From Preferences to Trees: From Social Choice to Biology

F.R. McMorris
Dean, College of Science and Letters
Department of Applied Mathematics
Department of Computer Science
Illinois Institute of Technology
Chicago, IL 60616
mcmorris@iit.edu
OUTLINE OF THIS PRESENTATION

1. Preference relations - examples.

2. Trees and classification schemes - examples.


4. Some analogs of Arrow’s theorem in bioconsensus.

5. Distance-based consensus.

6. A result unique to bioconsensus.
PREFERENCE RELATIONS

A weak order $\rho$ on $S$ is a complete ($x\rho y$ or $y\rho x$ for every $x, y \in S$) transitive ($x\rho y$ and $y\rho z$ imply $x\rho z$) relation on $S$. Defining a relation $\epsilon$ on $S$ by $x\epsilon y$ iff $x\rho y$ and $y\rho x$ an equivalence relation results. If the associated partition is $\{A_1, \ldots, A_k\}$, then the cells inherit a strict ordering: $A_1 < \ldots < A_k$. (The converse is true also.)

If $|A_i| = 1$ for all $i$ a linear order on $S$ results. Let $\mathcal{L}$ be the set of all linear orders on $S$. 
TREES AND CLUSTER SCHEMES

A tree quasi-order on $S$ is a transitive, reflexive relation $\tau$ on $S$ that satisfies the tree condition: If $x \tau y$ and $x \tau z$, then $y \tau z$ or $z \tau y$. Tree quasi-orders are formed from partitions of $S$ in the same manner as weak orders; by tree (forest) ordering the cells of the partition.

Let $Q$ be the set of all tree quasi-orders on $S$. 
A phylogenetic tree on $S$ is a graph-theoretic tree with $|S|$ end-vertices, each one labeled by an element of $S$, and having no vertices of degree two.

Let $\mathcal{T}$ denote the set of all phylogenetic trees on $S$. We assume $|S| > 4$. 
key:
(match, change, indel)

Edge Estimates by sampling for transthyretin cds (not to scale)

The famous saguaro cactus tree.
Features of every classification scheme of $S$ usually involve the notion of a *cluster*, where the clusters are constructed so that objects of the same cluster are more similar to each other than to objects of another cluster. Thus a classification on $S$ is just a set of nonempty subsets of $S$. In addition, if we let $H$ be a given classification of $S$, we require $S \in H$ as well as $\{x\} \in H$ for all $x \in S$. Since a (simple) *hypergraph* on $S$ is defined as a set of nonempty subsets of $S$, a classification scheme is foremost a type of hypergraph.
We refer to elements $A \in H$ as clusters of the hypergraph $H$. A cluster $A$ is non-trivial if $1 < |A| < |S|$.

A classification of $S$ will often be structured into something tree-like called a hierarchy. In our context, a hierarchy on $S$ is a hypergraph $H$ such that $A \cap B \in \{A, B, \emptyset\}$ for every cluster $A, B \in H$ and also $\{x\} \in H$ for all $x \in S, \emptyset \notin H$, and $S \in H$. Let $\mathcal{H}$ denote the set of all hierarchies on $S$. Let $H_\emptyset$ be the hierarchy with no non-trivial clusters.
The hierarchy on the left has non-trivial clusters \(\{a, b\}\), \(\{a, b, c\}\), \(\{a.b.c.d\}\).
SOCIAL CHOICE

Although the results to be mentioned have similar versions for weak orders, for simplicity let’s focus on $\mathcal{L}$. A consensus function on $\mathcal{L}$ is a map

$$C : \mathcal{L}^k \rightarrow \mathcal{L},$$

where $k$ is a positive integer.

(Later we will replace $\mathcal{L}$ with $\mathcal{H}, \mathcal{T},$ and $\mathcal{Q}$)
For $\mathcal{R}$ a binary relation on $S$ and $X \subseteq S$, let $R|_X$ denote the restriction of $R$ to $X \times X$. For $P$ and $P'$ in $\mathcal{R}^k$ we write $P|_X = P'|_X$ if they agree componentwise on $X$.

Terminology: $k$-tuples $P = (R_1, \ldots, R_k) \in \mathcal{R}^k$, $P' = (R'_1, \ldots, R'_k) \in \mathcal{R}^k$ are called profiles.
What properties might be satisfied by “good”, or “bad”, consensus functions? Some axioms from classical Social Choice Theory are the following:

- $C$ satisfies the *Pareto* condition if, for any $x, y \in S, (x, y) \in L_i$ for all $i$ implies $(x, y) \in C(P)$. 
• $C$ satisfies *non-imposition* if for any $x, y \in S$, there exist profiles $P$ and $P'$ such that $(x, y) \in C(P)$ and $(y, x) \in C(P')$. $C$ is *imposed* if it does not satisfy non-imposition. (An example of imposed is if $C$ is a constant map. An example of non-imposed is any $C$ that is Pareto.)

