefiniowanemu przez wzór (2) — odległość

$$(4) \quad r = 1 - \frac{w}{a+b-w} = \frac{a+b-2w}{a+b-w}.$$

We wzorze (4) licznikiem jest liczba gatunków występujących w jednym i tylko jednym z dwóch badanych zespołów, a mianownikiem — liczba wszystkich gatunków występujących w tych zespołach. Oznaczając przez A zbiór wszystkich gatunków występujących w \mathfrak{A} , a przez B zbiór wszystkich gatunków występujących w \mathfrak{B} i używając terminologii algebry zbiorów, możemy powiedzieć, że mianownikiem jest liczba elementów sumy zbiorów A i B , a licznikiem liczba elementów różnicy symetrycznej A i B . Oznaczając odległość A i B przez $\varrho(A, B)$, liczbę elementów zbioru X przez $m(X)$ i używając symboliki algebry zbiorów, możemy więc wzór (4) napisać w innej formie:

$$(5) \quad \varrho(A, B) = \frac{m(A \dot{-} B)}{m(A + B)}.$$

Definicja (4), czyli (5), da się przedstawić graficznie jak następuje. Przedstawmy rozważane zespoły roślinne — ściślej mówiąc dwa zbiory gatunków roślin A i B — jako obszary na płaszczyźnie tak, by każdy gatunek odpowiadał jednostce pola. Tak więc każdy z narysowanych obszarów A i B ma tyle jednostek pola, ile reprezentuje gatunków; również część wspólna ma tyle jednostek pola, ile oba zespoły mają wspólnych gatunków. Przy tej interpretacji odległość (4) jest równa stosunkowi pola obszaru zakreskowanego pojedynczo na rysunku 1 do pola całego zakreskowanego obszaru.

Odległość ϱ powinna spełniać następujące postulaty, jeśli ma odpowiadać intuicji (i być w zgodzie ze zwykłą terminologią matematyczną):

1. $\varrho(A, B) \geq 0$, przy czym $\varrho(A, B) = 0$ wtedy i tylko wtedy, gdy $A = B$,

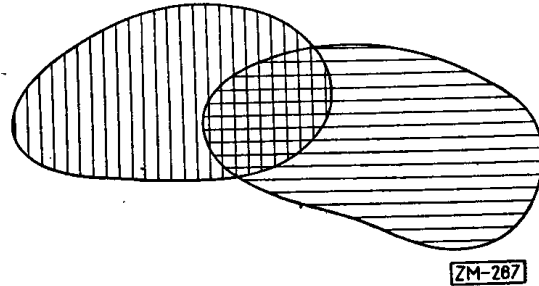
2. $\varrho(A, B) = \varrho(B, A)$ (tzw. warunek symetrii),

3. $\varrho(A, B) + \varrho(B, C) \geq \varrho(A, C)$ (tzw. warunek trójkąta).

Widać od razu, że odległość (5) spełnia warunek 1 (którego sens jest widoczny: odległość systematyczna dwóch zespołów roślinnych jest nieujemna, a zerem jest wtedy i tylko wtedy, gdy nie różnią się one pod względem gatunków), a także warunek 2.

Spełniony jest też warunek

1'. $\varrho(A, B) \leq 1$, przy czym $\varrho(A, B) = 1$ wtedy i tylko wtedy, gdy A i B nie mają elementów wspólnych.



Rys. 1

ZM-267

Inaczej mówiąc: odległość jest maksymalna (100%, czyli 1) dla zespołów bez gatunków wspólnych i tylko dla takich. Również odległość zdefiniowana wzorem (3) spełnia warunki 1, 1' i 2. Istotna różnica między dwiema definicjami ujawnia się przy badaniu warunku trójkąta. Odległość (3) nie spełnia go, co wynika z podanego poprzednio przykładu, w którym odległość od A do B równa się $\frac{1}{4}$, od B do C też $\frac{1}{4}$, podczas gdy od A do C aż 1.

