

On order equivalences between distance and similarity measures on sequences and trees

Martin Emms and Hector-Hugo Franco-Penya

February 6, 2012

Distance and Similarity

Distance

Similarity

Order-equivalence Notions

Alignment Duality

Neighbour and Pair Ordering

Distance to Similarity

Similarity to Distance

Empirical Investigation

- ▶ Often suggested that *similarity and distance* (on sequences and trees) are just **interchangeable** eg.

- ▶ Often suggested that *similarity and distance* (on sequences and trees) are just **interchangeable** eg.

To compare RNA structures, we need a score system, or alternatively a distance, which measures the similarity (or the difference) between the structures. These two versions of the problem – score and distance – are equivalent (Herrbach et al, 2006).

- ▶ Often suggested that *similarity and distance* (on sequences and trees) are just **interchangeable** eg.

To compare RNA structures, we need a score system, or alternatively a distance, which measures the similarity (or the difference) between the structures. These two versions of the problem – score and distance – are equivalent (Herrbach et al, 2006).

- ▶ We will distinguish **several distinct kinds of equivalence**

- ▶ Often suggested that *similarity and distance* (on sequences and trees) are just **interchangeable** eg.

To compare RNA structures, we need a score system, or alternatively a distance, which measures the similarity (or the difference) between the structures. These two versions of the problem – score and distance – are equivalent (Herrbach et al, 2006).

- ▶ We will distinguish **several distinct kinds of equivalence**
- ▶ and show that while **some kinds of equivalence hold, others do not**

Outline

Distance and Similarity

Distance

Similarity

Order-equivalence Notions

Alignment Duality

Neighbour and Pair Ordering

Distance to Similarity

Similarity to Distance

Empirical Investigation

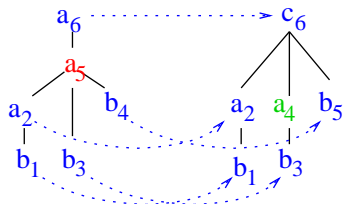
Distance on Sequences and Trees

a partial mapping $\alpha : S \mapsto T$ is a Tai mapping iff α respects left-to-right order and ancestry.

Distance on Sequences and Trees

a partial mapping $\alpha : S \mapsto T$ is a Tai mapping iff α respects left-to-right order and ancestry.

example Tai mapping α :

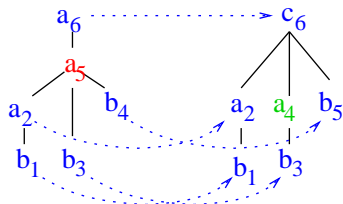


Distance on Sequences and Trees

a partial mapping $\alpha : S \mapsto T$ is a Tai mapping iff α respects left-to-right order and ancestry.

example Tai mapping α :

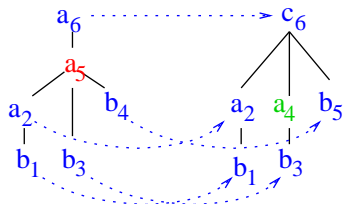
To score a mapping identify 3 sets



Distance on Sequences and Trees

a partial mapping $\alpha : S \mapsto T$ is a Tai mapping iff α respects left-to-right order and ancestry.

example Tai mapping α :



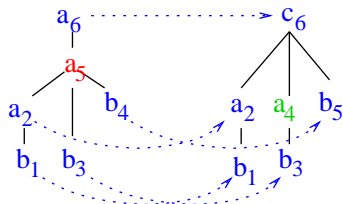
To score a mapping identify 3 sets

\mathcal{D} deletions: eg. a_5 has no image
the $i \in S$ s.t. $\forall j \in T, (i, j) \notin \alpha$

Distance on Sequences and Trees

a partial mapping $\alpha : S \mapsto T$ is a Tai mapping iff α respects left-to-right order and ancestry.

example Tai mapping α :



To score a mapping identify 3 sets

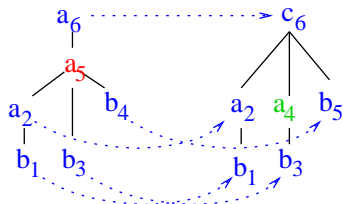
\mathcal{D} deletions: eg. a_5 has no image
the $i \in S$ s.t. $\forall j \in T, (i, j) \notin \alpha$