• $C$ is a *dictatorship* if there exists a $j$ such that for any profile $P$; if $(x, y) \in L_j$ for $x, y \in S$, then $(x, y) \in C(P)$. 
• $C$ is an *anti-dictatorship* if there exists a $j$ such that for any profile $P$; if $(x, y) \in L_j$ for $x, y \in S$, then $(y, x) \in C(P)$.

• $C$ satisfies *independence of irrelevant alternatives* (IIA) if, for every $X \subseteq S$ and profiles $P \& P'$, $P|_X = P'|_X$ implies $C(P)|_X = C(P')|_X$. (this is a notion of input agrees on $X \Rightarrow$ output agrees on $X$)
**Theorem** (R. Wilson, 1972): If $C$ is a consensus function on $\mathcal{L}$ that satisfies IIA, then either $C$ is a dictatorship, an anti-dictatorship, or imposed.

**Corollary** (K. Arrow, 1951, 1963): If $C$ is a consensus function on $\mathcal{L}$ that satisfies IIA and Pareto, then $C$ is a dictatorship.
Up to 1993, tournaments were the only other discrete structure for which a Wilson-type theorem existed. (Monjardet, 1978)

For $T \in \mathcal{T}$ and $X \subseteq S$ with $|X| \geq 4$, let $T|_X$ denote the tree obtained from $T$ by pruning $S - X$ and suppressing resulting vertices of degree two.
The obvious IIA analog for $T$ is to call a consensus function $C$ on $T$ stable on $X \subseteq S$ if, for profiles $P$ and $P'$,

$$ P|_X = P'|_X \Rightarrow C(P)|_X = C(P')|_X. $$

**Theorem** (McMorris & Powers, 1993): A consensus function $C$ on $T$ is stable on every $X \subseteq S$ if and only if $C$ is either a constant function or a projection.

The following corollary was used to prove this theorem.

24
First, two familiar definitions. \([ab|cd = (ab, cd)]\)

- \(C\) is a \textit{dictatorship} if there is a \(j\) such that for any profile \(P = (T_1, \ldots, T_k)\), \((ab, cd) \in T_j\) implies \((ab, cd) \in C(P)\).

- \(C\) satisfies the \textit{Pareto} condition if, for any profile \(P\), \((ab, cd) \in T_i\) for every \(i = 1, \ldots, k\) implies \((ab, cd) \in C(P)\).
**Corollary** (McMorris, 1985; McMorris & Powers, 1993): A consensus function $C$ on $T$ is stable on every $X \subseteq S$ and satisfies the Pareto condition if and only if it is a dictatorship.

It would be nice to find a proof of the theorem that does not use the corollary!
Let \( C : Q^k \rightarrow Q \).

- \( C \) is independent (stable) if, for any \( P, P' \in Q^k \) and \( X \subseteq S \), \( P|_X = P'|_X \) implies \( C(P)|_X = C(P')|_X \).

- \( C \) is Pareto if, for any \( P = (\tau_1, \ldots, \tau_k) \in Q^k \) and \( x, y \in S \), \( x \tau_i^* y \) for every \( i = 1, \ldots, k \) implies \( x C(P)^* y \).

(\( x \tau^* y \) means \( (x, y) \in \tau \) but \( (y, x) \notin \tau \))
• $C$ is *dictatorial* if there exists $j$ such that, for any $P = (\tau_1, \ldots, \tau_k) \in Q^k$ and $x, y \in S$,

$$x \tau_j^* y \text{ implies } xC(P)^* y.$$ 

**Theorem** (McMorris & Neumann, 1983): Let $|S| \geq 3$. If $C : Q^k \rightarrow Q$ is independent and Pareto, then $C$ is dictatorial.
DISTANCE BASED CONSENSUS

We next consider some POSSIBILITY results.