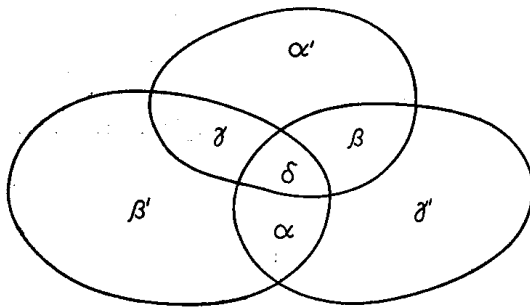
Otóż odległość (5) spełnia warunek trójkąta, co udowodnimy za chwilę i co jest główną racją przemawiającą za tą definicją.

Posługując się nią. S. Kuleczyński obliczył dendryt 128 obszarów roślinnych, na które można podzielić ziemię — ten dendryt potwierdza intuicyjny pogląd na ich pokrewieństwa pod względem ekologii roślin.

3. Warunek trójkąta. Udowodnimy, że

$$(6) \quad e(A, B) + e(B, C) \geq e(C, A).$$

Trzy dowolne zbiory A, B, C rozbijają uniwersum na 8 części co



ZM-268

Rys. 2

do przynależności dowolnego elementu: każdy element uniwersum może należeć lub nie należeć do A , przy czym te relacje wyłączają się wzajemnie; podobnie jest z przynależnością do B i do C — jest zatem 2^3 możliwości. Graficznie można je przedstawić jako 8 obszarów (7 trójkątów krzywoliniowych i obszar zewnętrzny, rys. 2).

Oznaczenia pól na rysunku małymi literami greckimi pozwa-

lają tezę sprowadzić do twierdzenia arytmetycznego o liczbach nieujemnych:

$$\frac{\alpha + \alpha' + \beta + \beta'}{\alpha + \alpha' + \beta + \beta' + \gamma + \delta} + \frac{\beta + \beta' + \gamma + \gamma'}{\beta + \beta' + \gamma + \gamma' + \alpha + \delta} \geq \frac{\gamma + \gamma' + \alpha + \alpha'}{\gamma + \gamma' + \alpha + \alpha' + \beta + \delta}.$$

Napiszmy

$$(7) \quad \begin{aligned} \alpha' + \alpha &= a, & \beta' + \beta &= b, & \gamma' + \gamma &= c, \\ \alpha + \delta &= u, & \beta + \delta &= v, & \gamma + \delta &= w, \end{aligned}$$

a twierdzenie to przyjmie postać

$$\frac{a + b}{a + b + w} + \frac{b + c}{b + c + u} \geq \frac{c + a}{c + a + v}$$

(oczywiście użyte tu oznaczenia a i b nie mają nic wspólnego z oznaczeniami w § 1 i § 2), to znaczy

$$\begin{aligned} (a+b)(b+c+u)(c+a+v) + (b+c)(c+a+v)(a+b+w) &\geq \\ &\geq (c+a)(a+b+w)(b+c+u), \end{aligned}$$

a po uproszczeniu

$$\begin{aligned} (a+b)(b+c)(c+a) + (a+b)(b+c)v + (a+b)uv + \\ + (b+c)(a+b)v + (b+c)vw \geq (c+a)wu, \end{aligned}$$

czyli

$$(a+b)(b+c)(c+a) + 2(a+b)(b+c)v + (a+b)uv + (b+c)vw \geq (c+a)wu.$$

Po uwzględnieniu (7) ostatnia nierówność przyjmie postać

$$\begin{aligned} (8) \quad (a+b)(b+c)(c+a) + 2(a+b)(b+c)(\beta+\delta) + \\ + (a+b)(a+\delta)(\beta+\delta) + (b+c)(\beta+\delta)(\gamma+\delta) \geq (c+a)(\gamma+\delta)(a+\delta). \end{aligned}$$

Nierówność tę można — po rozwinięciu względem potęg δ — napisać w postaci

$$P + Q\delta + R\delta^2 \geq p + q\delta + r\delta^2.$$

Otóż

$$\begin{aligned} P &\geq (a+b)(b+c)(c+a) \geq ac(c+a) \geq a\gamma(c+a) = p, \\ Q &= 2(a+b)(b+c) + (a+b)(a+\beta) + (b+c)(\beta+\gamma) \geq 2ac + aa + c\gamma \geq \\ &\geq ca + a\gamma + c\gamma + aa = q, \\ R &= (a+b) + (b+c) \geq a+c = r. \end{aligned}$$

Z tych porównań wynika — wobec $\delta \geq 0$ — nierówność (8), a więc i teza (6).

