NOTES ON THE STRUCTURE OF SCIENTIFIC LITERATURE

Joseph. D. Sneed
Center for Artificial Intelligence
Colorado School of Mines

Preliminary Draft

June 24, 1983

0. INTRODUCTION.

0.0 A method of using co-citation data to identify scientific specialties and trace their development through time is outlined here. At this point, the method is purely "theoretical" in the sense that no algorithms for implementing the method are suggested. Even at this theoretical level, the discussion is incomplete in several ways. Important formal properties of some of the constructs employed remain to be clarified. The specific way these constructs relate to "theory nets" [1] produced by philosophical "content analysis" remains to be specified in detail.

0.1 The method outlined here differs from earlier attempts [13],[14],[15], to use co-citation data to reveal the structure of science in the following ways:

A) The ultimate aim of the analysis is to compare the structure of scientific specialties (both synchronic and diachronic) revealed by co-citation analysis with the structure revealed by a certain kind of content analysis ([1]).

B) The synchronic analysis generalizes earlier efforts in that it countenances the possibility that some specialties may be sub-specialties of two or more higher level specialties.

C) The diachronic analysis:

1) takes as the units of analysis "temporal cross-sections" of the scientific literature, rather than the whole of the reference literature;

2) makes explicit the criteria it employs for identifying the same specialties in successive temporal cross-sections and locating emergent specialties in the synchronic structure of the cross-sections.

D) Uses the mathematical apparatus of fuzzy sets to describe the concepts of significance and proximity in the scientific literature as well as the concept of a scientific specialty.

E) Suggests a graphical approach to the fuzzy set analysis that pictures pieces of literature and specialties as "vectors in a multi-dimensional space whose direction indicates the nature of the specialty and magnitude indicates their relative importance

0.2 The mathematical apparatus required for this approach is developed in Secs. 1-5. The application to scientific literature is described in Secs. 6-11. The relation between this approach an other is indicated in Sec. 9. A generalization of one aspect of the approach is sketched in Secs. 12-14.

1. FUZZY SIMILARITY STRUCTURES.

1.0 A fuzzy similarity structure S is a crisp set Ω together with a fuzzy set ω and a fuzzy relation σ over ω that is reflexive and symmetric. Fuzzy similarity structures are ideally suited to dealing with situations in which the data yields numerical measures of "closeness", "similarity" or similarity" -- measured by σ -- over a set of objects Ω which differ in their "relative significance" or "weight" -- measured by ω. Formally,

$$S = \langle \Omega, \omega, \sigma \rangle$$

where

a) $$\Omega \in |\mathbf{SET}|$$

b) $$\omega \in \mathbf{FUZ}(\Omega)$$

c) $$\sigma \in \mathbf{FUZ}(\Omega \times \Omega)$$

d) $$\sigma(x,y) \leq \max(\omega(x),\omega(y)).$$

Here '$|\mathbf{SET}|$' denotes the class of all sets. '$\mathbf{SET}(X,Y)$' denotes the set of all functions from set X to set Y. For all $\Omega \in |\mathbf{SET}|$,

$$\nu \in \mathbf{SET}(\Omega,[0,1])$$

is a fuzzy set on the base set Ω. For all $x \in \Omega$, $\nu(x)$ is the degree of membership of x in the fuzzy set ν. The set of all fuzzy sets on Ω is denoted by

$$\mathbf{FUZ}(\Omega) := \mathbf{SET}(\Omega,[0,1]).$$

Condition d) requires that σ be a fuzzy relation *OVER* the fuzzy set ω. Intuitively, this means that the "importance" of similarities in the data can be no greater than the "importance" of the objects regarded as similar. It is this feature of fuzzy similarity structures that suit them for dealing with data in which everything is not equally important. The limiting case, in which everything *IS* equally important occurs when

$$\omega = \tilde{\Omega}$$

where $\tilde{\Omega}$ is the fuzzy set corresponding to the crisp set Ω, i.e., for all $x \in \Omega$

$$\tilde{\Omega}(x) = 1.$$

That σ is some kind "closeness" relation is assured be two additional requirements:

e) $$\sigma(x,x) = \omega(x)$$

f) $$\sigma(x,y) = \sigma(y,x).$$

1.1 Let '**|FSM|**' denote t he class of all fuzzy similarity structures. We take the category of fuzzy similarity structures to be

$$\mathbf{||FSM||} := \langle\ \mathbf{|FSM|},\ \mathbf{FSM},\ \mathbf{FSM_c}\ \rangle$$

here, for all S, S' $\in$ **|FSM|**

FSM(S,S') := { $\langle f,g \rangle$ | a) $f \in \mathbf{SET}(\Omega,\Omega')$

b) $g \in \mathbf{MAN}([0,1],[0,1])$

c) for all $x,y \in S$

i) $\omega'(f(x)) = g(\omega(x))$

ii) $\sigma'(f(x),f(y)) = g(\sigma(x,y))$ }

and $\mathbf{FSM_c}$ is $\mathbf{SET_c}$.

1.2 The full sub-category of **||FSM||** in which σ is a crisp relation we denote by '**||SM||**'. It is sometimes convenient to consider the full sub-categories of **||FSM||** and **||SM||** whose objects have the same "base set" Ω. These are denoted by '**||FSM(Ω)||**' and '**||SM(Ω)||**' respectively. members of **|SM|** and **|SM(Ω)|** are denoted by '$\langle \Omega, \sigma s \rangle$', suppressing the ω which, in this case, is just a crisp sub-set of Ω the "field" of σ.

1.3 Sub-categories of **||FSM||** obtained by restricting the class of morphisms -- **FSM** -- describe the the invariance properties that might be required of the functions σ and ω. We might consider:

Ordinal Fuzzy Similarity Structures:

$$\mathbf{||FSMO||} := \langle \mathbf{|FSM|}, \mathbf{FSMO}, \mathbf{FSM}_{\mathbf{c}} \rangle$$

where

$$\mathbf{FSMO}(S,S') \subset \mathbf{FSM}(S,S')$$

so that g is monotone non-decreasing.

Interval Fuzzy Similarity Structures

$$\mathbf{||FSMI||} := \langle \mathbf{|FSM|}, \mathbf{FSMI}, \mathbf{FSM}_{\mathbf{c}} \rangle$$

where

$$\mathbf{FSMI}(S,S') \subset \mathbf{FSM}(S,S')$$

so that g is positive linear.

Ratio Fuzzy Similarity Structures

$$\mathbf{||FSMR||} := \langle \mathbf{|FSM|}, \mathbf{FSMR}, \mathbf{FSM}_{\mathbf{c}} \rangle$$

where

$$\mathbf{FSMR}(S,S') \subset \mathbf{FSM}(S,S')$$

so that g is positive multiplicative.

1.4 In some situations, σ may have additional properties of interest. It may be transitive:

g) $$\sigma \bullet \sigma \subseteq_f \sigma$$

where '•' denotes max-min composition. That is, for all ν, $\mu \in$ **FUZ**$(\Omega \times \Omega)$, x, y, z $\in \Omega$,

$$(\nu \bullet \mu)(x,z) := \max_{y \in \Omega} | \min(\nu(x,y), \mu(y,z)) \quad .$$

Fuzzy similarity structures in which σ is transitive are called 'fuzzy equivalence structures'. Thus the category of

Fuzzy Equivalence Structures:

$$\|\mathbf{FEQ}\| := \langle |\mathbf{FEQ}|, \mathbf{FEQ}, \mathbf{FEQ}_c \rangle$$

The category of fuzzy equivalence structures is a full sub category of the category of fuzzy similarity structures;

$$\|\mathbf{FEQ}\| \in_f \|\mathbf{FSM}\|.$$

That is for E, E' $\in$ **|FEQ|**, **FEQ**(E,E') = **FSM**(E,E').

There are also ordinal, interval, ratio, etc. fuzzy equivalence structures. These will all be sub-categories of **||FEQ||** in which the morphisms are restricted as above.

1.5 For E $\in$ **|FEQ|**, the complement of σ

$$\bar{\sigma} := 1 - \sigma$$

which intuitively represents "separation" is an "ultrametric", i.e.,

h) $$\bar{\sigma}(x,z) \leq \max(\bar{\sigma}(x,y), \bar{\sigma}(y,z)).$$

and conversely, for S $\in$ **|FSM|** if $\bar{\sigma}$ is an ultrametric then S $\in$ **|ESM|**.

1.6 Note that if σ is an ultrametric it is also a metric, i.e.

i) $$\bar{\sigma}(x,z) \leq \bar{\sigma}(x,y) + \bar{\sigma}(y,z)$$

but not conversely. So we may distinguish the category of

Metric Fuzzy Similarity Structures

||MFSM|| in which $\bar{\sigma}$ is such that σ is a metric.

The sub-category relations are

$$\mathbf{||FEQ||} \in_f \mathbf{||MFSM||} \in_f \mathbf{||FSM||}.$$

1.7 There is a way of converting any fuzzy similarity structure into a metric fuzzy similarity structure. First, note that any $\nu \in \mathbf{FUZ}(\Omega)$ may be regarded as a vector in the "unit-positive hyper-cube" of $\mathbf{||R^nVEC||}$ where $n = N(\Omega)$. Any specific way of doing this requires that some way of ordering the members of Ω be specified so that we know which ν-values are to be plotted along which axis in $\mathbf{R^n}$. For simplicity, we will consider only finite n. Let

$$d \in \mathbf{ISET}(\mathbf{N}^{+n}, \Omega)$$

"index" Ω . For $i \in N^{+n}$, let 'x_i' denote d(i). Then for any $V \in \mathbf{|R^nVEC|}$, let

$$v_d \in \mathbf{SET}(\Omega, |V|)$$

be such that, for all $\nu \in \Omega$

$$v_d(\nu) := \langle\ \nu(x_1),\ \nu(x_2, \ldots, \nu(x_n)\ \rangle$$

Any metric on $\mathbf{||R^nVEC||}$ will clearly "impose" a metric on $\mathbf{FUZ}\ (\Omega)$.

1.8 Second, note that, for all $x \in \Omega$, we may define the "potential function of x", $\sigma_x \in \mathbf{FUZ}(\Omega)$, so that, for all $y \in \Omega$,

$$\sigma_x(y) := \sigma(x,y).$$

We may thus impose a metric on **FUZ**(Ω) by regarding the potential functions σ_x as vectors in some metric vector space over $\mathbf{R}^n$. More precisely, For all $\Omega \in$ **|SET|**, and indexing functions d there is a functor

$$\textbf{Met} : ||\textbf{FSM}(\Omega)|| \times ||\textbf{NR}^n\textbf{VEC}|| \to ||\textbf{MFSM}(\Omega)||,$$

where n = N(Ω) and **||MR^{n}VEC||** is the category of vector spaces over $\mathbf{R}^n$ cum metric, so that, for all

$$\langle \Omega, \omega, \sigma \rangle \in |\textbf{FSM}(\Omega)1$$

$$\langle V, \mu \rangle \in |\textbf{MR}^n\textbf{VEC}|$$

$$\textbf{Met}(\langle \Omega, \omega, \sigma \rangle, \langle V, \mu m \rangle) = \langle \Omega', \omega', \sigma ' \rangle$$

iff

$$\Omega = \Omega'$$

and, there is an indexing function d so that, for all x, y $\in \Omega$,

$$\omega'(x) = \mu\ (v_d(\sigma_x),v_d(\sigma_x))$$

$$\sigma(x,y) = 1 - \mu(v_d(\sigma_x),\ v_d(\sigma_y))$$

It is easy to see that the bilinearity of μ entails that σ ' is

a fuzzy relation over ω'. Further σ' is a metric on Ω' so that $\langle \Omega', \omega', \sigma' \rangle$ is a metric fuzzy similarity structure.

