Toward a Unified Theory of Relational Representations of Closed Set Systems and Phylogenetic Trees

(Preliminary Thoughts – 10/25/00)

Richard Cramer-Benjamin
Niagara University
Niagara, NY

Gary D. Crown
Wichita State University
Wichita, KS

Melvin F. Janowitz
Dimacs
Piscataway, NJ

$\mathcal{W}$, $\mathcal{V}$ are classes of structures.
Consensus function is mapping from

$$\mathcal{W}^n \mapsto \mathcal{V}.$$

How do we construct the consensus output? What are the building blocks?

For binary relations, form the ordered pairs that comprise the output.

For rooted trees, form the triples that lie in the ternary relation $|$. $ab|c$ means that for some cluster A, $a, b \in A$ and $c \notin A$.

For unrooted trees, form the quadruples that lie in the quaternary relation $ab||cd$ Here $ab||cd$ means that there is a split (A, B) such that $a, b \in A$, while $c, d \in B$.

Alternate: form the splits of the consensus object.

Working with subsets of finite nonempty X.

$\beta =$ positive integer.

$\mathcal{H} =$ collection of nonempty subsets of X
(called a β-system of sets)

$X \in \mathcal{H}$

$A \in \mathcal{H} \Longrightarrow |A| \geq \beta$

β is the *base* of $\mathcal{H}$.

$\mathcal{H}$ is *closed* if $A, B \in \mathcal{H}$ with

$$|A \cap B| \geq \beta \Longrightarrow A \cap B \in \mathcal{H}.$$

Members of $\mathcal{H}$ called *clusters*.

$\mathcal{H} =$ closed β-system of sets.

Set representation of $\mathcal{H}$:

mapping $\gamma : \mathcal{H} \mapsto \mathcal{P}(S)$,
where $S =$ fixed nonempty finite set.

Special representation for $\mathcal{H}$.

$$\gamma_\beta(\mathcal{H}) = \{(B, c) : |B| = \beta \text{ and for some } A \in \mathcal{H},\ B \subseteq A, \text{ and } c \notin A\}.$$

$(B, c) \in \gamma_\beta(\mathcal{H})$ denoted $B|c$.

$$\gamma^c_\beta(\mathcal{H}) = \{(B, c) : |B| = \beta, (B, c) \notin \gamma_\beta(\mathcal{H})\}.$$

$(B, c) \in \gamma^c_\beta(\mathcal{H})$ denoted $B \nmid c$.

Will work with Bases $\beta = 1,\ 2$.

Results extend to arbitrary β.

Base $\boldsymbol{\beta} = \mathbf{1}$

For each relation R, if
$AR = \{y : aRy \,\forall a \in A\}$, then

$$\mathcal{H}_R = X \cup \{AR : \emptyset \subset AR\}$$

is a closed 1-system of sets.
$\gamma_1(\mathcal{H})$ is a binary relation on X.
$C_x =$ smallest $A \in \mathcal{H}$ containing x,
If $R = \gamma_1^c(\mathcal{H})$, then $C_x = xR$.
Write $x|y$ if $(x, y) \in \gamma_1(\mathcal{H})$
and $x \nmid y$ otherwise.

Theorem 1(1) *For any* $\mathcal{H}, Q = \gamma_1^c(\mathcal{H})$ *is a quasiorder.*

(2) *For any quasiorder* Q, $Q = \gamma_1^c(H_Q)$.

(3) $\mathcal{H}$ *is of form* $\mathcal{H}_R$ *iff* $\mathcal{H}$ *is generated by clusters of the form* C_x.

A Concrete Example

Take $X = \{a, b, c, d, e\}$, and

$$\mathcal{H} = \{\{a\}, \{a, b\}, \{a, b, c, d\}, X\}.$$

The relation $|$ is then given by

$$a|b, a|c, a|d, a|e, b|c, b|d, b|e, c|e, d|e.$$

This is what is known as a strict weak order P, where $aPbPc, dPe$.

The relation $\nmid$ is defined by $x \nmid y \iff x|y$ fails, and it is a weak order in that it is reflexive, transitive and complete.