\mathcal{I} insertions: eg. a_4 has no source
the $j \in T$ s.t. $\forall i \in S, (i, j) \notin \alpha$

Distance on Sequences and Trees

a partial mapping $\alpha : S \mapsto T$ is a Tai mapping iff α respects left-to-right order and ancestry.

example Tai mapping α :



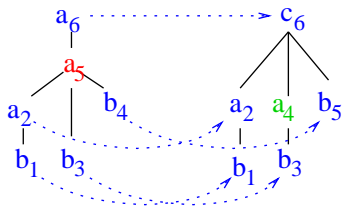
To score a mapping identify 3 sets

\mathcal{D} deletions: eg. a_5 has no image
the $i \in S$ s.t. $\forall j \in T, (i, j) \notin \alpha$

\mathcal{I} insertions: eg. a_4 has no source
the $j \in T$ s.t. $\forall i \in S, (i, j) \notin \alpha$

\mathcal{M} match/swaps: eg. a_6 goes to c_6
the $(i, j) \in \alpha$

example Tai mapping α :



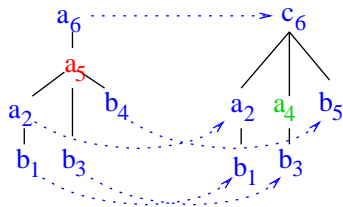
a 'cost' table C^Δ defines costs for members of $\mathcal{D}, \mathcal{I}, \mathcal{M}$

| | λ | a | b | c |
|-----------|-----------|-----|-----|-----|
| λ | | • | 1 | • |
| a | 1 | 0 | • | 1 |
| b | • | • | 0 | • |
| c | • | • | • | • |

Definition ('distance' scoring of an alignment)

$$\Delta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Delta(i^\gamma, j^\gamma) + \sum_{i \in \mathcal{D}} C^\Delta(i^\gamma, \lambda) + \sum_{j \in \mathcal{I}} C^\Delta(\lambda, j^\gamma)$$

example Tai mapping α :



a 'cost' table C^Δ defines costs for members of $\mathcal{D}, \mathcal{I}, \mathcal{M}$

| | λ | a | b | c |
|-----------|-----------|-----|-----|-----|
| λ | | • | 1 | • |
| a | 1 | 0 | • | 1 |
| b | • | • | 0 | • |
| c | • | • | • | • |

Definition ('distance' scoring of an alignment)

$$\Delta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Delta(i^\gamma, j^\gamma) + \sum_{i \in \mathcal{D}} C^\Delta(i^\gamma, \lambda) + \sum_{j \in \mathcal{I}} C^\Delta(\lambda, j^\gamma)$$

on the example $\Delta(\alpha : S \mapsto T) = 3$

Distance as min cost mapping

Definition ('distance' scoring of a tree pair)

The Tree- or Tai-distance $\Delta(S, T)$:

$$\Delta(S, T) = \min\{\Delta(\alpha : S \mapsto T) \mid \alpha \text{ is a Tai-mapping}\}$$

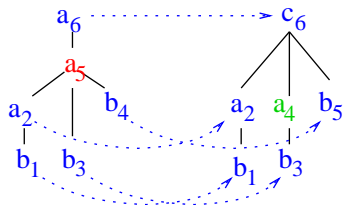
Distance as min cost mapping

Definition ('distance' scoring of a tree pair)

The Tree- or Tai-distance $\Delta(S, T)$:

$$\Delta(S, T) = \min\{\Delta(\alpha : S \mapsto T) \mid \alpha \text{ is a Tai-mapping}\}$$

the mapping α :



this mapping has minimal cost, for assumed C^Δ

hence $\Delta(S, T) = 3$

On order equivalences between distance and similarity measures on sequences and trees

└ Distance and Similarity

└ Distance

Minimal Constraints on Δ

- ▶ $\Delta(S, T)$ can be computed the Zhang/Shasha algorithm, and its correctness – ie. that it finds $\min\{\Delta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^Δ to satisfy any particular properties**

Minimal Constraints on Δ

- ▶ $\Delta(S, T)$ can be computed the Zhang/Shasha algorithm, and its correctness – ie. that it finds $\min\{\Delta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^Δ to satisfy any particular properties**