1.9 We may let $\langle V, \mu \rangle$ be fixed and consider

$$\textbf{Met}_\mu : ||\textbf{FSM}(\Omega)|| \to ||\textbf{MFSM}(\Omega)||.$$

For $\textbf{Met}_\mu(S) = \langle \Omega', \omega', \sigma ' \rangle$, consider the function

$$\Theta \in \textbf{SET}(\Omega' \times \Omega', [0,1])$$

so that

$$\Theta(u,v) := \frac{\bar{\sigma}'(u,v)}{\omega'(u)\ \omega'(v)}$$

In the case that μ is the familiar Euclidean metric, $\Theta(u,v)$ is just the cosine of the angle between the vectors u and v. It takes the value 1 when the angle between vectors is 0 and the value 0 when the angle is $\pi/2$ the maximum value in this situation. Θ is a similarity relation over the fuzzy set Ω' (but not over the fuzzy set ω') so that

$$\langle\ \Omega',\ \tilde{\Omega}'\ \Theta\ \rangle$$

is a fuzzy similarity structure.

1.10 Intuitively, $\Theta(v_d(x),v_d(y))$ tells us how close x and y are independently of how significant they are in terms of ω. The "magnitude" $\omega'(v_d(x))$ tells us how how significant x is in terms of w. Still more intuitively, in the vector representation of S the magnitude of the vectors tells us how ω-important they are and the angle between the vectors tells us how σ-similar they are -- with more vectors in more nearly the same direction representing more σ-similar members of Ω. That is, "directions" in the vector space represent regions of σ-similarity. Vectors of different ω-magnitudes may all lie in very nearly the same σdirection.

1.11 It is convenient to have some notation for the fuzzy similarity structure

$$\langle\ \Omega',\ \tilde{\Omega}',\ \Theta\ \rangle$$

associated with $\mathbf{Met}_{\mu}(S)$. We shall call it '$\mathbf{\Theta}_{\mu}(S)$'. We may as well think of $\mathbf{\Theta}_{\mu}$ as a functor

$$\mathbf{\Theta}_{\mu} : ||\mathbf{FSM}(\Omega)|| \rightarrow ||\mathbf{MFSM}(\Omega)||.$$

2. TYPOLOGIES.

2.0 Consider the partial ordering of crisp similarity on a finite set Ω.

$$\Sigma(\Omega) := \langle\ |\mathbf{SM}(\Omega)|,\ \subseteq\ \rangle$$

where $|\mathbf{SM}(\Omega)|$ is the set of all crisp similarity structures on Ω, and $\subseteq$ is set-theoretic inclusion. A "typology for Ω" is a sub-structure of

$$\langle\ |\mathbf{SM}(\Omega)|,\ \subseteq\ \rangle$$

which is a simple ordering containing the vacuous similarity structure on Ω, $S_0 := \langle\ \Omega,\ \Omega \times \Omega\ \rangle$, together with a "stratification function" g.

$|\mathbf{TY}(\Omega)| := \{\ \langle\ |\mathbf{SM}|,\ \subseteq,\ g\ \rangle\ |$

a) $\langle\ |\mathbf{SM}|,\ \subseteq\ \rangle \Subset \langle\ |\mathbf{SM}(\Omega)|,\ \subseteq\ \rangle$

b) $S_0 \in |\mathbf{SM}|$

c) $\langle\ |\mathbf{SM}|,\ \subseteq\ \rangle$ is a simple ordering

d) if $m = N(|\mathbf{SM}|)$ then

$$g \in \mathbf{SET}(|\mathbf{SM}|,\ \mathbf{N_0}^{+m-1})$$

so that

$$g(S_0) = 0$$

if S' $\subseteq$-covers S then

$$g(S') = g(S) - 1 \qquad \}\}$$

Members of $|\mathbf{TY}(\Omega)|$ may be conceived as sub-chains of the chains in the partial ordering $\Sigma(\Omega)$. In the case that the similarity structures are all equivalence structures $\Sigma(\Omega)$ is a "taxonomy". In this case, each level in the taxonomy is a partition of Ω.

2.1 For any $S = \langle\ \Omega,\ \sigma\ \rangle \in |\mathbf{SM}(\Omega)|$ we define the the set of S-types:

$S/\sigma := \{\ \tau \subseteq S\ |$ a) $\tau \times \tau \subseteq \sigma$

b) for all $x \in (S \sim \tau)$, there is a $y \in \tau$ so that $\langle x,y\rangle \notin \tau$ $\}$

S-types are sub-sets of Ω all of whose members are σ-related and which are "maximal" in that the contain all members of Ω that are σ-related to all their members. S-types are sometimes called 'S-cliques'. Generally, the the same sub-set of Ω may be as an S-type of two distinct similarity structures that appear in a typology for Ω. In the case that $\Sigma(\Omega)$ is a taxonomy S-types are equivalence classes and they can not intersect.

2.1 For any $T \in |\mathbf{TY}(\Omega)|$, we may consider the partial ordering of the S-types under set-theoretic inclusion.

$$\langle\ |\mathbf{SM}|//\sigma, \subseteq\ \rangle$$

where

$$|\mathbf{SM}|//\sigma := \cup\ \{\ S/\sigma \mid S \in |\mathbf{SM}|\ \}.$$

We may append to this partial ordering of S-types a level function

$$m \in \mathbf{SET}(\ |\mathbf{SM}|//\sigma,\ \mathbf{N_0}^{+m^{-1}})$$

so that, for all $\tau \in |\mathbf{SM}|//\sigma$

$$m(\tau) := \max\ \{\ m \mid \text{there is an } S \in |\mathbf{SM}| \text{ so that } \tau \in |S|/\sigma \text{ and } m = g(S)\ \}.$$

That is, $m(\tau)$ is lowest g-level in the typology at which a similarity structure appears that has τ as an S-type. Thus, for $T \in |\mathbf{TY}(\Omega)|$, we may define the:

Type Structure for a Typology

$$T//\sigma := \langle\ |\mathbf{SM}|//\sigma, \subseteq, m\ \rangle$$

$T//\sigma$ depicts the "specialization" or "inclusion" relation among S-types in the typology T. In addition, it tells us the lowest level in the typology at which an S-type occurs. For an S-type that appears in different g-levels, the m-level tells us the the "strongest" degree of similarity that still yields this S-type. Any stronger similarity relation "splits" this S-type.

2.2 The $\subseteq$-inclusion relation on $|\mathbf{SM}|//\sigma$ has the following property:

$$\text{if } \tau \neq \tau' \text{ and } \tau \subseteq \tau' \text{ then } m(\tau) < m(\tau').$$

That is whenever τ is a distinct sub-type of τ' it must be in a lower level in the type structure. However, a type τ may appear in level m without being a sub-type of any higher level type in the type structure. This happens when τ is in S/σ for more than one S in the typology. τ appears in the type structure only at the level corresponding to lowest level in the typology in which it appears.

3. TYPOLOGIES FOR FUZZY SIMILARITY STRUCTURES.

3.0 We may define a functor

$$\mathbf{Ty} : ||\mathbf{FSM}(\Omega)|| \rightarrow ||\mathbf{TY}(\Omega)||$$

$$\mathbf{Ty}(\langle \Omega, \omega, \sigma \rangle) = \langle |\mathbf{SM}|, \subseteq, g \rangle$$

iff there is a $c \in \mathbf{ISET}(D_2(\sigma), |\mathbf{SM}|)$ so that, for all $r \in D_2(\sigma)$; $x, y \in \Omega$, if $c(r) = \langle \Omega, \sigma (r) \rangle$, then

$$\langle x, y \rangle \in c(r) \text{ iff } \sigma(x,y) \geq r$$

3.1 Note that a consequence of this definition is:

$$\text{for all } r_1 \leq r_2 \in D_2(\sigma),\ g(r_1) \leq g(r_2).$$

That is, lower level in the typology correspond to increasing values of the fuzzy similarity relation σ. A bit more precisely, when

$$\mathbf{Ty}(\langle \Omega, \omega, \sigma \rangle) = \langle |\mathbf{SM}|, \subseteq, g \rangle$$

we may define

$$c' \in \mathbf{ISET}(\mathbf{N_0}^{+m^{-1}}, D_2(\sigma))$$

and

$$r := c' \bullet g$$

so that, for all S. S' ϵ **|SM|**,

$$r(S) \leq r(S') \text{ iff } g(S) \leq g(S').$$

In effect, the r-function maps the integer g-level values by their corresponding values of σ in $D_2(\sigma)$. It tells us how big the σ-similarity value that corresponds to the crisp similarity in S.

3.2) Note that the number of levels in in the typology **Ty**(S) is just the number of distinct values of the fuzzy similarity relation σ in S. In some cases, we might expect this to yield a typology that was "too fine" to be useful. That is, there would be too many levels and too many distinct types. This problem is somewhat less serious when we consider type structures, rather than typologies, because identical types appearing at different levels are no longer distinguished. Some method of further aggregating types at the same level would serve to make the resulting type structure more tractable even though the number of levels in the typology was not reduced. One way of reducing the number of levels is simply to arbitrarily choose a number of levels, say m, and divide [0,1] into m equal intervals. Then define the crisp similarity relations using values of the fuzzy similarity relation in these intervals.

3.3 Consider now, the type structure corresponding to **Ty**(S) for the fuzzy similarity structure S = $\langle \Omega, \omega, \sigma \rangle$.

$$\mathbf{Ty}(S)//\sigma = \langle \, |\mathbf{TY}(S)//\sigma|, \subseteq , m \, \rangle$$

We may define the r-level of types in **Ty**(S)//σ in the following way. With c' defined as before, define

$$r \in \mathbf{SET}(|\mathbf{TY}(S)//\sigma|, D_2(\sigma))$$

$$r := c' \bullet m.$$

That is, the r-value of τ is the value of the fuzzy σ-similarity corresponding to the lowest level S of the typology in with τ appears as an S-type. Consider S-types at the same r-level that is,

$$\{ \, \tau \in |\mathbf{TY}(S)//\sigma| \mid r(\tau) = r \, \}.$$

There will be some $S \in$ **|SM|** so that

$$S/\sigma(r) = \{ \tau \in |TY(S)//\sigma| \mid r(\tau) = r \}.$$

Thus S-types at the same r-level will not be $\subseteq$-related. Further, $\tau \in S/\sigma(r)$ is such that:

a) for all x, y $\in \tau$, $\sigma(x,y) \geq r(\tau)$

b) if $z \notin \tau$ then there is some $x \in \tau$ so that $\sigma(x,z) < r(\tau)$.

Or equivalently:

$$\min_{x,y \in \tau} \sigma(x,y) \geq r(\tau) > \max_{z \notin \tau} \min_{w \in \tau} \sigma s(w,z) .$$

Next, consider $\tau \neq \tau'$ and $\tau \subseteq \tau'$ so that $m(\tau) > m(\tau')$. Thus

$$r(\tau) > r(\tau')$$

so that members of τ have σ-similarity values at least as great as $r(\tau)$ which is strictly greater than $r(\tau')$.

3.4 It is of some interest to note that we may also obtain fuzzy similarity structures from typologies. That is, we may define a functor:

$$\mathbf{Fsm}: ||\mathbf{TY}(\Omega)|| \rightarrow ||\mathbf{FSM}(\Omega)||$$

so that

$$\mathbf{Fsm}(\langle |\mathbf{SM}|, \subseteq, g \rangle = \langle \Omega, \omega, \sigma \rangle$$

iff, if $g_m = \max \{ g(S) \mid S \in |\mathbf{SM}| \}$, then for all x, y $\in \Omega$,

$$\sigma(x,y) = 1/g_m \ (\max \{ g(S) \mid \langle x,y \rangle \in S \})$$

$$\omega(x) = \max \{ \sigma(x,z) \mid z \in \Omega \}$$

Note that **Fsm** is not a left inverse for **Ty**. Generally,

$$\mathbf{Fsm}(\mathbf{Ty}(X) = X.$$

However **Ty** is a left inverse for **Fsm** since, for all $Y \in |\mathbf{TY}(\Omega)|$,

$$\mathbf{Ty}(\mathbf{Fsm}(Y) = Y.$$

3.5 **Ty** gives a typology for any fuzzy similarity structure, but the typology it gives may not be the one we are interested in. In some cases, the "interesting" typology for fuzzy similarity structure S is obtained form **Ty** only after S has be "transformed" into another fuzzy similarity structure. We may represent this transformation by an "auto-functor"

$$\mathbf{A} : ||\mathbf{FSM}(\Omega)|| \rightarrow ||\mathbf{FSM}(\Omega)||.$$

A may be composed with **Ty** to give

$$\mathbf{Ty} \bullet \mathbf{A}$$

which also takes fuzzy similarity structures on Ω into fuzzy Ω-typologies. In particular, for any metric μ on $||\mathbf{R^n VEC}||$ where $n = N(\Omega)$,

$$\mathbf{Ty} \bullet \mathbf{Met}_\mu$$

yields a Ω-typology as does

$$\mathbf{Ty} \bullet \Theta_\mu.$$

It is a relative this latter functor that will be of most interest to us.

