Ein Modell phylogenetischer Diversität

Klaus Nehring¹ und Clemens Puppe²

Zusammenfasung:

Ziel des Beitrags ist es, einen kurzen Überblick über das in Nehring/Puppe 2003a vorgeschlagene Modell phylogenetischer Diversität zu geben. Das Modell basiert auf der "Theory of Diversity" (Nehring/Puppe 2002), deren Grundgedanken ebenfalls kurz vorgestellt werden. Im Gegensatz zu Weitzman's (1992) rein kladistischer Herangehensweise erlaubt das hier vorgestellte Modell, neben auch evolutionsgeschichtlicher taxonomische Information berücksichtigen. Trotz der dadurch gewonnenen größeren Flexibilität und Anwendbarkeit lässt sich die Diversität einer Menge von Arten immer noch auf einfache Weise rekursiv aus den paarweisen Distanzen ihrer Elemente ableiten.

1 Einleitung

Ziel dieses Beitrags ist es, einen kurzen Überblick über das in Nehring/Puppe (2003a) vorgeschlagene Modell phylogenetischer

¹ Department of Economics, University of California Davis, CA 95616, U.S.A., kdnehring@ucdavis.edu

² Wirtschaftstheoretische Abteilung III, Universität Bonn Adenauerallee 24-42, D- 53113 Bonn, clemens.puppe@wiwi.uni-bonn.de

Diversität zu geben. Die grundlegenden Begriffe und Resultate der "Theory of Diversity" (Nehring/Puppe 2002), auf denen das Modell aufbaut, werden ebenfalls kurz vorgestellt (siehe hierzu auch Nehring/Puppe 2001). Im Gegensatz zu Weitzman's (1992) rein kladistischer Herangehensweise erlaubt das hier betrachtete Modell, neben evolutionsgeschichtlicher Information auch taxonomische Information zu berücksichtigen. Das kladistische Modell ergibt sich dabei als ein Spezialfall. Trotz des höheren Grades an Flexibilität und Allgemeinheit lässt sich die Diversität einer Menge von Arten immer noch rekursiv aus den paarweisen Distanzen ihrer Elemente ableiten (es handelt sich also um ein *azyklisches* Modell im Sinne von Nehring/Puppe 2002).

2 Das Multi-Attribut-Modell

Es bezeichne X eine endliche Grundgesamtheit von Objekten. In dem hier betrachteten biologischen Kontext können diese Objekte einzelne Individuen, Arten ("Spezies") oder aber auch in sich abgeschlossene Ökosysteme ("Habitats") sein. Im folgenden werden wir uns auf die Spezies-Interpretation beschränken. Das Ziel der Analyse ist, jeder Teilmenge $S \subseteq X$ von Arten einen Diversitätswert v(S) zuzuordnen. Ein naiver erster Vorschlag könnte sein, die Diversität einer Menge S einfach mit der Anzahl der Arten in S zu identifizieren. Das Problem dieses Vorschlags ist, dass dabei Ähnlichkeiten zwischen Arten unberücksichtigt bleiben. Sind beispielsweise x und y einander sehr ähnlich und beide sehr verschieden von z, so sollte intuitiv die Diversität der Menge $\{x, y\}$ kleiner sein als die Diversität der Menge $\{x, z\}$ ungeachtet der Tatsache, dass beide Mengen zwei Elemente enthalten.