Z porównania wyrazów wolnych wynika, że nierówność jest ostra w przypadku ogólnym, to jest tym, który przedstawia rysunek: wtedy bowiem litery greckie i łacińskie wszystkie przedstawiają liczby dodatnie. Znalezienie warunku koniecznego i dostatecznego na ostrość (6) byłoby interesujące, bo gdy (6) zamienia się na równość, możemy mówić o kolineacji zbiorów A, B, C i stwierdzić, że jeden z nich leży między pozostałymi. W zastosowaniach biologicznych może to mieć interpretację (np. może oznaczać drogę ewolucyjną zespołów).

4. Odległość zbiorów. Oczywiście koncepcję odległości, zdefiniowanej wzorem (5), można stosować do dowolnych zbiorów skończonych:

$$e(A, B) = \frac{m(A \dot{-} B)}{m(A + B)},$$

gdzie m oznacza moc (czyli liczbę elementów) zbioru.

Oznaczając przez 1 przynależność danego elementu do zbioru, a przez 0 nieprzynależność, możemy ułożyć tablicę 1, gdy np. zbiór A składa się z elementów a, b, c i d (w lesie \mathfrak{A} rośnie sosna, dąb, brzoza i olcha), a zbiór B z elementów b, c, d i e (w lesie \mathfrak{B} rośnie dąb, brzoza, olcha i świerk).

TABLICA 1

	a	b	c	d	e	m
A	1	1	1	1	0	
B	0	1	1	1	1	
$A \dot{-} B$	1	0	0	0	1	2
$A + B$	1	1	1	1	1	5

Tak więc $m(A \dot{-} B) = 2$, $m(A + B) = 5$, a zatem $\varrho(A, B) = \frac{2}{5}$.

Koncepcję tę można dalej uogólnić, rozpatrując zbiory A i B nieskończone, a nawet nieprzeliczalne, a za m przyjmując pole, objętość lub inną miarę (ogólnie: miarę abstrakcyjną)⁽¹⁾.

5. Odległość funkcji i jej zastosowania do zespołów roślinnych.

Dalszym uogólnieniem, interesującym praktycznie i teoretycznie, niech będzie odległość funkcji, analogiczna do opisanej odległości zbiorów.

Niech więc f i g będą dwiema funkcjami o tej samej przestrzeni zmiennej niezależnej, w której jest określona pewna miara m . Mogą to być na przykład dwie funkcje określone w odcinku $\langle 0, 1 \rangle$, w którym jako miarę m rozważa się zwykłą długość, mogą być funkcje określone w zbiorze złożonym z pięciu elementów: a, b, c, d i e , w którym jako miarę m rozważa się liczbę elementów itd. Względem miary m można całkować; odległość funkcji f i g nieujemnych definiujemy wtedy wzorem

$$(9) \quad \varrho(f, g) = \frac{\int |f - g| dm}{\int \max(f, g) dm},$$

i ogólniej — już bez założenia o znaku funkcji f i g — wzorem

$$(10) \quad \varrho(f, g) = \frac{\int |f - g| dm}{\int \max(|f|, |g|, |f - g|) dm}.$$

Gdy przestrzeń ma skończoną liczbę elementów, całkowanie przechodzi w sumowanie. Niech np. w przestrzeni złożonej z pięciu elementów a, b, c, d i e będą określone funkcje f i g jak następuje

$$\begin{aligned} f(a) &= 4, & f(b) &= 2, & f(c) &= 3, & f(d) &= 1, & f(e) &= 0, \\ g(a) &= 0, & g(b) &= 2, & g(c) &= 1, & g(d) &= 2, & g(e) &= 5, \end{aligned}$$