4. TYPOLOGIES FOR METRIZED FUZZY SIMILARITY STRUCTURES.

4.0 Consider the fuzzy similarity structure S, a metrization $\mathbf{Met}_\mu(S)$ and the associated similarity structure $\Theta_\mu S)$. If we are interested in a typology for S that ignores ω-importance then $\mathbf{Ty} \bullet \Theta_\mu(S)$, for some μ might be appropriate. Intuitively, this typology groups members of Ω into σ-similarity classes according to how nearly "in the same direction" they lie in the vector representation of σ. It ignores the magnitude of the vectors.

4.1 There is, however, still another possibility suggested by the vector representation of S. We may aggregate members of $\mathbf{Ty} \bullet \Theta_{\mu}(S)$ in a quite natural way to obtain fewer $\Theta_{\mu}(S)$-types. Let

$$\Theta_{\mu}(S) = \langle \Omega', \tilde{\Omega}', \Theta \rangle$$

and consider

$$\mathbf{Ty}(\langle \Omega', \tilde{\Omega}', \Theta \rangle) = \langle |\mathbf{SM}|, \subseteq, g \rangle.$$

Let

$$S_1 = \langle \Omega', \tilde{\Omega}', \sigma_1 \rangle = g^{-1}(1).$$

That is, S_1 is the first, non-trivial crisp similarity structure in the typology. The crisp similarity relation σ_1 is such that

$$\sigma_1(u,v) \text{ iff } \Theta(u,v) \geq r_1$$

where r_1 is the smallest non-zero value that Θ takes in Ω'.

4.2 We may aggregate each member of Ω'/σ_1 in the following way. For all $\tau^1 \in \Omega'/\sigma_1$, let

$$\bar{\tau}^1 = \bar{+} \{ v \mid v \in \tau^1 \}$$

where

$$u \bar{+} v := \tfrac{1}{2}(u + v).$$

That is, $\bar{\tau}^1$ is the vector average of all the vectors in τ^1. It is also a vector in the unit-positive hyper-cube of $\|\mathbf{R}^n\mathbf{VEC}\|$ with $n = N(\Omega)$ and thus a fuzzy set on Ω. In effect, we have transformed the crisp type τ^1 into a fuzzy type $\bar{\tau}^1$. We may now consider

$$\overline{\Omega'/\sigma_1} := \{ \bar{\tau}^1_i \mid \tau^1_i \in \Omega'/\sigma_1 \}$$

and regroup these average vectors into similarity classes using the same criterion of closeness that produced the original similarity classes Ω'/σ_1. That is, we consider

$$\overline{\Omega'/\sigma_1}/\sigma_1.$$

We may then reaggregate the members of these similarity classes. In general, this procedure may be iterated n-times to produce

$$\overline{\Omega' /^n \sigma_1}^n.$$

We may halt the procedure when it ceases to produce new equivalence classes, i.e. when

$$\overline{\Omega' /^n \sigma_1}^n = \overline{\Omega' /^n \sigma_1}^n /\sigma_1.$$

This will generally have the effect of reducing the number of "significantly different" directions representing σ_1-similarity. Note, however, that resulting $\overset{n}{\bar{\tau}}{}^1_i$ and $\overset{n}{\bar{\tau}}{}^1_j$ may still overlap in the sense of both having non-zero values for the same members of Ω. They are maximally aggregated only in the the angle between them is greater than r_1.

4.3 We may do the same thing at every lower level of $\mathbf{Ty} \bullet \Theta_{\mu}(S)$ thus producing a new structure of similarity classes.

$$\mathbf{Ty} \bullet \Theta_{\mu} S) \widetilde{\widetilde{//}}\sigma = \langle \, |\mathbf{Ty} \bullet \Theta_{\mu}(S) \widetilde{\widetilde{//}}\sigma|, \subseteq, m \, \rangle$$

As well as looking at the ordering of aggregated similarity classes by set theoretic inclusion $\subseteq$, we can also look at the ordering of their aggregation under fuzzy-set inclusion $\subseteq_f$. That is, let

$$|K(S)| := \{ \bar{\tau} \mid \tau \in |\mathbf{Ty} \bullet \Theta_{\mu}(S) \widetilde{\widetilde{//}}\sigma| \}.$$

Then let

$$K(S) := \langle |K(S)|, \subset_{-f}, m \rangle$$

where $m(\bar{\tau}) = m(\bar{\tau})$. We will call K(S) 'the fuzzy type structure for S' It is also convenient to have some notation for the set of types at a given m-level. We let, for all $m \in (D_2(\sigma))$,

$$|K(S)_m| := \{ \bar{\tau} \in |K(S)| \mid m(\bar{\tau}) = m \}.$$

4.4) Once we have decided to deal with fuzzy types, instead of crisp types, it is possible to consider other partial ordering relations on |K(S)| besides $\subset_{-f}$ together with more general "identity criteria" for fuzzy types. Some idea of how to do this is described below in Secs. 12-14.

5. TYPOLOGY KINEMATICS.

5.0 Consider a sequence of fuzzy similarity structures over disjoint sub-sets of the same set Ω with an additional fuzzy relation c on Ω that may serve to measure similarity of members of Ω across these sub-sets.

$$SQ = \langle \Omega, \bar{\Omega}, \omega, \sigma, c \rangle$$

so that

$$\bar{\Omega} \in \mathbf{SET}(N_0^+, \mathrm{Pot}(\Omega))$$

so that, for all t, t' $\in \mathbf{N_0}^+$, $t \neq t'$,

$$\Omega(t) \cap \Omega(t') = \Lambda$$

$$\omega \in \mathbf{SET}(\mathbf{N_0}^+, \mathbf{FUZ}\,\Omega))$$

$$\sigma \in \mathbf{SET}(\mathbf{N_0}^+, \mathbf{FUZ}(\Omega \times \Omega)$$

so that, for all $t \in \mathbf{N_0}^+$,

$$\omega(t) \in \mathbf{FUZ}(\Omega(t))$$

$$\sigma(t) \in \mathbf{FUZ}(\Omega(t) \times \Omega(t))$$

so that for all $t \in \mathbf{N}_0^+$

$$S(t) = \langle \Omega(t), \omega(t), \sigma(t) \rangle.$$

is a fuzzy similarity structure in the previous sense and

$$c \in \mathbf{FUZ}(\Omega \times \Omega)$$

so that, for all $x, y \in \Omega$,

$$c(x,y) = c(y,x).$$

There is no need to require that c be reflexive since we shall only consider its values for members of disjoint sets.

5.1 Consider the sequence of aggregated type structures for SQ

$$K(S(t))$$

Relate members of

$$|K(S(t))| \quad \text{and} \quad |K(S(t+1))|$$

using the c-relation by defining

$$C \in \mathbf{FUZ}(\mathbf{FUZ}(\Omega) \times \mathbf{FUZ}(\Omega))$$

so that for all $\nu, \mu \in \mathbf{FUZ}(\Omega)$

$$C(\nu, \mu) := \frac{\sum|_{x,y \in \bar{\Omega}} \nu(x)\ c(x,y)\ \mu(y)}{\sum|_{x,y \in \bar{\Omega}} \nu(x)\ \mu(y)}$$

In particular

$$C(\tau(t),\ \tau(t+1)) := \frac{\sum_{\substack{x \in \Omega(t) \\ y \in \Omega(t+1)}} | \quad \tau(t,x)\ c(x,y), \tau(t+1,y)}{\sum_{\substack{x \in \Omega(t) \\ y \in \Omega(t+1)}} | \quad \tau(t,x)\tau(t+1,y)}$$

At this point, we drop the '$^{-}$' notation over the 'τ''s. It is to be understood that the 'τ''s denote fuzzy types.

5.3 For each m-level, define the s-level successor relation on

$$|K(S(t)_m| \times |K(S(t+1)_m|$$

by, for all $s \in [0,1]$,

$$S_s(\tau(t),\tau(t+1)) \text{ iff } C(\tau(t),\tau(t+1)) > s$$

Note that the S_s relation holds only between fuzzy types of the same m-level in successive members of the sequence of fuzzy typologies. Note also that m-levels in successive fuzzy type structures in the sequence may correspond different levels of σ-similarity since $D_2(\sigma(t))$ will generally be different for different t. There may not even be the same number of m-levels in successive members of the sequence. The m-level just reflects the order of these values. That this order is "significant" enough to use as a method of identifying types through time is an empirical hypothesis about the data. In the case that the number of m-levels differs, it may be convenient to redefine m-levels in terms of intervals in in [0,1]. For the sake of intuitive clarity, we do not do this here.

5.4 We categorize change from K(S(t)) to K(S(t+1)) using properties of S_s in roughly the following way.

A) $\tau(t)$ has a unique successor $\tau(t+1)$ in K(S(t+1)) and $\tau(t+1)$ has a unique predecessor -- τ (t) is genidentical with $\tau(t+1)$

B) $\tau(t)$ has no successor in K(S(t+1)) -- $\tau(t)$ dies at t

C) $\tau(t+1)$ has no predecessor in K(S(t)) -- $\tau(t+1)$ is born at t

i) there is some $\tau'(t+1)$ genidentical with a $\tau'(t)$ and $\tau(t+1) \subseteq_f \tau'(t+1)$) $\tau(t+1)$ is a new specialization of $\tau'(t)$. The 'ed fuzzy clusters will be of a higher m-level than $\tau(t+1)$.

ii) there is some $\tau'(t+1)$ genidentical with a $\tau'(t)$ and $\tau'(t+1) \subseteq_f \tau(t+1)$) -- $\tau(t+1)$ is a new generalization of $\tau'(t)$. (Maybe we need to require that there be other specializations of $\tau(t+1)$ perhaps only in a later period.) The 'ed fuzzy clusters will be of lower level than $\tau(t+1)$.

iii) both i) and ii)

Note that there may be new kinds that are specializations (generalizations) of more than one old type.

D) $\tau(t)$ has multiple successors $\tau_i(t+1)$ each of which is preceded only by $\tau(t)$.

i) the $\tau_i(t+1)$ develop into specializations (generalizations) at later stages -- slow development of new specializations

ii) $\tau(t)$ fragments into different specializations

E) $\tau(t+1)$ has multiple predecessors $\tau_i(t)$ each of which is succeeded only by $\tau(t+1)$

F) $\tau(t)$ has multiple successors some of which have multiple predecessors.