Die grundlegende Idee des in Nehring/Puppe (2002) entwickelten Multi-Attribut-Modells ist, dass die Diversität einer Menge von Objekten durch die Gesamtheit der relevanten Eigenschaften der enthaltenen Objekte bestimmt ist. Es sei $\{f, f', ...\}$ eine Liste von Eigenschaften, die die Arten in X besitzen können. Beispiele relevanter Eigenschaften sind in unserem Kontext: ein bestimmtes Gen zu besitzen, eine natürliche Arznei zu enthalten, die Zugehörigkeit zu einer bestimmten biologischen Gattung

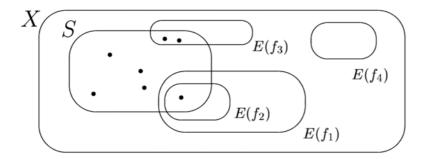
oder die Abstammung von einem bestimmten Vorfahren. Für jede Eigenschaft bezeichne $\lambda_f \ge 0$ ihr *(relatives) Gewicht.* Weiterhin bezeichne E(f) die Extension von f, d.h. die Menge aller Arten in X, die die Eigenschaft f besitzen. Ist beispielsweise f die Eigenschaft "ein Säugetier zu sein", so ist E(f) einfach die Teilmenge aller Säugetiere in X. Eine Art x besitzt also die Eigenschaft f genau dann, wenn $x \in E(f)$. Wir sagen, dass eine Menge S eine Eigenschaft f realisiert, wenn es mindestens eine Art in S gibt, die die Eigenschaft f besitzt, d.h. wenn der Durchschnitt der Mengen S und E(f) nicht leer ist.

Definition Eine Funktion ν , die jeder Teilmenge $S \subseteq X$ einen Wert $\nu(S)$ zuordnet, heißt eine *Diversitätsfunktion*, falls es eine Liste von Eigenschaften und Gewichte $\lambda_f \ge 0$ gibt derart, dass für alle S gilt:

$$v(S) = \sum_{f: E(f) \cap S \neq \emptyset} \lambda_f$$
 (1)

Gemäß Gleichung (1) ist also der Diversitätswert einer Menge S von Arten durch die Summe der Gewichte aller Eigenschaften gegeben, die von S realisiert werden, d.h. die von mindestens einer Art in S verkörpert werden. Insbesondere unterscheiden wir also nur, ob eine Eigenschaft von wenigstens einer Art erfüllt wird oder nicht, unabhängig davon, auf wie viele Arten sonst noch diese Eigenschaft zutreffen mag. Die folgende Abbildung illustriert diese Definition. In Abbildung 1 werden die Eigenschaften f_1 , f_2 , f_3 von der Menge S realisiert, die Eigenschaft f_4 aber nicht. Daher geht auch das Gewicht λ_{f4} nicht in den Diversitätswert von S ein. Man beachte außerdem, dass die Eigenschaft f2 spezieller als die Eigenschaft f_I ist (weil $E(f_2) \subset E(f_I)$). Zum Beispiel könnte f_I für die Eigenschaft "Wirbeltier" stehen und f_2 für die Eigenschaft "Säugetier". Da jedes Säugetier gleichzeitig ein Wirbeltier ist, ist die Existenz eines in S hinreichend dafür, dass beide Eigenschaftsgewichte λ_{f2} und λ_{f1} in den Diversitätswert von S eingehen.

Abbildung 1: Die Menge S realisiert die Eigenschaften f₁, f₂ und f₃



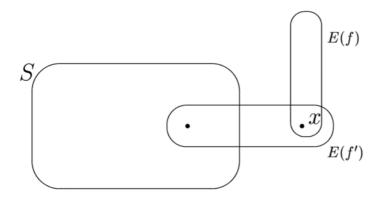
Was folgt aus diesem Vorschlag für die *marginale* Diversität eines zusätzlichen Objektes?

Offensichtlich fügt ein Objekt x dem Diversitätswert von S die Gewichte aller derjenigen Eigenschaften hinzu, die x besitzt, die aber von keinem anderen Objekt in S bereits verkörpert werden. Mit anderen Worten, es gilt:

$$v(S \cup \{x\}) - v(S) = \sum_{f: x \in E(f), E(f) \cap S = \emptyset} \lambda_f$$
 (2)

Die folgende Abbildung verdeutlicht die Gleichung (2). Die Art x in Abbildung 2 besitzt sowohl die Eigenschaft f als auch die Eigenschaft f' (weil $x \in E(f)$ und $x \in E(f')$). Da aber f' bereits von einer anderen Art in S verkörpert wird, trägt nur die Eigenschaft f mit dem Gewicht λ_f zum marginalen Wert $v(S \cup \{x\}) - v(S)$ bei.