⁽¹⁾ Trzeba tu przypomnieć, że w wielu pracach rozważana była tzw. odległość Frécheta, Nikodyma i Aronszajna zbiorów, określona jako licznik wyrażenia (5), tj. miara różnicy symetrycznej zbiorów. Por. np. P. R. Halmos, *Measure theory*, New York 1950, str. 168 i 169.

co można, z myślą o zastosowaniu wzoru (9), ująć w tabelę 2. Odczytujemy na niej

$$e(f, g) = \frac{12}{16} = \frac{3}{4}.$$

TABLICA 2

	a	b	c	d	e	f
<i>f</i>	4	2	3	1	0	
<i>g</i>	0	2	1	2	5	
$ f-g $	4	0	2	1	5	12
$\max(f, g)$	4	2	3	2	5	16

I to pojęcie ma interpretację dla gatunków roślinnych. Funkcje *f* i *g* charakteryzować będą te same dwa lasy \mathfrak{A} i \mathfrak{B} , co w tabelicy 1, ale występowanie poszczególnych gatunków jest teraz scharakteryzowane także ilościowo: w pierwszym lesie przypadają na każdym 10 drzew: 4 sosny,

2 dęby, 3 brzozy i 1 olcha, a w drugim 2 dęby, 1 brzoza, 2 olchy i 5 świerków. Są to dane⁽²⁾ z Wrocławskiego Rejonu Lasów Państwowych (nadleśnictwo Miękinia: oddział 41, pododdział d i oddział 98, pododdział g).

Oczywiście odległość (5) jest szczególnym przypadkiem odległości (9): wystarczy za funkcje *f* i *g* wziąć funkcje charakterystyczne zbiorów *A* i *B* (tj. funkcje równe 1 w danym zbiorze, a 0 poza nim).

Można też i na odwrót: sprowadzić w pewien sposób do odległości (5) odległość (9) (a także (10)), a dzięki temu sprawdzić dla niej warunek trójkąta, skoro dla (5) został już wcześniej sprawdzony⁽³⁾.

6. Zasada wzajemności. Definicja podobieństwa (2) oraz odległości (4) ma jeszcze inną interpretację. Można mianowicie postawić sobie w botanice (a także i w innych badaniach, w których występują biotopy, a nawet w badaniach lingwistycznych lub jeszcze odleglejszych od biologii) problemat poniekąd odwrotny. Na przykładzie lasów można go sformułować tak: Pytamy nie o podobieństwo wzajemne dwóch obszarów leśnych (np. obszarów \mathfrak{A} i \mathfrak{B} zacytowanych przed chwilą) na tle gatunków drzew, ale o podobieństwo (cenotopiczne, czyli systematyczne) dwóch gatunków drzew, na przykład brzozy i olchy, na tle obszarów, w których występują. W tym celu posługujemy się tymi samymi informacjami, które nam służyły w pierwotnym zadaniu, np. danymi z Wrocławskiego Rejonu Lasów Państwowych, które teraz będziemy czytali odwrotnie: stwierdzimy, ile w tym rejonie jest lasów, w których występuje brzoza (*a*), ile takich, w których występuje olcha (*b*) i — w końcu — ile takich, w których występują oba gatunki (*w*). Odczytawszy liczby *a*, *b*, *w* z rejestrów rejonu obliczymy z wzoru (4) odległość systematyczną *s* olchy od brzozy na tle lasów tego rejonu. To pojęcie można również

⁽²⁾ Zawdzięczamy je mgrowi F. Szczotce.

⁽³⁾ Te i inne rozważania teoretyczne podane są w pracy E. Marczewskiego i H. Steinhausa *On a certain distance of sets and the corresponding distance of functions*, Coll. Math. 6 (1958), str. 319-327.

спrecyzować przez uwzględnienie także częstości występowania (np. w lesie \mathcal{A} , jak czytaliśmy, brzoza występuje 0,3 razy, a olcha 0,1 razy).