For any \( P = (H_1, \ldots, H_k) \in \mathcal{H}^k \) and \( X \subseteq S \), set
\[
\gamma_P(X) = \frac{|\{i : X \in H_i\}|}{k}.
\]

Define \( Maj \), the majority rule on \( \mathcal{H} \) by
\[
Maj(P) = \{X \subseteq S : \gamma_P(X) > \frac{1}{2}\}.
\]
Key Result (Margush & McMorris, 1981): For any \( P = (H_1, \ldots, H_k) \in \mathcal{H}^k \), \( \text{Maj}(P) \) is a hierarchy, so \( \text{Maj} \) is a well-defined consensus rule \( \text{Maj} : \mathcal{H}^k \rightarrow \mathcal{H} \).
Before placing $Maj$ into context, consider the general situation: Let $(X,d)$ be a finite metric space. Of interest to us is the case when $X = \mathcal{D}$, some fixed class of discrete structures. For the consensus of a profile $P$ it is appealing to utilize the metric $d$ and attempt to obtain those points in the space that are, in some sense, closest to $P$. Another important example is when $X$ is the vertex set of a connected graph and $d$ is the usual shortest path metric.
Let $\mathcal{D}^* = \bigcup_{k \geq 1} \mathcal{D}^k$. We need a measure of remoteness $\delta : \mathcal{D} \times \mathcal{D}^* \rightarrow \mathbb{R}$ so that for appropriate functions $\delta$, $\delta(D, P)$ can be thought of as a measure of how close $D \in \mathcal{D}$ is to a given profile $P$.

A (distance based) consensus function on $\mathcal{D}$ is a function of the form $C : \mathcal{D}^* \rightarrow 2^\mathcal{D} - \{\emptyset\}$, where

$$C(P) = \{D : \delta(D, P) \text{ is minimum}\}.$$
Let $P = (D_1, \ldots, D_k)$ be a profile and $D \in \mathcal{D}$. Three natural measures of remoteness are

- $\delta_1(D, P) = \sum_{i=1}^{k} d(D, D_i)$,

- $\delta_2(D, P) = \max\{d(D_i, D) : i = 1, \ldots, k\}$,

- $\delta_3(D, P) = \sum_{i=1}^{k} d(D, D_i)^2$. 
Current terminology is as follows: The median function is that consensus function, $Med$, where $Med(P) = \{ D : \delta_1(D,P) \text{ is minimum} \}$, the center function is the function $Cen$ where $Cen(P) = \{ D : \delta_2(D,P) \text{ is minimum} \}$, while the mean function, $Mea$, is defined by $Mea(P) = \{ D : \delta_3(D,P) \text{ is minimum} \}$. For studies involving classifications and hierarchies, the most widely used consensus function among these three is the median function.
If the vertex set of a connected graph is \( X = \{x_1, \ldots, x_m\} \) and \( P = (x_1, \ldots, x_m) \), then \( Cen(P) \) is the usual center of the graph, and \( Med(P) \) is the usual median of the graph.
MEDIAN FUNCTION ON HIERARCHIES

Med : \mathcal{H}^* \rightarrow 2\mathcal{H} - \{\emptyset\} such that, for any \( P = (H_1, \ldots, H_k) \in \mathcal{H}^* \),

\[
Med(P) = \{H \in \mathcal{H} : \sum_{i=1}^{k} d(H, H_i) \text{ is minimum}\},
\]

where \( d \) is the symmetric difference distance on \( \mathcal{H} \), i.e.

\[
d(H, H') = |H \cup H'| - |H \cap H'|.
\]
Theorem (Margush & McMorris, 1981): $\text{Maj}(P) \in \text{Med}(P)$ for any $P = (H_1, \ldots, H_k) \in \mathcal{H}^k$, and if $k$ is odd then $\text{Med}(P) = \{\text{Maj}(P)\}$. 
Theorem (Barthélemy & McMorris, 1986): For any $P \in \mathcal{H}^k$, all medians of $P$ in $\mathcal{H}$ are of the form $\text{Maj}(P) \cup \{A_1, \ldots, A_m\}$ such that, for $1 \leq l \leq m$, $\gamma_P(A_l) = \frac{1}{2}$ and $A_l$ is compatible with $\text{Maj}(P) \cup \{A_1, \ldots, A_{l-1}\}$.

($A \subseteq S$ is compatible with $H$ if $A \cap X \in \{\emptyset, A, X\}$ for any $X \in H$.)

Using this theorem it is easy to show that it is possible to find just two hierarchies $H_1$ and $H_2$ such that if $P = (H_1, H_2)$, we have $|\text{Med}(P)| = 2^n$. 
Let $S = abcd$. If $P = (H_1, H_2) = (H_\emptyset \cup \{ab, cd\}, H_\emptyset \cup \{ac, bd\})$, then $\text{Maj}(P) = H_\emptyset$. By the previous result, the median set has all hierarchies whose non-trivial clusters are subsets of $\{ab, cd\}$ or $\{ac, bd\}$, so $\text{Med}(P) = \{H_\emptyset\}
\cup\{H \in \mathcal{H} : H \in \{\{ab\}, \{cd\}, \{ab, cd\}, \{ac\}, \{bd\}, \{ac, bd\}\}\}$. This is the smallest example in a family for which $|\text{Med}(P)| = 2^{n+1} - 1$, where $n = |S|$. In the worst case the number of median hierarchies increases exponentially with $n$, even though calculating $\text{Maj}(P) \in \text{Med}(P)$ requires time at most polynomial in $n$. 
A CHARACTERIZATION OF \textit{Med} ON $\mathcal{H}$