6. LITERATURE.

6.0 Let us begin by conceiving "scientific literature" rather broadly. Let us view this literature as a set of scientific documents -- $\mathbb{L}$. The paradigm example of a scientific document is an article in a refereed scientific journal. But we may construe $\mathbb{L}$ broadly enough so that it contains books, journal articles, letters -- perhaps even films, phonograph recordings, floppy disks and tapes. Intuitively, a scientific document in this broad sense is an abstract linguistic object that is "about" a scientific subject. A document is abstract in that it may, and typically does, have several "instances". For example, each copy of a article in a scientific journal is an instance of a single document. It may not always be so clear as it is in this example whether two physical objects are instances of the same or different documents. Scholars may dispute whether two hand-written manuscripts are "copies" or "instances" of the same original document. Formally, |L is just a finite, non-empty set.

$$|L \in \mathbf{|SET|};\ \neq \Lambda \text{ and finite}$$

|L ~ the scientific literature; the set of all scientific documents

6.1 Let us understand |L to contain ALL scientific documents "from the beginning up to now". Each document has associated with it exactly one "date". Ordinarily, we might take the date of a document is its date of publication. More fundamentally though, a document's date is the time at which it was produced or completed. Some documents may be produced before they are published (but not conversely) -- e.g. the forgotten manuscript at the bottom of the desk drawer. Formally, "date" is a function from the set of documents into the non-negative real numbers.

$$d \in \mathbf{SET}(|L, \mathbf{R_0}^{+})$$

d ~ the date function

We may think of d(x) as the date of production of document x measured in years from some, not quite arbitrarily chosen, zero point. We agree to choose our "zero time" so that no documents were produced before this time. We could, of course, allow negative dates. But the notation is simpler if we do not. We shall need to consider sets of documents produced between two dates.

$$L \in \mathbf{SET}(\ \mathbf{R_0}^{+} \times \mathbf{R_0}^{+},\ \mathrm{Pot}(|L)\ )$$

$$L(t,t') := \{\ x \in |L \mid t \leq d(x) \leq t'\ \}$$

~ the literature between times t and t'

This also provides us with a way of referring to the literature produced up to date t.

$$L(0,t) \sim \text{the literature up to } t$$

6.2 In addition to the date function, the scientific literature also comes equipped with with a citation relation. The paradigm example of the citation relation holding is provided by journal articles x and y where x is cited in the list of references occurring in y. We might countenance an somewhat broader interpretation of citation in which, for example, implicit references to the contents of one document in another might count as citation. Whether one document cites another, in this broader sense, is a question to be decided by close examination of the contents of the two documents by historians. Only in the paradigm cases, can questions about the citation relation be decided on the basis of something close to purely syntactical data. Formally, the citation relation c is just a binary relation on $\mathbb{L}$.

$$c \subseteq \mathbb{L} \times \mathbb{L}$$

$$c(x,y) \sim x \text{ is cited by } y$$

It is plausible to think that documents can not cite themselves. That is, for all $x \in \mathbb{L}$,

$$\text{not } c(x,x).$$

It is also plausible to think that document x can not cite documents produced later than x. That is, for all $x, y \in \mathbb{L}$,

$$c(x,y) \text{ only if } d(x) \leq d(y).$$

There may be counter examples to this if d(x) is understood to be the publication date. But we take d(x) to be the production date and this proposition to be an analytic truth about "production" of documents. Though we co not countenance self-citation. We do countenance mutual citation. That is, for some $x, y \in \mathbb{L}$ it might be that

$$c(x,y) \text{ and } c(y,x).$$

However, by virtue of our convention relating citation and production dates, mutual citation entails identical production dates.

6.3 The analysis of the kinematics of scientific literature outlined here will focus on "literature cross-sections". A literature cross-section is just the set of all scientific documents produced during a time period of "thickness" Δ. We may envision all of $\mathbb{L}$ sliced into cross-sections of thickness Δ. These cross-sections are numbered in temporal sequence so that we may locate each cross-section by its number in the sequence. More precisely, consider

$$D \in \mathbf{SET}(\mathbf{N}^+ \times \mathbf{R}^+, \mathrm{Pot}(\mathbb{L}))$$

$$D(p,\Delta) := L((p-1)\backslash D, p\Delta)$$

~ the literature cross-section of thickness Δ at time $p\Delta$

The integer argument p in $D(p,\Delta)$ indicates the temporal position of the literature cross-section $D(p,\Delta)$. The real number Δ indicates the thickness of the cross-section. For example, $D(1982,1)$ is the set of scientific documents produced during the year 1882 (assuming we measure time in years with the customary zero point).

6.4 There is no compelling reason to take $\Delta = 1$ year. Indeed, there is some reason to think that a somewhat larger Δ might be more appropriate. Consider

$c\langle D(p,\Delta)$ ~ documents cited in $D(p,\Delta)$

$D(p-m,\Delta) \cap c\langle D(p,\Delta)$ ~ documents in cross-section $D(p-m,\Delta)$ cited in $D(p,\Delta)$

$$D(p,\Delta,m) := \frac{N(D(p-m,\Delta) \cap c\langle D(p,\Delta))}{N(c\langle D(p,\Delta))}$$

~ fraction of documents cited in $D(p,\Delta)$ that are in $D(p-m,\Delta)$

$$\Sigma_{m=1}^{m-5} D(p,\Delta,m) := PI(p,\Delta)$$

~ fraction of documents cited in $D(p,\Delta)$ that are produced in the 5 cross-sections of thickness Δ immediately preceding $D(p,\Delta)$

~ for $\Delta = 1$, the "Price Index" for the entire literature $|L$.

The value of $PI(p,1)$, for recent values of p, has been observed to be about 0.32. It might make some sense to look at $PI(p,\Delta)$ as a function of Δ and think about whether this could tell us what the optimal choice of Δ for the analysis of literature kinematics should be.

6.5) For some purposes, one might want to use a "corrected" set of documents.

$$\eta_2 : |L \rightarrow \mathbf{N}^+$$

$$\eta_2(y) := N(c\langle(y))$$

$$N(\{ x \in |L \mid c(x,y) \})$$

~ the number of documents cited in y

$$|L_c : \mathbf{N}^+ \times \mathbf{N}^+ \rightarrow Pot(|L)$$

$$|L_c(p_1,p_2) := \{ x \in |L \mid p_1 \leq \eta_2(x) \leq p_2 \}$$

~ the literature excluding review articles ($p_2(x) > p_2$) and "pontificating" ($\eta_2(x) < p_1$)

$$D : \mathbf{R} \times \mathbf{R}^+ \rightarrow Pot(|L)$$

$$D(t,\Delta) := L(t) \sim L(t - \Delta)$$

~ the source literature at t relative to interval Δ

$$D_c : \mathbf{R} \times \mathbf{R}^+ \times \mathbf{N}^+ \times \mathbf{N}^+ \rightarrow \mathrm{Pot}(|L)$$

$$D_c(t,\Delta,p_1,p_2) := D(t,\Delta) \cap |L_c(p_1,p_2)$$

~ corrected source literature

However, the method of analysis proposed here attempts to deal in a different way with the concerns that motivate using a corrected set of documents.

7. INFLUENTIAL LITERATURE.

7.0 Let us begin to characterize the "influence" of a document x by considering $q(p,\Delta,x)$ -- the number of documents in the literature cross section $D(p,\Delta)$ that cite document x. Formally,

$$q \in \mathbf{SET}(\mathbf{N}^+ \times \mathbf{R}^+ \times |L,\mathbf{N}^+)$$

so that

$$q(p,\Delta,x) := N(\{ y \mid y \in D(p,\Delta) \text{ and } c(x,y) \}).$$

~ the number of times x is cited in the literature cross section of thickness Δ at time $p\Delta$.

Alternatively,

$$q(p,\Delta,x) = N(c\rangle(x) \cap D(p,\Delta))$$

where $c\rangle(x)$ is the set of all documents that cite x. Note that $q(p,\Delta)$ is defined on all of $|L$, including $D(p,\Delta) \subseteq |L$. Documents in $D(p,\Delta)$ may cite other documents in $D(p,\Delta)$.

7.1 We may convert the function

$$q(p,\Delta) \in \mathbf{SET}(|L,\mathbf{N}^+)$$

into a measure of the relative influence of "reference documents" in the "source literature" $D(p,\Delta)$ by normalizing with respect to the number of citations received by the document most frequently cited in $D(p,\Delta)$. Consider,

$$\omega \in \mathbf{SET}(\mathbf{R} \times \mathbf{R}^+ \times |L, [0,1])$$

so that

$$\omega(p,\Delta,x) := \frac{q(p,\Delta,x)}{\max|_{y \in |L} (q(p,\Delta,y))}$$

ω -- influential literature -- is a time dependent fuzzy set on $|L$ whose degree of membership is the relative frequency of citation in the source literature of thickness Δ at time $p\Delta$. Note that the base set of ω is $|L$ and not $L(0,p\Delta)$, but for $t' > p\Delta$, $\omega(t',\Delta,x) = 0$, for all x.

7.2 The time dependent fuzzy set $\omega(p,\Delta)$ plays a role analogous to the "highly cited literature" H in Small and Griffith [15]:

$$H \in \mathbf{SET}(\mathbf{R} \times \mathbf{R}^+ \times \mathbf{N}^+, |L)$$

$$H(p,\boldsymbol{\Delta},n) := \{ x \in L(0,p\boldsymbol{\Delta}) \mid q(p,\boldsymbol{\Delta},x) \geq n \}$$

~ the documents cited more than n times in the source literature $D(p,\boldsymbol{\Delta})$

8. CO-CITATION AND PROXIMITY MEASURES.

8.0 The fundamental data we use to characterize the "topology" of scientific literature is the co-citation relation. Documents x and y are co-cited by document z just when both x and y are cited by z. Formally, the co-citation relation c^2 is a tertiary relation on $|L$:

$$c^2 \subseteq |L \times |L \times |L$$

so that, for all $x, y, z \in |L$,

$$c^2(x,y,z) \text{ iff } c(x,z) \text{ and } c(y,z)$$

$\sim$ x and y are co-cited by z

Clearly, the c^2-relation is symmetric and transitive in the first two arguments. That is, for all $x, y, w, z \in |L$,

$$c^2(x,y,z) \text{ iff } c^2 2(y,x,z).$$

and,

$$\text{if } c^2(x,y,z) \text{ and } c^2(y,w,z) \text{ then } c^2(x,w,z).$$

According to our definition, all documents are co-cited with themselves whenever they are cited. That is, for all $x, z \in |L$,

$$c^2(x,x,z) \text{ iff } c(x,z).$$

Thus, for all z, $c^2(-,-,z)$ is an equivalence relation with the two equivalence classes -- those documents cited in z and those documents not cited in z. Since we do not countenance self-citation, we have, for all $x, y \in |L$

$$\text{not } c^2(x,y,x).$$

Thus, mutual citation --

$$c(x,y) \text{ and } c(y,x)$$

-- does not entail co-citation of x and y by some z. Also note that the set of documents that co-cite x and y is just the intersection of the set of documents that cite x and the set of documents that cite y. That is, for all $x, y \in |L$,

$$c^2{>}\ (x,y) = c{>}(x) \cap c{>}(y).$$

Note that we could generalize the co-citation relation by defining c^n ($n \geq 1$) to be an n+1-ary relation on $|L$ so that

$$c^n(x_1,\ldots,x_n,z)$$

is true just when z cites all of $x_1,\ldots,x_n$. The citation relation c is c^1 in this notation.