Abbildung 2: Marginale Diversität von x bei Hinzufügung zu S



Ein wesentlicher Aspekt der vorgeschlagenen Form der Diversitätsbestimmung ist die Nicht-Negativität der Eigenschaftsgewichte λ_f . Diese Einschränkung garantiert, dass die Funktion v tatsächlich als eine Bestimmung der Diversität – im Gegensatz zum subjektiven Gesamtnutzen einer Menge - aufgefasst werden kann: Die Realisierung zusätzlicher Eigenschaften kann den Wert einer Menge im Hinblick auf ihre Diversität niemals verringern. Es ist dabei durchaus möglich, dass gewisse Eigenschaften zwar unerwünscht sind gleichzeitig aber zur Vielfalt beitragen können. Beispielsweise kann eine Person jede Begegnung mit einer Spinne als unangenehm empfinden, gleichzeitig aber einräumen, dass die Existenz der Gattung der Spinnen zur Vielfalt des Lebens beiträgt.

Zwei fundamentale Eigenschaften einer Diversitätsfunktion lassen sich aus der Nicht-Negativität der Eigenschaftsgewichte ableiten. Aus (2) folgt unmittelbar, dass der marginale Wert einer zusätzlichen Art nie negativ ist ("Monotonie"). Außerdem folgt, dass der marginale Wert einer Art umso kleiner ist je umfassender die Menge ist, zu der die Art hinzugefügt wird, d.h. der marginale Wert von Diversität ist fallend ("Submodularität"):

Monotonie $v(S \cup \{x\}) - v(S) \ge 0$

Submodularität
$$S \subseteq T \Rightarrow [v(S \cup \{x\}) - v(S) \ge v(T \cup \{x\}) - v(T)]$$

Die Eigenschaft der Submodularität lässt sich im Kontext von Diversität wie folgt interpretieren: je vielfältiger eine Menge von Arten bereits ist, desto weniger *zusätzliche* Diversität wird durch die Existenz weiterer Arten erzeugt.

3 Attribute und die extensionale Eindeutigkeit der Eigenschaftsgewichte

Die Essenz des Multi-Attribut-Modells ist es, die Bestimmung von Artendiversität auf eine Bewertung der Eigenschaften der entsprechenden Arten zurückzuführen bzw. aus dieser abzuleiten: Jede gegebene Bewertung von Eigenschaften induziert gemäß der Formel (1) einen entsprechenden Diversitätswert. Man kann nun umgekehrt zeigen, dass sich jede gegebene Funktion v, die jeder Teilmenge $S \subseteq X$ einen Wert zuordnet, in der Form (1) mit geeignet gewählten Eigenschaftsgewichten darstellen lässt. Mit anderen Worten, jede Bewertung von Teilmengen lässt sich als eine Aggregation impliziter Eigenschaftsgewichte auffassen. Dabei sind die entsprechenden Eigenschaftsgewichte extensional eindeutig bestimmt. Dies wollen wir nun genauer erläutern. Es sei $A \subseteq X$ eine beliebige Teilmenge von Arten. Es ist möglich, dass es mehrere Eigenschaften gibt, die genau von den Arten in A verkörpert werden, d.h. deren Extension die Menge A ist (z.B. könnten in einem geographisch begrenzten Habitat die karnivoren Arten gerade diejenigen Arten sein, die zu einer bestimmten taxonomischen Klassifikation gehören). Man kann nun jede Teilmenge A mit der Gesamtheit aller Eigenschaften, deren Extension genau A ist, identifizieren. Jede Teilmenge A kann auf diese Weise als ein mögliches Attribut aufgefasst werden. Das Gewicht eines Attributs $A \subseteq X$ bestimmt sich gemäß der Formel

$$\lambda_A := \sum_{f: E(f)=A} \lambda_f$$

Das Gewicht λ_A des Attributs A ist also die Summe der Gewichte aller Eigenschaften mit Extension A. Falls es keine Eigenschaft gibt, deren Extension gleich A ist, so setzen wir $\lambda_A = 0$.