$\boldsymbol{\beta = 2}$

X_2 is set of 2 element subsets of X.

Write xy for $\{x, y\}$.

$\gamma_2(\mathcal{H}) \subseteq X_2 \times X$.

C_{xy} denotes smallest $A \in \mathcal{H}$ containing xy.
If $R = \gamma_2^c(\mathcal{H})$, then $C_{xy} = xyR$.
For each relation R,

$$\mathcal{H}_R = X \cup \{AR : |AR| > 1\}$$

is a closed 2-system of sets.

Definition 2 $R \subseteq X_2 \times X$ satisfies
Reflexive law if $xy \in X_2$, then

$$xyRx, \; xyRy.$$

Transitive law if $xy, ab \in X_2$, then

$$abRx, \; abRy, \; xyRw \Longrightarrow abRw.$$

Theorem 3(1) *For any* $\mathcal{H}$, $Q = \gamma_2^{\mathcal{C}}(\mathcal{H})$ *is reflexive and transitive.*

(2) *For any reflexive transitive relation* Q,

$$Q = \gamma_2^{\mathcal{C}}(\mathcal{H}_Q).$$

(3) $\mathcal{H}$ *is of the form* $\mathcal{H}_R$ *iff* $\mathcal{H}$ *is generated by clusters of the form* C_{xy}.

Write $xy|z$ if $(xy, z) \in \gamma_2(\mathcal{H})$, and $xy \nmid z$ otherwise.

For $E \subset X$, $|E| > \beta$,
removal restriction of $\mathcal{H}$ to E, $\mathcal{H}|_E = X \cup \{Y \cap E : Y \in \mathcal{H}, \beta \leq |Y \cap E| < |E|\}$.

Assume $\mathcal{W}$ is a collection of closed β-systems of sets, and that $\mathcal{W}$ is closed under the formation of removal restrictions.

Definition 4 The γ_β representation is *injective* on $\mathcal{W}$ if $\gamma_\beta(\mathcal{H}_1) = \gamma_\beta(\mathcal{H}_2)$ with $\mathcal{H}_1, \mathcal{H}_2 \in \mathcal{W} \Longrightarrow \mathcal{H}_1 = \mathcal{H}_2$.

$A \in \mathcal{H}$ is minimally generated iff $A = C_B$ with $|B| = \beta$.

Theorem 5 *When $\beta = 2$, the following are equivalent:*

(1) γ_2 *is injective on* $\mathcal{W}$

(2) *The removal restriction of any* $\mathcal{H} \in \mathcal{W}$ *to a proper cluster is a closed weak hierarchy.*

(3) *If* $\mathcal{H} \in \mathcal{W}$, *and* $A_1, A_2, A_3, A_4 \in \mathcal{H} - X$, *then* $A_1 \cap A_2 \cap A_3 \cap A_4$ *is one of* $A_1 \cap A_2 \cap A_3$, $A_1 \cap A_2 \cap A_4$, $A_2 \cap A_3 \cap A_4$.

(4) $A \in \mathcal{H} - X \Longrightarrow A$ *minimally generated.*

(5) $\mathcal{H} \in \mathcal{W} \Longrightarrow \mathcal{H} = \mathcal{H}_R$, *where* $R = \gamma_2^c(\mathcal{H})$.

So this is what is needed to be able to recapture $\mathcal{H}$ from $\gamma_2(\mathcal{H})$.

Theorem 6 *When $\beta = 1$, the following are equivalent:*

(1) γ_1 *is injective on* $\mathcal{W}$

(2) *The removal restriction of any* $\mathcal{H} \in \mathcal{W}$ *to a proper cluster is a weak order.*

(3) *If* $A_1, A_2, A_3 \in \mathcal{H} - X$, *then*

$$A_1 \cap A_2 \cap A_3 \in \{A_1 \cap A_2, A_1 \cap A_3\}.$$

(4) $A \in \mathcal{H} - X \Longrightarrow A$ *minimally generated.*

(5) $\mathcal{H} \in \mathcal{W} \Longrightarrow \mathcal{H} = \mathcal{H}_R$, *where* $R = \gamma_1^c(\mathcal{H})$.