8.1 We may obtain indicators of the proximity of documents x and y, as perceived by the producers of documents in the literature cross-section $D(p,\Delta)$ by considering the number of documents in $D(p,\Delta)$ that co-cite x and y -- $b(p,\Delta,x,y)$. Formally,

$$b \in \mathbf{SET}(\mathbf{N}^+ \times \mathbf{R}^+ \times |L \times |L, \mathbf{N}^+)$$

$$b(p,\Delta,x,y) := N(c^2\rangle(x,y) \cap D(p,\Delta))$$

~ the number of documents in literature cross section $D(p,\Delta)$ that co-cite x and y

Note that $b(p,\Delta)$ is defined for all of $|L$, including $D(p,\Delta) \subseteq |L$. Members of $D(p,\Delta)$ may be co-cited by other members of $D(p,\Delta)$. Among the properties of $b(p,\Delta)$ worth noting are the following. For all $x, y \in |L$,

$$b(p,\Delta,x,y) = b(p,\Delta,y,x)$$

$$b(p,\Delta,x,y) \leq \min (q(p,\Delta,x), q(p,\Delta,y))$$

$$b(p,\Delta,x,x) = q(p,\Delta,x).$$

8.2 We can obtain an indicatior for the relative proximity of the document pair $\langle x,y \rangle$ by normalizing $b(p,\Delta)$ with number of citations received by the document most frequently cited in $D(p,\Delta)$.

$$\lambda \in \mathbf{SET}(\mathbf{N}^+ \times \mathbf{R}^+ \times |L \times |L, [0,1])$$

$$\lambda(p,\Delta,x,y) := \frac{b(p,\Delta,x,y)}{\max|_{z \in |L} \; q(p,\Delta,z)} .$$

$\lambda(p,\Delta)$ is the relative frequency of co-citation in the source literature of thickness Δ at time $p\Delta$. Note that normalizing $\lambda(p,\Delta)$ with the maximum number of citations is the same as normalizing with respect to the maximum number of co-citations. These maxima are, in fact, the same number since we count documents as being co-cited with themselves. We could, of course define $\lambda(p,\Delta)$ only on pairs of distinct documents and then normalize with

$$\max_{z,w \in \mathbb{L};\ z \neq w} b(p,\Delta,z,w) \ .$$

8.3 λ is a time dependent fuzzy relation on the time dependent fuzzy set Δ [8] in the sense that, for all $p \in \mathbf{N}^{+}$,

$$\lambda(p,\Delta,x,y) \leq \min(\lambda(p,\Delta,x),\lambda(p,\Delta,y)).$$

λ is reflexive in λ in the sense that, for all $p \in \mathbf{N}^{+}$,

$$\lambda(p,\Delta,x,x) = \lambda(p,\Delta,x).$$

λ is symmetric in the sense that, for all $p \in \mathbf{N}^{+}$,

$$\lambda(p,\Delta,x,y) = \lambda(p,\Delta,y,x).$$

Thus $G = \langle \omega, \lambda \rangle$ is a time dependent fuzzy graph.

8.4 Other measures of the strength of co-citation links have been suggested. In [15] Small and Griffith use simply the un-normalized co-citation number $b(p,\Delta)$. This measure is clearly unnatural in the fuzzy set formulation since it is not a fuzzy relation on $\mathbb{L}$. Small [13] uses essentially

$$\mu(p,\Delta,x,y) := \frac{b(p,\Delta,x,y)}{[\ q(p,\Delta,x)\ q(p,\Delta,y)\ ]^{\frac{1}{2}}}$$

$$= \frac{\lambda(p,\Delta,x,y)}{[\ \lambda(p,\Delta,x)\ \lambda(p,\Delta,y)\ ]^{\frac{1}{2}}} \ .$$

First, note that, though $\mu(p,\Delta)$ is a fuzzy relation on $\mathbb{L}$, it is not a fuzzy relation on the fuzzy set $\lambda(p,\Delta)$ since $\mu(p,\Delta,x,y)$ may be greater than $\min(\lambda(p,\Delta,x),\lambda(p,\Delta,y))$. Second, note that $\mu(p,\Delta,x,y)$ grows large as either $\lambda(p,\Delta,x)$ or $\lambda(p,\Delta,y)$ become small. This has the effect of making the strength of co-citation links inversely proportional to the "impact" of the co-cited documents. Garfield [4] views this as an advantage of μ over b:

> "...(μ) makes the co-citation threshold less restrictive, (than b) which permits fields of relatively low activity to become visible."

It should be noted that normalizing b with q_{max} has somewhat of the same effect of raising the measure of relative co-citation strength between documents with relatively few total citations compared to what would be indicated by an unnormalized b. It does not however exaggerate this effect as μ would seem to do. Note that this exaggeration is avoided if one (as Small does) arbitrarily truncates the λ function at a certain level and considers only "highly cited" documents. A second virtue that Garfield [4] attributes to μ over b is that it gives less effect than b to co-citation links between highly cited methodology papers and other papers. Note that, for a methodology paper m, $q(m) \sim q_{max}$ so that λ, as well as μ, would have this advantage over b. In Sec. 10.4 below a proximity measure somewhat similar to μ will be suggested.

8.5 The basic fuzzy relation λ may be used to develop a number of "proximity measures" on $|L$. First, consider max-min composition of fuzzy relations on $|L$.

$$(\mu \bullet \nu)(x,z) := \max_{y \in |L} (\min (\mu(x,y), \nu(y,z)))$$

It is easy to show that $\bullet$ is associative. So we may define (suppressing the p and Δ arguments for notational convenience):

$$\lambda^1 := \lambda$$

$$\lambda^2 := \lambda \bullet \lambda$$

$$\lambda^n := \lambda^{n-1} \bullet \lambda$$

$$\lambda^\star := \sup_{n \in \{1,2,\dots\}} \lambda^n$$

Intuitively, we may think of

$$\min(\lambda(x,y), \lambda(y,z))$$

as the strength of the 2^{nd}-order co-citation link or co-citation chain between x and z through y. The idea is that a chain is no stronger than its weakest link. We may then think of $\lambda^2(x,z)$ as the strength of the strongest chain of length 2 connecting x an z. Similarly,

$$\min(\lambda(x,y_1, \lambda(y_1y_2,\ldots, \lambda(y_{n-2}y_{n-1}, \lambda(y_{n-1}z) \;)$$

is the strength of the n^{th}-order co-citation link or chain between x and z through y_1 $y_2\ldots$, $y_{n-1}$and $\lambda^n(x,z)$ is the strength of the strongest n^{th}-order co-citation link or chain of length n connecting x and z.

8.6 Since λ is reflexive

$$\lambda^{n-1} \lambda^n.$$

Note that $\lambda^n(x,z)$ is directly proportional to the max-min of the number of co-citations of pairs of documents in "chains" connecting x and z, even though the value of $\lambda^n(x,z)$ itself is normalized to be in [0,1]. Note that $\lambda^\star$ is always transitive, in the sense that

$$\lambda^\star \bullet \lambda^\star \leq \lambda^\star$$

but the λ^n are generally not transitive. It is also worth noting that no λ^n can be converted into a metric by the transformation

$$\overline{\lambda^n} := 1 - \lambda^n.$$

First,

$$\begin{aligned}\overline{\lambda^n}(x,x) &= 1 - \lambda^n(x,x)\\ &= 1 - \lambda(x)\\ &= 0\end{aligned}$$

Second, depending on y,

$$\overline{\lambda}^n(x,z) \text{ may be } \leq \text{ or } \geq \overline{\lambda}^n(x,y) + \overline{\lambda}^n(y,z).$$

Intuitively, the co-citation strength between x and z is not related in any systematic way to the the co-citation strength between x and y and between y and z.

8.7 The proximity measures λ^n separate the effect of co-citation into different fuzzy relations for each length of co-citation chain. It would be convenient to combine these relations into a single relation that summarized the whole effect of co-citation between documents. Consider now the fuzzy relation $\sigma(p,\Delta)$.

$$\sigma(p,\Delta) \in \mathbf{FUZ}(|L \times |L)$$

$$\sigma(p,\Delta)(x,y) := \bigcup_{f}\Big|_{n \geq 0} \left(\lambda^n(p,\Delta,x,y) \,/\, n \right)$$

where $\cup_f$ is the union operation for fuzzy sets defined by:

$$(\nu \cup_f \mu)(x) := \max \left(\nu(x), \mu(x) \right)$$

so that

$$\sigma(x,y) = \max\Big|_{n \geq 0} \left(\lambda^n(x,y) \,/\, n \right).$$

We may think of the degree of membership of $\langle x,y \rangle$ in σ as a measure of how close y is to x in terms of the number of co-citations of x and y. Co-citation links count in inverse proportion to their order -- hence the factor '1/n'. That is lower order co-citation links count more toward regarding y as near to x. More precisely, $\sigma(x,y)$ is the maximum value of $\lambda^n(x,y)$ / n.

8.8 Note that, for all x, y $\in$ |L

$$\sigma(x,x) = \omega(x)$$

and

$$\sigma(x,y) = \sigma(y,x)$$

Further, when

$$c(x,z) \text{ iff } c(y,z)$$

then

$$\sigma(x,y) = \omega(x) = \omega(y).$$

For the same reason as the λ^n, σ can not be converted into a metric in the obvious way. Thus, in the terminology of Sec. 1,

$$\langle\, |L, \omega, \sigma \,\rangle$$

is a non-metric fuzzy similarity structure. So we may use all the analytical tools on it that were developed in Secs. 2-5.

9. λ-CLUSTERING LITERATURE CROSS-SECTIONS.

9.0 There are a number of ways to use the λ^n to group documents in $D(p,\Delta)$ into crisp clusters (See [17].) We want to look briefly at some of them just to contrast them with the method suggested in the next section.

9.1 After Rosenfeld [12], consider

$$K^n(p,\Delta) := \{\, K(p,\Delta) \subseteq D(p,\Delta) \mid \min_{x,y \in K} \lambda^n(p+1,\Delta,x,y) \geq \max_{z \notin K} (\min_{w \in K} \lambda^n(p+1,w,z)) \,\}.$$

(Having once made it explicit, we suppress the p and Δ dependency where no confusion can result.) For $z \notin K$,

$$\min_{w \in K} \lambda^n(w,z)$$

is the weakest λ^n-link that z has with any member of K. Intuitively, it is the strength of z's λ^n-link with K.

$$\max_{z \notin K} \left(\min_{w \in K} \lambda^n(w,z) \right)$$

is then the strength of the strongest λ^n-link with K among non-members of K.

9.2 Thus, K is determined by the property that the strength of the weakest of the λ^n-links linking members of K is greater than the strength of the strongest λ^n-link between any non-member of K and K, where the strength of a non-member's link to K is the strength of its weakest link to a member of K. More explicitly then, the strength of the weakest of the λ^n-links linking members of K is greater than the strength of any non-member's weakest λ^n-link to a member of K.

9.3 It is relatively easy to see that two members of K^n, say K and K', may intersect. For n = 1, consider:

y

$\lambda^1 = .5$ $\lambda^1 = .7$

x z

$\lambda^1 = .3$

Here

$$\{x, y\} \text{ and } \{y, z\} \in K^1 \text{ and } \{x, z\} \notin K^1.$$

9.4 Generally, for K ∈ K, $S \subseteq K^n$ it will not be the case that $S \in K^n$. But if $S = K' \in K^n$, then

$$\min_{x,y \in K} \lambda^n(x,y) \geq \min_{x,y \in K'} \lambda^n(x,y).$$

That is, nested sets in K^n may be ordered according to the minimum λ^n-values that characterize them. Recall that $\lambda^n(x,z)$ is directly proportional to the max-min of the number of co-citations of pairs of documents in "chains" connecting x and z. Thus we may order members of nested sets in K^n according to the m-values where m is smallest max-min value of the number of co-citations in n^{th}-order chains connecting the pairs. That is,

$$m : K^n \rightarrow \mathbf{N}^+$$

$$m(K) := q_{max} \; [\min_{x,y \in K} \; \lambda^n(x,y)]$$

where

$$q_{max} := \max_{y \in L} \; q(p,\Delta,y)$$

Intuitively, we may think of the m-value of K as the smallest number of co-citations linking members of the in the strongest chains connecting pairs in K. The smaller the members of K^n become, the bigger the m-value characterizing them becomes. Note that all singletons {x} are in K^n for any n with the corresponding m-value being $q_{max}\sigma(x)$. Note also that intersections of members of K^n need not be in K^n.

9.5 Consider the largest K^n clusters. They include documents linked by n^{th}-order chains whose weakest links have only 1 co-citation as well as documents linked by n^{th}-order chains whose weakest links have $m > 1$ co-citations. We may think of the members of K^n arranged in the following kind of array, according to m-value

m-value	
1	largest K^n clusters; not a partition generally
2	sub-sets of level 1 clusters; possibly subsets of more than one level 1 cluster; perhaps with some intersections
3	sub-sets of level 2 clusters; etc.
	:
	:
	:
q_{max}	singletons and documents co-cited whenever cited

9.6 This suggests it is expedient to consider the sets:

$$K^n_m(p,\Delta) := \{ K(p,\Delta) \in K^n(p,\Delta) \mid \min|_{x,y \in K} \lambda^n(p+1,\Delta x,y) = m / q_{max} \}$$

Consider first K^n_m and K^n_{m+1} As we have seen, each $K \in K^n_{m+1}$ will be a sub-set of at least one $K' \in K^n_m$ obtained by successively adding documents to K that are linked to all its members by a λ^n link of strength m. There will generally be several ways to do this since, the documents eligible for addition at any stage, though linked in this way to all members of K, may not be so linked to each other. Thus, which document is added at a particular step may determine which documents may be added at subsequent steps. Likewise, each $K' \in K^n_m$ may have several sub-sets that are members of K^n_{m+1}.