Es gilt nun folgendes Resultat (vgl. Nehring/Puppe 2002, Fact 2.1): Zu jeder Funktion ν gibt es eine Gewichtung λ aller Teilmengen von X derart, dass für alle S gilt:

$$v(S) = \sum_{A: A \subset X, A \cap S \neq \emptyset} \lambda_A$$

Dabei sind die Attributgewichte λ_A eindeutig durch v bestimmt.³ Die Attributgewichte λ_A sind alle nicht-negativ genau dann, wenn die Funktion v monoton und submodular ist und eine gewisse Regularitätsbedingung erfüllt (vgl. Nehring/Puppe 2002, Fact 2.2). Mit anderen Worten: eine Diversitätsfunktion v ist im wesentlichen durch die Bedingungen der Monotonie und Submodularität charakterisiert.

Wegen der Eindeutigkeit der Attributgewichte "offenbart" also eine Diversitätsfunktion v, welche Attribute als relevant angesehen werden, nämlich genau diejenigen Attribute, deren Gewicht strikt positiv ist. Für eine Diversitätsfunktion v bezeichnen wir mit $\Lambda := \{A \subseteq X : \lambda_A > 0\}$ die Menge der Attribute.

4 Diversität als Funktion paarweiser Distanzen: Azyklizität

Weitzman (1992) hat folgenden Vorschlag zur Bewertung von Diversität gemacht. Gegeben sei eine Metrik d, die für je zwei Arten x und y in X deren Distanz ("dissimilarity") d(x, y) angibt. Weitzman schlägt vor, die marginale Diversität von x bei Hinzufügung zur Menge S mit der minimalen Distanz von x zu den Elementen von S zu identifizieren, d.h.

$$v(S \cup \{x\}) - v(S) = \min_{y \in S} d(x, y)$$
 (3)

Diese Gleichung erlaubt es, die Diversität einer Menge von Arten *rekursiv* zu bestimmen: Ist die Menge $S = \{x_1, ..., x_n\}$ gegeben, bestimme zunächst gemäß Formel (3) die marginale Diversität von x_1 bei Hinzufügung zur Menge $\{x_2, x_3, ..., x_n\}$; dann bestimme die marginale Diversität von x_2 bei Hinzufügung zu $\{x_3, ..., x_n\}$ usw. Auf diese Weise lässt sich die Diversität von $\{x_1, ..., x_n\}$ auf die paarweisen Distanzen und den Wert aller einelementigen Mengen zurückführen. Das Problem dieser rekursiven Methode ist, dass sie im allgemeinen *pfadabhängig* ist, d.h.

³ Es wird hier die Normalisierung $v(\emptyset) = \lambda_\emptyset = 0$ vorausgesetzt. Die Attributgewichte lassen sich aus v gemäß der folgenden Formel berechnen: $\lambda_A = \sum_{S \subseteq A} (-1)^{\#A - \#S + I} v(X \setminus S)$. Implizit hängen die Attributgewichte also auch davon ab, welche Menge X von Arten der Bewertung zu Grunde liegt.

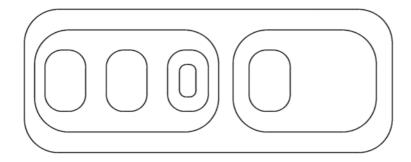
für verschiedene Reihenfolgen der Elemente ergeben sich im allgemeinen verschiedene Diversitätswerte. Anders ausgedrückt: die Rekursionsformel (3) kann nur für sehr spezielle Distanzmetriken erfüllt sein.