Phylogenetic Trees

Need a common framework that includes closed 2-set systems and unrooted trees. Unrooted tree $\mathcal{T}$ on X: connected acyclic graph with no vertex of degree 2, and with the degree 1 vertices labeled by X.

Fact: Deleting any edge from $\mathcal{T}$ produces a bipartition of X called a *split*. The splits determine $\mathcal{T}$.

Fact: Choosing any vertex v as a *root* for $\mathcal{T}$ turns it into a rooted tree.

Fact: There is a notion of restriction, but it is different from the earlier version. For $A \subset X$, $\mathcal{T}|_A$ is formed by looking at subgraph generated by A, then removing any degree 2 vertex. But this is an unrooted tree on A, not on X.

Another View- (Bandelt and Dress)

split $S = (A, B)$, where A, B are complements and neither is a singleton.

$\mathcal{S}$ = collection of splits of X.

For each split S, and each $x \in X$, $S(x)$ is class of S containing x; $\overline{S}(x)$ remaining class.

$\overline{\mathcal{S}}(x)$ = closed 2-set system generated by $\{\overline{S}(x) : x \in X\}$.

Define relation $\|$ on X_2 by $ab\|cd$ if for some split (A, B), have $a, b \in A$, and $c, d \in B$.

For fixed $a \in X$,

$$\{ab\|cd : b, c, d \in X - \{a\}\}$$

determines $\overline{\mathcal{S}}(a)$ by defining $cd|b$ if $ab\|cd$.

More generally, we shall take as our starting point a relation $\mathfrak{Q}$ on X_2. For any fixed $a \in X$, want to write $cd|b$ when $ab\|cd$. Want a closed 2-set system on $X - \{a\}$ so that $|$ is the γ_2 relation. This means that $\|$ must satisfy

(1) $ab\|cd \Longrightarrow cd\|ab$,

(2) $ad\|cd$ must fail,

(3) $ab\|cd \Longrightarrow ax\|cd$ or $ay\|cd$ or $ab\|xy$ whenever $b, c, d, x, y \in X - \{a\}$.

Question: Do above conditions imply existence of a system of splits that induce $\|$?

Example 7 $X = \{a, b, c, d, e\}$. Take as the set of splits $S_1 = (abc)(de), S_2 = (bcd)(ae)$.

The quaternary representation for this is

$\mathcal{S} = ab\|de, ac\|de, bc\|de, bc\|ae, bd\|ae, cd\|ae.$

We will examine $\overline{\mathcal{S}}(a)$, $\overline{\mathcal{S}}(c)$, and $\overline{\mathcal{S}}(e)$.

$\overline{\mathcal{S}}(a)$.
Representation: $de|b, de|c, bc|e, bd|e, cd|e$
Clusters: $bcde, bcd, de$

$\overline{\mathcal{S}}(c)$.
Representation: $de|a, de|b, ae|b, ae|d.$
Clusters: $abde, ae, de.$

$\overline{\mathcal{S}}(e)$.
Representation: $ab|d, ac|d, bc|d, bc|a, bd|a, cd|a$
Clusters: $abcd, abc, bcd, bc.$

Sample Result:

Theorem 8 *To say that $\overline{\mathcal{S}}(a)$ is a rooted tree for all a is to say that*

$ab \| cd \Longrightarrow cd \| ab.$

$ad \| cd$ *fails.*

$ab \| cd \Longrightarrow at \| cd$ *or* $ab \| ct.$

At most one of $ab \| cd$, $ad \| bc$, $ac \| bd$ *can hold.*

Candidates for consensus functions:

Input is collection of relations on X_2. Reasonable assumptions might be:

$ab||cd$ in all inputs $\Longrightarrow ab||cd$ in the output.

For each $a \in X$, the ternary restriction of $||$ to a is defined by fixing a and mapping $ab||cd$ to $cd|b$. This is a ternary relation on $X - \{a\}$. If each input $||$ restricted to a in an input profile agrees with that of a second input profile, then the outputs restricted to a must be the same.

We can then use consensus methods designed for closed set systems.