9.7 Consider now K^n_m and K^{n-1}_m. Since λ is reflexive

$$\lambda^{n-1} \leq\!< \lambda^n.$$

Thus, every $K \in K^{n-1}_m$ is a sub-set (perhaps improper) of at least one

$K' \epsilon K^n_m$. Moreover, every $K' \epsilon K^n_m$ has at least one sub-set in K^{n-1}_m, though all these sub-sets may be singletons. Intuitively, every $K \epsilon K^{n-1}_m$ can be extended to a member of K^n_m by adding all documents that have λ^n links to of strength m to all members of K^{n-1}_m but fail to have λ^{n-1} links of strength m to all members. It may be possible to extend K in this manner to distinct members of K^n_m. Some documents may be linked to all members of K by λ^n links of strength m, but still not linked to each other by such links. If such an expansion is not possible then K is itself a member of K^n_m. Conversely, every $K' \epsilon K^n_m$ can be restricted to a member of K^{n-1}_m by deleting members that are not linked to every other member by a Ll^{n-1} link of strength m. There may be several ways to do this, depending on where you start. If there is no way to do it then K' is itself a member of K^{n-1}.

9.8 Consider now K^n_{m+1} and K^{n-1}_m. As we have seen, members of both these sets will generally be sub-sets of members of K^n_m. What are the possible sub-set relations among these? There appears to be no general way to sort these out. Members of K^n_{m+1} and K^{n-1}_m may intersect in all possible ways. They may even be identical.

9.9 Consider finally the limiting case of $K^\star$. Since $\lambda^\star$ is transitive, the members of $K^\star_m$ for any m partition |L. Consider $K^\star_1$. Members of $K^\star_1$ contain pairs that are linked by arbitrarily long chains whose individual links need be no stronger than 1 co-citation. This is the coarsest partition provided by the $K^\star_m$. $K^\star_2$ also partitions |L. Moreover, $K \epsilon K^\star_2$ is a sub-set (not necessarily proper) of exactly one $K' \epsilon K^\star_1$ and the members of $K^\star_2$ that are sub-sets of K' partition K'. Note though that some members of this partition of K' may be singletons. The $K^\star_m$'s for various mvalues are the clusters of Small and Griffith [15] restricted to the cross section $D(p,\Delta)$. The $\subseteq$-relation on them forms a set of "trees" or "dendograms" whose branches do not intersect. In the language of cluster analysis, they are a hierarchical taxonomy. This type of clustering is not appropriate for our purposes since it rules out the possibility that cluster at level m being a specialization of more than one cluster at level m - 1.

9.10 One possible way to see document clusters as having the structure of a specialization net in which one cluster may be a specialization of several others is just to consider the partial ordering

$$\langle \mathbb{K}(p,\Delta), \subseteq \rangle.$$

where

$$\mathbb{K}(p,\Delta) := \cup \{ m \leq m_{max}, n \geq 0 \mid K^m_n \}.$$

This would of course be the "specialty structure" of all of science at $p\Delta$ relative to the cross-section thickness Δ. Sub-nets would correspond to the specialty structures of various disciplines and sub-disciplines. It is at the level of sub-disciplines and disciplines that we should expect to see this net structure become isomorphic to a theory element net.

9.11 This structure would be a generalization of the partition structure of Griffith and Small [15] in that structures of the following kind might be present in the partial ordering.

```
[ ]                 [ ]
  \                 /
   \               /
    \             /
     \           /
      \         /
         [ ]
```

This makes it possible to think that $\langle \mathbb{K}(p,\Delta)$, @ $\rangle$ could be isomorphic to the most general kind of theory element specialization net [1].

9.12 There is however a problem with this. The additional structure is obtained by introducing different lengths of citation chains -- the parameter 'n'. This has the effect that one and the same cluster may be arrived at in two different ways. In particular, K^n_{m+1} and K^{n-1}_m may intersect. While not fatal to this approach, this suggests that there may be a good deal of redundancy in |K. More generally, some of members of |K, though distinct, may overlap to such a degree that we would want to say they represented the same speciality. This is perhaps most obvious in the case of members of K^n_{m+1} and K^{n-1}_m. When there are identical members of these two sets there is no problem with identifying these identical members with the same specialization of a member of K^n_m. We have simply characterized the same literature cluster in two different ways. But when non-identical members have intersections that a large relative to the members themselves, we might have trouble deciding whether we have one speciality or two. Less obviously, even when members of K^n_{m+1} intersect, it will not always be possible to say when these members represent one specialty or two. What we need is some way to say how much overlap among intersecting clusters is required before we may conclude that they represent the same specialty.

10. σ-CLUSTERING LITERATURE CROSS SECTIONS.

10.0 The basic idea here is to use the $\sigma(p+1,\Delta)$ relation determined by co-citation in the "source literature" in the cross-section $D(p+1,\Delta)$ to cluster documents in the immediately preceding literature cross-section $D(p,\Delta)$. The document clusters in $D(p,\Delta)$ and the relations among them represent the "structure of the scientific literature" at time $p\Delta$ as perceived by the producers of the literature in the immediately succeeding cross-section $D(p+1,\Delta)$. Intuitively, the structure of the clusters in $D(p,\Delta)$ is determined by the most myopic hindsight. By clustering only literature cross-sections of thickness Δ -- rather than the whole of the preceding literature $L(0,t)$ as Small and Griffith [15] do -- we can observe changes in the structure of the literature over time in a somewhat more convenient way.

10.1 How much of the "reference literature" we capture in $D(p,\Delta)$ depends, of course, on how big Δ is. The scientific literature is characterized by a kind of myopic hindsight in the sense that one-half the documents cited in D(p,1) will have dates later than about p-10 years [11]. Thus, if we take $\Delta = 1$, we might expect to capture somewhat more than 5 per-cent of all the literature cited in D(p+1,1) in the previous year's literature section.

10.2 For each p,

$$S(p,\Delta) := \langle D(p,\Delta),\ \omega(p+1,\Delta)|_{D(p,\Delta)},\ \sigma s(p+1,\Delta)|_{D(p,\Delta)} \rangle$$

is a fuzzy similarity structure in the sense of Sec. 1. Thus we may consider the sequence of fuzzy similarity structures (in the sense of Sec. 5)

$$SQ := \langle \mathbb{L}, S, \omega, \sigma \rangle$$

and the corresponding sequence of structures fuzzy type structures

$$K(S(p,\Delta)).$$

This is the thing we take to depict the development scientific specialties over time as reflected in the scientific literature. This is the structure that we hypothesize will turn our to be isomorphic to the theory net evolutions that represent the development of the "intellectual structure" of science as revealed by a certain kind of "content analysis" [1].

10.3 Intuitively, K(S(p,Δ)) is a structure whose elements are fuzzy sets over the basic set of documents in the literature cross-section D(p,Δ). We arrive at these "fuzzy clusters" by aggregating the documents in D(p,Δ), themselves represented as fuzzy sets on D(p,Δ). The fuzzy clusters in K(S(p,Δ)) are structured into specialties of one another by fuzzy set inclusion and organized into levels representing different degrees of proximity.

10.4 The method of aggregation employed to arrive at the fuzzy clusters is most naturally envisioned by considering the documents in D(p,Δ) to be vectors in a vector space whose dimension is just the number of documents in D(p,Δ). We may do this simply by considering σ_x for document x to be a vector whose coordinates along axes representing each of the other documents y are just $\sigma_x(y)$. Intuitively, this locates document x in the (unit-positive hyper cube of) the "intellectual space" formed by the whole collection of documents. If we impose some metric on this space, we may speak both of the magnitude and the direction of σ_x. It is not immediately clear that the common Euclidean metric is the most natural one to use here. Some thought should be given to the implications of selecting different metrics. But, for purposes of illustrating the intuitive ideas, we may stick with the Euclidean metric. The magnitude of σ_x (Sec. 1.8)

$$\mu(v_d(\sigma_x),v_d(\sigma_x))$$

is a measure of the relative importance of document x as indicated by its relative citation frequency (in cross section $D(p+1,\Delta)$. (The exact relation to the relative citation frequency is somewhat intricate in view of the definition of σ_x.) The angle between σ_x and σ_y as indicated by its cosine

$$\Theta(v_d(\sigma_x),v_d(\sigma_y)) = \frac{\sigma'(v_d(\sigma_x),v_d(\sigma_y))}{\omega'(v_d(\sigma_x))\ \omega'(v_d(\sigma_y))}$$

$$= \frac{1 - \mu(v_d(\sigma_x),v_d(\sigma_y))}{\mu(v_d(\sigma_x),v_d(\sigma_x))\ \mu(v_d(\sigma_y),v_d(\sigma_y))}$$

is a measure of the proximity of documents x and y as indicated by relative co-citation frequency. It appears to have properties similar to the proximity indicator used by Small [13] (See Sec. 8.4) though the details of how it related to other proximity indicators remain to be explored.

10.5 This way of representing documents -- viewing the potential function σ_x of document x as a vector -- suggests that intuitively we may view "directions" in a vector space of dimension $N(D(p,\Delta))$ as specifying the "subject matter orientation" of specialties. Such a direction is just an "orientation" with respect to all the literature in $D(p,\Delta)$. Supplied with a magnitude indicating its relative significance, -- a direction becomes a fuzzy set representing the specialty.

10.6 These intuitive ideas suggest first, that the appropriate way to identify specialties is to cluster the σ_x's according to the magnitude of the angle between their vector representations. Θ is, in fact, a fuzzy similarity relation so all we know about typologies for fuzzy similarity structures may be employed to construct Θ-similarity clusters at different Θ-values.

10.7 Second, the vector representation makes it intuitively natural to think of the clusters of s_x-vectors as fuzzy sets, rather than crisp sets. We make a crisp set of σ_x-vectors into into a fuzzy set simply by taking the "vector average" of all the σ_x-vectors in the set. Intuitively, it is easy to see how this average vector characterizes the "orientation" of the initial crisp set of σ_x-vectors.

10.8) Finally, it is possible to see a natural way of reducing the number of Θ-similarity clusters at a given level. You just form Θ-similarity clusters at the same level on the result of aggregating the initial crisp clusters. This will result in non-singleton clusters whenever the aggregate vectors are within the range of Θ-similarity required for cluster membership at the given level. This procedure may be iterated until it yields only singleton clusters, thereby possibly reducing the number of distinct clusters at this level.

11. σ-CLUSTER KINEMATICS.

11.0 Basically, what the sequence of type structures

$$K(S(p,\Delta))$$

does for us it provide an intertemporal identity criterion of genidentity criterion that allows us to recognize the same literature type as successive stages and also, in some situations, to identify specializations and generalizations of literature types. The way in which this sequence of type structures describes scientific change may be made more explicit as follows.

11.1 We relate clusters in $D(p,\Delta)$ and $D(p+1,\Delta)$ using relative citation intensity:

$$C(\tau(p,\Delta),\tau(p+1,\Delta)) := \frac{\Sigma_{x,y}\ \tau(p,\Delta,x)\ \hat{c}(x,y)\ \tau(p+1,\Delta,y)}{\Sigma_{x,y}\ \tau(p,\Delta,x)\tau(p+1,\Delta,y)}$$

where $\hat{c}$ is the characteristic function of the citation relation c. Thus the citation relation here plays the role of the fuzzy relation over Ω (here = |L) that connects successive fuzzy type structures in the sequence. Intuitively, the fuzzy C-relation measures the relative degree to which members of type $\tau(p+1,\Delta)$ cite members of $\tau(p,\Delta)$. 'Relative' means two things here. First, members of the types contribute to the C-value in proportion to their values as members of these types. Roughly, citations both to and from more typical members of types contribute more to the C-value for the types. Note that the fuzzy set representation of the types is essential here. Second, the C-value is normalized to reflect the overall size or significance of the types. The basic idea is that C-values for types indicated the extent to which the types are the same or "genidentical".