Im Rahmen des Multi-Attribut-Modells lassen sich paarweise Distanzen zwischen Arten auf natürliche Weise wie folgt definieren. Für je zwei Arten x und y in X sei

$$d(x, y) := v(\{x, y\}) - v(\{y\}) = \sum_{f: x \in E(f), y \notin E(f)} \lambda_f$$

Mit anderen Worten, die Distanz von x zu y wird definiert als das Gewicht aller Eigenschaften (oder Attribute) die x besitzt, die aber y nicht besitzt. Es gilt nun das folgende Resultat (Nehring/Puppe 2002, Theorem 3.1). Eine Diversitätsfunktion v erfüllt die Rekursionsformel (3) genau dann, wenn die zugehörige Menge Λ der relevanten Attribute "hierarchisch" geordnet ist. Dabei heißt eine Menge $\{A, A', \dots\}$ von Attributen hierarchisch geordnet, falls je zwei Attribute entweder einen leeren Durchschnitt haben oder eines der beiden Attribute ganz im anderen Attribut enthalten ist. Die hierarchische Anordnung schließt also aus, dass sich zwei Attribute "überlappen" können. Die folgende Abbildung illustriert eine hierarchisch geordnete Menge von Attributen.

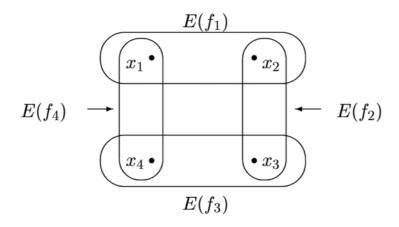
Abbildung 3: Eine hierarchische Anordnung von Attributen



⁴ Eine ausführliche Analyse des hierarchischen Modells und des allgemeineren "line model" findet sich in Nehring/Puppe (2003b).

Die prinzipielle Idee, die Bestimmung der Diversität einer Menge von Arten auf eine Bestimmung der paarweiser Distanzen ihrer Elemente (und die Bewertung aller einelementigen Mengen) zurückzuführen, lässt sich verallgemeinern. Eines der Hauptresultate der "Theory of Diversity" (Nehring/Puppe 2002, Theorem 6.2) zeigt, dass dies genau dann möglich ist, wenn die relevanten Eigenschaften *azyklisch* sind, d.h. wenn die zugehörigen Attribute keine Zyklen aufweisen. Die folgende Abbildung zeigt beispielhaft einen Zyklus von vier Attributen:

Abbildung 4: Ein Zyklus von vier Attributen

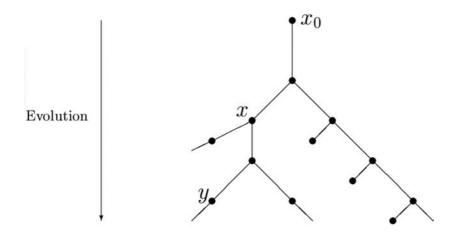


Der "Zyklus" in Abbildung 4 besteht darin, dass es zu den vier Arten x_1 , x_2 , x_3 und x_4 vier Eigenschaften f_1 , f_2 , f_3 und f_4 gibt derart, dass von der Menge $\{x_1, x_2, x_3, x_4\}$ genau die Arten x_1 und x_2 die Eigenschaft f_1 besitzen, genau die Arten x_2 und x_3 die Eigenschaft f_2 , genau die Arten x_3 und x_4 die Eigenschaft f_3 , sowie (und hier schließt sich der Zyklus) genau die Arten x_4 und x_1 die Eigenschaft f_4 . Analog dazu liegt ein Zyklus von drei Attributen dann vor, wenn es drei Arten gibt derart, dass jedes Paar dieser Arten eine gemeinsame Eigenschaft besitzt, die die dritte Art nicht besitzt. Die Azyklizität schließt alle Zyklen von k Attributen für jede ganze Zahl $k \ge 3$ aus.