Gdy obliczymy wzajemne odległości wszelkich gatunków drzew rejonu, będziemy mogli stwierdzić, które z tych odległości są duże w porównaniu z innymi. Jeżeli stanie przed nami zadanie obliczenia odległości obszarów roślinnych, które zawierają wiele gatunków, to przy pierwszym oszacowaniu wystarczy uwzględnić we wszystkich obszarach tylko te gatunki, których wzajemne odległości uprzednio obliczone, okazały się duże, gdyż te gatunki charakteryzują owe obszary. Do wybrania tych gatunków służy metoda dendrytów (tzw. taksonomia wrocławska)⁽⁴⁾.

Koncepcję przedstawioną w tym paragrafie można nazwać *zasadą wzajemności* lub *zasadą dualizmu*⁽⁵⁾.

Praca wpłynęła 9. 4. 1958

Э. МАРЧЕВСКИЙ и Х. ШТАЙНХАУЗ (Вроцлав)

ОБ СИСТЕМАТИЧЕСКОМ РАССТОЯНИИ РАСТИТЕЛЬНЫХ МАССИВОВ

РЕЗЮМЕ

Если в растительном массиве (например в лесу) \mathcal{A} выступает a видов, в массиве \mathcal{B} — b видов, а w есть число общих видов, т. е. видов, выступающих так в массиве \mathcal{A} как и в массиве \mathcal{B} , то расстоянием этих массивов называем число (4).

Число это характеризует разницу между массивами \mathcal{A} и \mathcal{B} . В частности, если в \mathcal{A} и \mathcal{B} выступает точно те же виды ($a = b = w$), то расстояние массивов равно нулю, а если \mathcal{A} и \mathcal{B} не имеют никакого общего вида ($w = 0$), то расстояние принимает наибольшее значение 1.

Вообще, если μ есть произвольная фиксированная мера, т. е. неотрицательная, аддитивная функция множества, расстояние между двумя μ -измеримыми множествами определяется равенством (5), где $A+B$ обозначает соединение множеств, а $A \div B$ — симметрическую разность.

Ещё более обще, расстояние двух неотрицательных μ -измеримых функций определяется формулой (9).

Определенное таким образом расстояние удовлетворяет условию треугольника и разным условиям регулярности.

(4) Por. zbiorową pracę Ogólnej Grupy Zastosowań Państwowego Instytutu Matematycznego (K. Florek, J. Łukaszewicz, J. Perkal, H. Steinhaus, S. Zubrzycki), *Taksonomia wrocławska*, Przegląd Antropologiczny 17 (1951), str. 193-211.

(5) Koncepcja ta przedstawiona była po raz pierwszy w cytowanej pracy zbiorowej, str. 205-207.

Расстояние (9) может служить для определения расстояния растительных массивов не только относительно числа видов, но и относительно численности различных видов, выступающих в этих массивах.

E. MARCZEWSKI and H. STEINHAUS (Wrocław)

ON THE SYSTEMATIC DISTANCE OF BIOTOPES

SUMMARY

If a species appear in a biotope \mathfrak{A} , e. g. in a wood, b species appear in biotope \mathfrak{B} and w is the number of common species, i. e. those occurring both in \mathfrak{A} and in \mathfrak{B} , the distance of those biotopes is the number (4).

This number shows how the biotopes \mathfrak{A} and \mathfrak{B} differ from each other. In particular, if the same species ($a = b = w$) are in \mathfrak{A} and \mathfrak{B} , then their distance is zero and if \mathfrak{A} and \mathfrak{B} have no species in common, then it assumes its greatest value, namely 1.

Generally, when μ is an arbitrary fixed measure, i. e. a non-negative additive set function, then the distance of two sets measurable by measure μ is defined by equality (5) where $A + B$ denotes the sum and $A \dot{-} B$ the symmetric difference of the sets A and B .

Still more generally, the distance of two non-negative functions measurable by the measure μ is defined by formula (9).

The distances thus defined satisfy the condition of the triangle and various conditions of regularity.

Distance (9) may serve to define the dissimilarity of biotopes with respect not only to the number of species but also to the number of individuals of each species occurring in the given biotopes.