11.2 Ideally, we might like to have some way of dealing with this fuzzy genidentity relation without arbitrarily attenuating it to a crisp genidentity relation. Right now, I can't see a neat way to do this. It is however, relatively easy to see how to arbitrarily pick some "cut-off value" and take two successive types to be genidentical when their C-value exceeds this cut-off value. For each m-level, define the s-level successor relation on

$$|K(Sp,\Delta))_m| \times K(S(p+1,\Delta))_m|$$

$$S_s(\tau(p),\tau(p+1)) \text{ iff } C(\tau(p),\tau(p+1)) > s$$

Recall that the S_s relation holds only between fuzzy types of the same m-level at successive time periods. Intuitively, then successive types are "candidates" for being "the same" type only when their C-values are greater than s.

11.3 There is a bit more that must be said, for exceeding cut-off s will not always be sufficient for genidentity. We may however categorize change from $\tau(p)$ to $\tau(p+1)$ using properties of S_s.

A) $\tau(p)$ has a unique S_s-successor $\tau(p+1)$ in $K(S(p+1,\Delta))$ and $\tau(p+1)$ has a unique S_s-predecessor in $K(S(p,\Delta))$ -- $\tau(p)$ is genidentical with $\tau(p+1)$.

In this simple case there is no doubt that $\tau(p+1)$ is the same type of literature as $\tau(p)$. It is a candidate and the only candidate. There is one other equally simple case.

B) $\tau(p)$ has no S_s-successor in $K(S(p+1,\Delta)$ -- $\tau(p)$ dies.

There just is no type of the same level as $\tau(p)$ in $D(p+1,\Delta)$ that cites and $\tau(p)$ with sufficient intensity to be regarded as its successor.

11.4 Let's now consider somewhat more complicated cases.

C) $\tau(p+1)$ has no S_s-predecessor in $K(S(p,\Delta))$ -- $\tau(p+1)$ is born.

Here $\tau(p+1)$ can not be identified with any P-type at the same level. But it may be identifiable as a specialization or generalization of some previous type.

C-i) there is some $\tau'(p+1)$ genidentical with a $\tau'(p)$ and $\tau(p+1) \subset_f \tau'(p+1)$ -- $\tau(p+1)$ is a new specialization of $\tau'(p)$.
Note that the 'ed fuzzy clusters will be of a higher m-level than $\tau(p+1)$.

C-ii) there is some $\tau'(p+1)$ genidentical with a $\tau'(p)$ and $\tau'(p+1) \subset_f \tau(p+1)$ -- $\tau(p+1)$ is a new generalization of $\tau'(p)$.

Here the 'ed fuzzy clusters will be of lower level than $\tau(p+1)$. It is not entirely evident the this is sufficient for counting $\tau(p+1)$ as a generalization of $\tau'(p)$. We might want to require as well that there be other specializations of $\tau(p+1)$ --either at p+1 or in some later period. The intuitive idea is that a genuine generalization can't be just a "diffusion" of the literature of $\tau'(p)$ -- it must be a diffusion that ultimately results in a "concentration" in a different direction.

C-iii) both C-i) and C-ii).

Here a new specialty appears in the "interval" between two specialization linked members of $K(S(p,\Delta))$. Once again, we might want to tighten this condition by requiring that the new type ultimately exhibit specializations. Note that, in all these cases, there may be new types that are specializations (generalizations) of more than one old type.

11.5) We have considered cases where there exist unique S_s-successors or S_s-predecessors at some level that can be identified either with $\tau(p+1)$ or some specialization or generalization of it. We now consider cases where we do not have such uniqueness.

D) $\tau(p)$ has multiple S_s-successors $\tau_i(p+1)$ each of which is preceded only by $\tau(p)$.

Here the $\tau_i(p+1)$ are distinct types at the same m-level as $\tau(p)$. The can not be regarded as specializations or generalizations of $\tau(p)$, but they may well be "emerging" specializations of $\tau(p)$. That is,

D-i) the $\tau_i(p+1)$ develop into specializations (generalizations) of $\tau(p)$ at later stages.

This might characterize the slow development of new specialities and generalizations. In contrast we might have:

D-ii) $\tau(p)$ fragments into different types.

A looser identity criterion for types, might reveal all these to be equivalent (See Sec. 14 below.) But we might be driven to recognizing that, at this m-level, the development was not "normal". There was genuine fragmentation that could not be seen as specialization or generalization of preceding types. Of course the fragments of $\tau(p)$ could continue to fit into the the entire structure of $K(S(p,\Delta))$ and $K(S(p+1,\Delta))$ in the same way that $\tau(p)$ did. The fragmentation or "revolutionary" development could be localized.

11.6 We might also observe the "converse" of the situation just considered.

E) $\tau(p+1)$ has multiple S_S-predecessors $\tau_i(p)$ each of which is succeeded only by $\tau(p+1)$

Again, a weaker identity criterion for types might reveal the $\tau_i(p)$ are equivalent. If not, then this would be and example of "survival of the fittest" or what Kuhn might term "the emergence of normal science". A somewhat more intricate case of this sort would be:

F) $\tau(p)$ has multiple successors some of which have multiple predecessors

Clearly F) and perhaps D-ii) and E) do not count as "normal science" in the sense of Kuhn.

11.7 The possible scenarios for the development of types over time considered here are not offered as an exhaustive "typology of scientific change". They simply suggest some kinds of development that have been discussed by historians and philosophers of science that could be identified using the apparatus described here. It is not difficult to sketch other identifiable scenarios that would not fit so easily into available accounts of scientific change. For example, consider $\tau''(p+1)$ and $\tau'(p+1)$, both new specializations of the higher level $\tau(p)$ which is genidentical with $\tau(p+1)$. Suppose there are $\tau''(p+2)$ and $\tau'(p+2)$ genidentical with $\tau''(p+1)$ and $\tau'(p+1)$ respectively, but no $\tau(p+2)$ genidentical with $\tau(p+1)$. Do we still want to say that the "evolutions" e" and e' are specializations of the evolution e ?

12. GENERALIZED FUZZY TYPE STRUCTURES.

12.0 Given a set of fuzzy types K that sort out members of Ω into "kinds" or "types" according to their degree of σ-similarity, there are more general ways of thinking about their "logical structure" than the one we have employed for $|K(S)|$ in Sec. 10 above.

12.1 More generally, we may regard kinds of things in Ω to be structured by two fuzzy relations on fuzzy sets - an "identity relation" Θ and type inclusion relation Π. Intuitively, Θ tells us the extent to which two fuzzy sets are the same kinds of things. That is, $\tau_1$and τ_2 may be to some degree the same kind of thing as in "Shrubs and bushes are pretty much the same." Π tells us the extent to which one fuzzy set is a kind of another. That is, τ_1 may be to some degree a kind of τ_2 as in "Labrador Retrievers are largely a kind of hunting dog." Finally, kinds of things in Ω are grouped into levels of strata that correspond to how much σ-closeness is required to be very much of the kind in question.

12.2 Formally, a fuzzy type structure for the similarity structure $S = \langle \Omega, \omega, \sigma \rangle$ should have at least the following properties. It should be an ordered 4-tuple

$$\langle K, \Theta, \Pi, m \rangle$$

where

a) $K \subseteq \mathbf{FUZ}(\Omega)$ and finite

b) $\omega \in K$

c) for all $\tau \in K$, $\tau \subseteq_f \omega$

Intuitively, c) means that the degree to which something is of kind τ can be no greater than its significance in the data -- as indicated by ω. Note, however that c) does not rule out the possibility that max-values of τ may be much smaller than σ. That is, there may be identifiable kinds of Ω's that are relatively insignificant as Ω's go. For example, there might be coherent kinds of scientific literature that were relatively insignificant, as science, in comparison to other kinds.

12.3 We require of the "identity relation" Θ that

d) $\Theta \in \mathbf{FUZ}(K \times K)$

and, for all $\tau, \tau' \in K$

e) $\Theta(\tau,\tau) = 1$

f) $\Theta(\tau,\tau') = \Theta(\tau',\tau)$

g) $\Theta \bullet_+ \Theta \subseteq_f \Theta$

In g) '$\bullet_+$' denotes "max-av composition" defined by, for all ν, $\mu \in$ **FUZ**(K × K), x, y ∈ K

$$(\nu \bullet_+ \mu)(x,z) := \tfrac{1}{2} \max_{y \in K} (\nu(x,y) + \mu(y,z))$$

That is, Θ is a reflexive, symmetric and $\bullet_+$-transitive fuzzy relation on K - what is might be called 'a strong fuzzy similarity relation' on K. Requiring $\bullet_+$-transitivity, rather than •- or max-min-transitivity has the effect of imposing a stronger kind of identity criterion for kinds in K. Intuitively, Θ is some kind of a measure of how much two fuzzy sets "overlap". Requiring $\bullet_+$-transitivity for Θ in K means that the overlap of τ with τ' must be greater than the maximum over all $\tau'' \in K$ of the average overlap of τ with τ'' and τ'' with τ'. Thus, if there is some τ'' that overlaps significantly with both τ and τ' then τ must overlap significantly with τ'.

12.4 We require of the type inclusion relation that

h) $$\Pi \in \mathbf{FUZ}(K \times K)$$

so that, for all τ, $\tau' \in K$

i) $$\text{if } \tau \subseteq_f \tau' \text{ then } \Pi(\tau,\tau') = 1$$

Requirement i) means that $\subseteq_f$-inclusion entails maximum Π-inclusion, but lesser degrees of Π-inclusion may obtain even when $\subseteq_f$-inclusion does not hold. Note that i) entails that, for all $\tau \in K$,

$$\Pi(\tau,\tau) = 1$$

and together with b) that

$$\Pi(\tau,\sigma) = 1.$$

That is, Π is symmetric and σ is the "maximum element" for the Π-ordering. We further require:

j) $\quad$ if $\Pi(\tau,\tau') = \Pi(\tau',\tau)$ then $(\Theta(\tau,\tau'))^2 \geq \Pi(\tau,\tau')$.

k) $$\Pi \bullet_+ \Pi \leq \Pi$$

The idea of j) is to require that kinds that mutually include each other are very nearly the same kinds. Intuitively, Π will be some kind of measure of the portion of the difference between τ and τ' is attributable to τ''s being larger than τ. Then $\Pi(\tau,\tau') = \Pi(\tau',\tau)$ just when they have the value $\frac{1}{2}$ and i) amounts to requiring that $\Theta(\tau,\tau') \geq \frac{1}{2}^2$. Stronger versions of i) are obtained by replacing the exponent 2 by some $e > 2$. There does not appear to be any obviously "natural" choice for the strength of i). The fuzzy set $\tilde{\Omega}$ corresponding to the crisp set Ω is the maximal element for $\langle K, \Pi \rangle$, i.e.

12.5 $\quad$ K is "stratified" by m, i.e.

l) $$m \in \mathbf{SET}(K,\mathbf{N_0}^{+n})$$

so that

m) $$m(\sigma) = 0.$$

The m-function partitions K into n + 1 levels $\{ 0, 1,\ldots, n \}$ or n non trivial levels $\{ 1, 2,\ldots, n \}$, excluding ω. We need some notation for m-levels. For all $m \in \mathbf{N_0}^{+n}$, let

$$K^m := \{ \tau \in K \mid m(\tau) = m \}.$$

Intuitively, the fuzzy sets in K are intended to characterize the degree to which members of Ω satisfy some criterion of σ-closeness. The stratification function m is intended to partition K according to the strength of the criterion of σ-closeness whose degree of satisfaction the fuzzy sets characterize. That is, we want level n to contain fuzzy sets characterizing the degree to which members of Ω satisfy a criterion of σ-closeness that is stronger than the criterion of σ-closeness whose degree of satisfaction is characterized by the fuzzy sets in in level n - 1. Just how we produce fuzzy sets satisfying this intuitive requirement, we may leave open. Our discussion here need not be tied to the specific way of constructing typologies for S considered above. Thus, members of K^m characterize the degree of satisfaction of a stronger criterion of σ-closeness than do members of K^{m-1}.