5 Phylogenetische Diversität

Im Gegensatz zu genetischer oder funktionaler Diversität basiert *phylogenetische* Diversität wesentlich auf evolutionsgeschichtlicher Information. Diese Information wird mit Hilfe einer partiellen Ordnung \geq_{ev} beschrieben. Konkret bedeute im folgenden $y \geq_{ev} x$ "die Art y stammt von der Art x ab" (siehe Abbildung 5). Wie allgemein üblich nehmen wir an, dass \geq_{ev} die Struktur eines Baumes erzeugt, d.h. für jede Art x ist die Menge aller Vorfahren von x total geordnet; insbesondere hat also jede Art (außer der "Ur-Art" x_0) genau einen unmittelbaren Vorfahren.

Abbildung 5: Die Baumstruktur der Ordnung ≥_{ev}



Für jede Art x bezeichne C_x die Menge aller Nachfahren von x; die Menge dieser Arten wird auch als die zu x gehörige Klade bezeichnet. Weitzman's (1992) Vorschlag ist durch die Annahme charakterisiert, dass die relevanten Attribute genau den Kladen entsprechen. Das heißt also, dass alle relevanten Eigenschaften von der Form "stammt von der Art x ab" sind ("kladistisches Modell"). Es ist unmittelbar klar, dass die Menge aller Kladen eine hierarchisch geordnete Mengenfamilie bildet, insbesondere kann die Diversität einer Menge von Arten im kladistischen Modell rekursiv gemäß der Gleichung (3) bestimmt werden.

Das kladistische Modell hat eine Reihe von nicht wünschenswerten Restriktionen. Die damit verbundenen Probleme bei der Anwendung des kladistischen Modells lassen sich im wesentlichen darauf zurückführen, dass die zugrundeliegenden Attribute "unsterblich" sind: die Eigenschaft "stammt von Art x ab" vererbt sich automatisch auf jeden Nachfahren von x. Mit anderen Worten schließt das kladistische Modell also alle Eigenschaften aus, die zwar manche aber eben nicht alle Arten einer Klade besitzen. Solche Eigenschaften können aber durchaus von Bedeutung sein, wie das folgende Beispiel zeigt (eine ausführlichere Kritik des kladistischen Modells und eine Reihe weiterer Beispiele findet sich in Nehring/Puppe 2003a; vgl. dazu auch Faith 1992):

Abbildung 6: Ein Ausschnitt des evolutionären Baumes

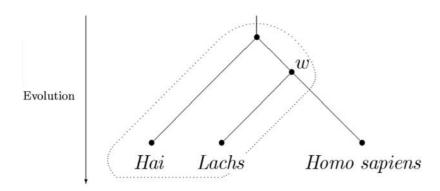


Abbildung 6 zeigt einen Ausschnitt der Stammesgeschichte der Arten. Die für unsere Zwecke wichtige Beobachtung ist, dass die Arten "Mensch" und "Lachs" einen gemeinsamen Vorfahren haben (w in der Abbildung), der kein Vorfahr der Art der Haie ist.⁵ Im kladistischen Modell hat dies unmittelbar zur Folge, dass (i) "Mensch" und "Lachs" eine gemeinsame Eigenschaft besitzen, die "Hai" nicht besitzt (nämlich von w abzustammen), während (ii) jede gemeinsame Eigenschaft von "Hai" und "Lachs" auch eine Eigenschaft der Art "Mensch" ist. Haie und Lachse haben aber (mindestens) eine wichtige gemeinsame Eigenschaft,

⁵ Wir benutzen den Begriff der "Art" hier nicht im streng zoologischen Sinn. Wir könnten aber ohne weiteres beispielsweise "Lachs" durch die Spezies "Salmo salar" (Atlantischer Lachs) und "Hai" durch die Spezies "Lamna nasus" (Heringshai) ersetzen. Wir danken Herrn Dr. A. Hoffmann für diesen Hinweis.