Consider the following properties for a consensus function $C : \mathcal{H}^* \rightarrow 2^{\mathcal{H}} - \{\emptyset\}$:

\textbf{Consistency (C):} If $C(P) \cap C(P') \neq \emptyset$ for profiles $P$ and $P'$, then $C(PP') = C(P) \cap C(P')$, where $PP'$ is the concatenation of $P$ and $P'$, i.e., if $P = (H_1, \ldots, H_k)$ and $P' = (H'_1, \ldots, H'_m)$ then $PP' = (H_1, \ldots, H_k, H'_1, \ldots, H'_m)$.

\textbf{Faithfulness (F):} $C((H)) = \{H\}$ for all $H \in \mathcal{H}$.
The next property has been named after the Marquis de Condorcet who suggested using the median function in 1785 for voting in the French Academy.

$\frac{1}{2}$-condorcet: For any $A \subseteq S$ and profile $P$ such that $\gamma_P(A) = \frac{1}{2}$, the following holds:

$H \in C(P)$ if and only if $H \cup \{A\} \in C(P)$ provided $H \cup \{A\}$ is a hierarchy.
The following was proved in a much more general setting.

**Theorem** (McMorris, Mulder & Powers, 2000): $C$ is the median function if and only if $C$ satisfies properties (C) and (F) and is $\frac{1}{2}$-condorcet.
Although there are some results for $Cen$ and $Mea$, there are no characterization theorems for these functions on $\mathcal{H}$.

These remain interesting open problems.
AN EXAMPLE OF A RESULT UNIQUE TO BIOCONCENSUS

Now for $X \subseteq S$ and $|X| \geq 4$, let $\mathcal{T}|_X$ be the set of all phylogenies on $X$, and let $\mathcal{T}_s = \bigcup_{X \subseteq S} \mathcal{T}|_X$. Analogous to the consensus rules $C: \mathcal{T}^k \rightarrow \mathcal{T}$, an agreement or subtree rule on $\mathcal{T}$ is a function $C: \mathcal{T}^k \rightarrow \mathcal{T}_s$. (A supertree rule is a function $C: \mathcal{T}_s^k \rightarrow \mathcal{T}$)
For $T, T' \in \mathcal{T}$, $T$ is said to resolve $T'$ if $T'$ can be obtained from $T$ by contracting edges. $T$ is said to display $T'|_X$ if $T|_X = T'|_X$ or $T|_X$ resolves $T'|_X$. A profile $P = (T_1, \ldots, T_k)$ is said to display $T'|_X$ if $T'|_X$ is displayed by each $T_i$, $i = 1, \ldots, k$. Let $D(P)$ be the set of all nontrivial phylogenies (i.e., those having at least one resolved quartet) that are displayed by $P$. 
$T$ resolves $T'$

$T^*$ displays $T'$

since $\frac{4}{2, 1, 2, 3, 4, 5, 6} \star 3 = T$
Let $C$ be a subtree rule on $\mathcal{T}$, $C : \mathcal{T}^k \rightarrow \mathcal{T}_s$.

GOOD PROPERTIES FOR $C$ TO SATISFY?

- $C$ satisfies agreement if for any profile $P$, $D(P) \neq \emptyset$ implies $C(P) \in D(P)$.

- $C$ satisfies $S$-neutrality if for any profile $P$ and permutation $\phi$ of $S$, $C(\phi(P)) = \phi(C(P))$. 
**Theorem** (Day & McMorris, 2003): No subtree rule on phylogenies with six or more leaves can satisfy both agreement and $S$-neutrality.

**Proof:** Consider $P = (T_1, T_2)$ with $T_1$ and $T_2$ as in the Figure.
Readers can verify that $D(P)$ consists of $ab|de$, $af|cd$, and $bc|ef$. Then the agreement axiom requires that $C(P) \in \{ab|de, af|cd, bc|ef\}$. Consider the permutation $\phi = (aec)(bfd)$. Then $\phi(T_1) = T_1$ and $\phi(T_2) = T_2$, whence $\phi(P) = P$. Now $S$-neutrality requires that $C(\phi(P)) = \phi(C(P))$, so that $C(P) = \phi(C(P))$; but this is impossible since $\phi(ab|de) = bc|ef$, $\phi(bc|ef) = af|cd$, and $\phi(af|cd) = ab|de$. 