13. TYPE INCLUSION II.

13.0 The Π-relation is intended to represent "type inclusion" or "specialization". That is '$\Pi(\tau, \tau') = r$' means roughly 'τ is a kind of τ'to degree r'. We explicitly countenance the possibility that fuzzy types may overlap -- not just in the sense that τ and τ' may have a non-null intersection, but also in the sense that τ'' may be a kind of τ as well as a kind of τ' to degree r, even though τ and k' do not stand in the Π-relation to any degree approaching r, in either direction.

13.1 There are at least two obvious candidates for the Π-relation. The first is just to identify Π with fuzzy set inclusion $\subseteq_f$, i.e.

$$\Pi(\tau, \tau') = 1 \text{ iff } \tau \subseteq_f \tau'.$$

This makes Π a crisp relation on K and assures that it will be a fuzzy partial ordering in the sense of i) - k) above. The second [16] is to let

$$\Pi(\tau,\tau') := \frac{\sum_{x \in \Omega} \max(0,\tau'(x) - \tau(x))}{\sum_{x \in \Omega} |\tau'(x) - \tau(x)|}$$

for $\tau \neq \tau'$ and

$$\Pi(\tau,\tau') := 0$$

for $\tau = \tau'$.

Intuitively, $\Pi(\tau,\tau')$ is a measure of the degree to which τ is "contained in" τ'. The numerator is a measure of the degree to which τ' is greater than τ, while the denomenator is a measure of the total difference between τ and τ'. Thus $\Pi(\tau,\tau')$ is a measure of the proportion of the total difference between τ and τ' attributable to τ's being greater than τ'.

13.2 This interpretation of Π has the following properties:

1) $\Pi(\tau,\tau') \in [0,1]$

2) $\Pi(\tau,\tau') = 0$ iff, for all $x \in \Omega$, $\tau(x) \leq \tau'(x)$

3) $\Pi(\tau,\tau') = 1$ iff, for all $x \in \Omega$, $\tau'(x) \leq \tau(x)$

4) $\Pi(\tau,\tau') = 1 - \Pi(\tau',\tau)$.

5) $\Pi(\tau,\tau') = \Pi(\tau',\tau)$ iff $\Pi(\tau,\tau') = \frac{1}{2}$

13.3 Defined in this way Π, is a fuzzy relation on K, but it is not guaranteed to be a fuzzy partial ordering of K. Whether it is depends on K. Condition i) above is assured by definition. Whether condition j) holds depends intuitively on how finely we distinguish members of K -- our identity criteria Θ for fuzzy types. If we lump together putative fuzzy types that overlap to a "significant degree" then j) is likely to hold for Π. Similarly, k) -- transitivity -- will hold provided that when $\Pi(\tau,\tau')$ is big, it is big because τ' is bigger than τ on x's with big values for both τ and τ'.

13.4 Intuitively, this seems to be roughly what we want. Though $\subseteq_f$ will always give us a (trivial) fuzzy partial ordering, it will exclude from the ordering much that is should intuitively be in it. (See [7] for some empirical evidence suggesting this.) Better that we have a stronger concept of type-inclusion and let it be an empirical fact about the proximity structure that it can be provided with a fuzzy typology in which type-inclusion is transitive.

13.5 For any $\Pi \in$ **FUZ**$(K \times K)$, we may define a crisp relation $\leq\bullet$ on K $\times$ K by:

$$\text{For all } \tau, \tau' \in K, \tau \leq\bullet \tau' \text{ iff } \Pi(\tau',\tau) \subseteq_f \Pi(\tau,\tau')$$

That is, $\tau \leq\bullet \tau'$ just when τ' is Π-included in τ to a lesser degree than τ is Π-included in τ'. When Π is identified as above, this entails that:

$$\tau \leq\bullet \tau' \text{ iff } \Pi(\tau,\tau') > \tfrac{1}{2}.$$

That is $\tau \leq\bullet \tau'$ just when τ is Π-included in τ' to a degree greater than $\frac{1}{2}$.

13.6 It is easy to see that if Π satisfies i) above then, for all $\tau \in K$.

$$\tau \leq\bullet \tau$$

and, if Π satisfies k) then, for all $\tau, \tau' \in K$,

$$\text{if } \tau \leq\bullet \tau' \text{ and } \tau' \leq\bullet \tau'' \text{ then } \tau \leq\bullet \tau''$$

That is, $\leq\bullet$ is reflexive and transitive when Π is reflexive and transitive. However, the converse does not hold. Condition j) on Π does not entail, for all $\tau, \tau', \tau'' \in K$

$$\text{if } \tau \leq\bullet \tau' \text{ and } \tau' \leq\bullet \tau \text{ then } \tau = \tau'.$$

The closest we can come to this is

$$\text{if } \tau \leq\bullet \tau' \text{ and } \tau' \leq\bullet \tau \text{ then } (\Theta(\tau,\tau'))^2 \geq \Pi(\tau,\tau').$$

Thus Π's being a fuzzy partial ordering (relative to Θ) in the sense of i) k) does not quite entail that $\leq\bullet$ is a crisp partial ordering. It does entail that $\leq\bullet$ is a "weak partial ordering" in the sense of being reflexive and transitive.

13.7 We then define $\sim\bullet$-equivalence.

$$\text{For all } \tau, \tau' \in K \; \tau \sim\bullet \tau' \text{ iff } \tau \leq\bullet \tau' \text{ and } \tau' \leq\bullet \tau.$$

Since $\leq\bullet$ is transitive $\sim\bullet$ is an equivalence relation. We may then say that τ' covers τ ($C(\tau',\tau)$ just when τ' $\leq\bullet$-includes τ, τ is not $\sim\bullet$-equivalent to τ' and there is no $\sim\bullet$-distinct τ'' between τ and τ' in the $\leq\bullet$-ordering. That is,

For all $\tau, \tau' \in K$, $C(\tau,' \tau)$ iff $\tau \leq\bullet \tau'$, not $\tau' \sim\bullet \tau$ and there is no $\tau'' \in K$, not $\tau'' \sim\bullet \tau$, not $\tau'' \sim\bullet \tau'$ so that $\tau \leq\bullet \tau'' \geq\bullet \tau''$

Since $\leq\bullet$ is transitive, we have

For all $\tau, \tau' \in K$, if $C(\tau',\tau)$ then not-$C(\tau,\tau')$.

14. TYPE SIMILARITY Θ.

14.0 Among the ways of measuring the relative amount of "overlap" or "intersection" among members of **FUZ**(Ω) that have been suggested ([2], [16]) is this. Let

$$\Theta(\nu,\mu) := \frac{\sum_{x \in \Omega} \nu(x)\mu(x)}{\left(\sum_{x \in \Omega} \nu(x)^2 \sum_{x \in \Omega} \mu(x)^2 \right)^{\frac{1}{2}}}$$

Θ then has the following properties:

1) $\Theta \in$ **FUZ**(**FUZ**(Ω) × **FUZ**(Ω))

2) $\Theta(\nu,\mu) = 1$ iff, for all $x \in \Omega$, $\nu(x) = \mu(x)$

3) $\lambda(\nu,\mu) = 0$ iff, for all $x \in \Omega$, $\nu(x) = 0$ iff $\mu(x) \neq 0$.

4) $\Theta(\nu,\mu) = \Theta(\mu,\nu)$

Θ will not necessarily be $\bullet_+$-transitive, nor will it necessarily be related to Π-inclusion as required by j) above. But, it is plausible to think that these conditions must be satisfied by the candidates for the Θ and Π relations we are considering in situations where the data warrant imposing a fuzzy typology.

BIBLIOGRAPHY

[1] Balzer, W. and Sneed, J., 'Generalized Net Structures of Emprical Theories, I and II", STUDIA LOGICA XXX, (3), 1977, 195-212; and STUDIA LOGICA XXXVII (2). 1978, 168-194.

[2] Bezdek, J. C ., 'Numerical Taxonomy with Fuzzy Sets', JOURNAL OF MATHEMATICAL BIOLOGY, 1, (1974), 57-71.

[3] Bezdek, J. C., and Harris, J. D., 'Fuzzy Partitions and Relations: and Axiomatic Basis for Clustering', FUZZY SETS AND SYSTEMS, 1 (1978), 111-127.

[4] Garfield, E., 'Mapping the Structure of Science', Ch. 8 in CITATION INDEXING: ITS THEORY AND APPLICATION IN SCIENCE, TECHNOLOGY AND HUMANITIES. Wiley, New York, (1979), 98-147.

[5] Gilbert, G. N., and Woolgar, S., 'The Quantitative Study of Science: an Examination of the Literature', SCIENCE STUDIES, 4, (1974), 279-294.

[6] Jardine, N. and Sibson, R., MATHEMATICAL TAXONOMY, Wiley, New York, (1971).

[7] Kempton, W., 'Category Grading and Taxonomic Relations: a Mug is a Sort of Cup', AMERICAN ETHNOLOGIST 5, (1978) 44-65.

[8] Lientz, B. P., 'On Time Dependent Fuzzy Sets', INFORMATION SCIENCES, 4, (1972) 367-376.

[9] Nadel, E., 'Citation and Co-Citation Indicators of a Phased Impact of the BCS Theory in the Physics of Super-Conductivity', SCIOMETRICS, 3, (1981), 203-221.

[10] Nadel, E., 'Incommensurability and Change: A Theoretical Discussion of Some Co-Citation Results', (unpublished manuscript) (1982).

[11] Price, D. de S., 'The Citation Cycle', COLLECTED PAPERS, American Society for Information Science, 8th Mid-Year Meeting, (1979), 195-209, reprinted in KEY PAPERS IN INFORMATION SCIENCE, B. C. Griffith (ed.) Knowledge Industry Publications, White Plains, New York, (1980)

[12] Rosenfeld, A., 'Fuzzy Graphs', PROC. U. S.-JAPAN SEMINAR ON FUZZY SETS AND THEIR APPLICATIONS, L. A. Zadeh, K. S. Fu, K. Tanka, M. Shimura (eds.), Academic Press New York, (1975), 125149.

[13] Small, H. G., 'A Co-Citation Model of a Scientific Specialty: A Longitudinal Study of Collagen Research', SOCIAL STUDIES OF SCIENCE, 7, (1977), 139-166.

[14] Small, H. G., and Greenlee, E., 'Citation Context Analysis of a Co-citation Cluster: Recombinant-DNA', SCIOMETRICS, 2, (1980), 277301.

[15] Small, H. G., and Griffith, B. C., 'The Structure of Scientific Literatures I: Identifying and Graphing Specialties', SCIENCE STUDIES. 4, (1974), 17-40.

[16] Smithson, M., 'Applications of Fuzzy Set Concepts to Behavorial Sciences', MATHEMATICAL SOCIAL SCIENCES, 2, (1982), 257-274.

[17] Yeh, R. T., and Bang, S. Y., 'Fuzzy Relations, Fuzzy Graphs and their Application to Clustering Analysis', PROC. U. S.-JAPAN SEMINAR ON FUZZY SETS AND THEIR APPLICATIONS, L. A. Zadeh, K. S. Fu, K. Tanka, M. Shimura (eds.), Academic Press New York, (1975), 125-149.

[18] Zadeh, L. A., 'Fuzzy sets and their Application to Pattern Classification and Clustering Analysis', CLASSIFICATION AND CLUSTERING, J. van Ryzin (ed.) Academic Press, New York, (1977).