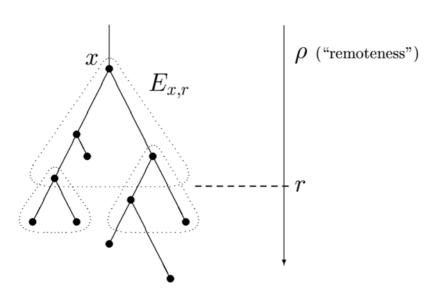
die "Mensch" nicht hat, nämlich die Eigenschaft zur Gruppe der Fische zu gehören (entsprechend dem gepunkteten Attribut "Fisch" in Abbildung 6).

Das in Nehring/Puppe (2003a) vorgeschlagene "phylogenetic tree model" entspricht der Annahme, dass die Menge der relevanten Attribute durch *alle Kladen und alle Taxa sowie deren Kombinationen* gegeben sind. Mit "Kombinationen" sind dabei Konjunktionen der entsprechenden Eigenschaften gemeint, also Eigenschaften der Form "stammt von x ab und gehört zum Taxon F". Man kann zeigen, dass das resultierende Modell durch eine geeignete Abschwächung der hierarchischen Struktur charakterisiert ist (der sogenannten "weak nesting property", siehe Nehring/Puppe 2003a, Theorem 1). Eine alternative Beschreibung lässt sich mit Hilfe eines Entferntheitsindex ρ ("remoteness") wie folgt geben (vgl. Nehring/Puppe 2003a, Theorem 2): Zu jedem phylogenetischen Baum-Modell gibt es einen Index ρ , derart dass alle relevanten Attribute von der folgenden Form sind:

$$E_{x,r} = \{y : y \ge_{ev} x \text{ und } \rho(y) \le r\}$$

Die zugrundeliegenden relevanten Eigenschaften sind also von der Form "stammt von x ab und hat einen ρ - Index kleiner oder gleich r" wie in der folgenden Abbildung gezeigt:

Abbildung 7: "Remoteness" Repräsentation des phylogenetischen Baum-Modells



Offensichtlich ist das phylogenetische Baum-Modell allgemeiner und flexibler als das kladistische Modell, da es eine größere Klasse relevanter Eigenschaften zulässt. Trotz dieses höheren Grades an Flexibilität, erlaubt aber auch das phylogenetische Baum-Modell die Bestimmung der Diversität einer Menge von Arten aus den paarweisen Distanzen ihrer Elemente. Aus Abbildung 7 ist nämlich unmittelbar ersichtlich, dass das phylogenetische Baum-Modell ein azyklisches Modell ist. Konkret lässt sich folgende Rekursionsformel angeben (Nehring/Puppe 2003a, Proposition 5.1). Es sei $\{y_1,...,y_m\}$ eine beliebige Menge derart, dass $\rho(y_1) \le \rho(y_2) \le ... \le \rho(y_m)$; dann gilt

$$v(\{y_1,...,y_m\}) = v(\{y_1\}) + \sum_{i=2,...,m} \ d(y_i,,\,\eta_i)$$

für geeignet gewählte $\eta_i \in \{y_1,...,y_{i-1}\}.$

Literaturverzeichnis

- Faith, D. (1992): Conservation Evaluation and Phylogenetic Diversity, in: Biological Conservation, 61, S. 1-10.
- Nehring, K., Puppe, C. (2001): Ein Modell von Vielfalt, in: Zeitschrift für angewandte Umweltforschung (ZAU), Sonderheft 13, S. 192-202.
- Nehring, K., Puppe, C. (2002): A Theory of Diversity, in: Econometrica, 70, S. 1155-1198.
- Nehring, K., Puppe, C. (2003a): Modelling Phylogenetic Diversity, erscheint in: Resource and Energy Economics.

Nehring, K., Puppe, C. (2003b): Diversity and Dissimilarity in Lines and Hierarchies, erscheint in: Mathematical Social Sciences.

Weitzman, D. (1992): On Diversity, in: Quarterly Journal of Economics, 107, S. 363-405.