Minimal Constraints on Δ

- ▶ $\Delta(S, T)$ can be computed the Zhang/Shasha algorithm, and its correctness – ie. that it finds $\min\{\Delta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^Δ to satisfy any particular properties**
- ▶
- ▶ but some settings of C^Δ make little sense, in particular

negative deletion/insertion cost-entries \Rightarrow a supertree (or subtree) of S is 'closer' to S than S itself

Minimal Constraints on Δ

- ▶ $\Delta(S, T)$ can be computed the Zhang/Shasha algorithm, and its correctness – ie. that it finds $\min\{\Delta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^Δ to satisfy any particular properties**
- ▶
- ▶ but some settings of C^Δ make little sense, in particular

negative deletion/insertion cost-entries \Rightarrow a supertree (or subtree) of S is 'closer' to S than S itself

- ▶ this motivates the nearly universal adopted non-negativity assumption

$$\forall x, y \in \Sigma (C^\Delta(x, y) \geq 0, C^\Delta(x, \lambda) \geq 0, C^\Delta(\lambda, y) \geq 0) \quad (1)$$

Outline

Distance and Similarity

Distance

Similarity

Order-equivalence Notions

Alignment Duality

Neighbour and Pair Ordering

Distance to Similarity

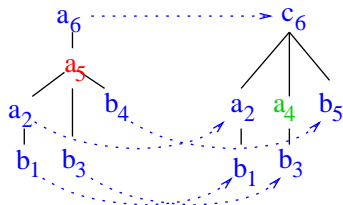
Similarity to Distance

Empirical Investigation

'similarity': a widely followed alternative, seeks to **maximize** a score assigned to an alignment

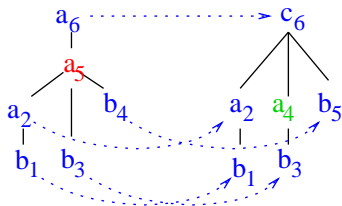
'similarity': a widely followed alternative, seeks to **maximize** a score assigned to an alignment

example Tai mapping α :



'similarity': a widely followed alternative, seeks to **maximize** a score assigned to an alignment

example Tai mapping α :

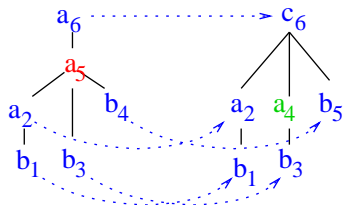


uses a 'similarity' table to assign scores to members of \mathcal{D} , \mathcal{I} , \mathcal{M} eg.

| | λ | a | b | c |
|-----------|-----------|-----|-----|-----|
| λ | | ● | 0 | ● |
| a | 0 | 2 | ● | 1 |
| b | ● | ● | 2 | ● |
| c | ● | ● | ● | ● |

'similarity': a widely followed alternative, seeks to **maximize** a score assigned to an alignment

example Tai mapping α :



uses a 'similarity' table to assign scores to members of \mathcal{D} , \mathcal{I} , \mathcal{M} eg.

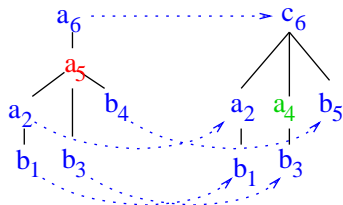
| | λ | a | b | c |
|-----------|-----------|-----|-----|-----|
| λ | | ● | 0 | ● |
| a | 0 | 2 | ● | 1 |
| b | ● | ● | 2 | ● |
| c | ● | ● | ● | ● |

Definition ('similarity' scoring of an alignment)

$$\Theta(\alpha : \mathcal{S} \mapsto \mathcal{T}) = \sum_{(i,j) \in \mathcal{M}} C^\ominus(i^\gamma, j^\gamma) - \sum_{i \in \mathcal{D}} C^\ominus(i^\gamma, \lambda) - \sum_{j \in \mathcal{I}} C^\ominus(\lambda, j^\gamma)$$

'similarity': a widely followed alternative, seeks to **maximize** a score assigned to an alignment

example Tai mapping α :



uses a 'similarity' table to assign scores to members of \mathcal{D} , \mathcal{I} , \mathcal{M} eg.

| | λ | a | b | c |
|-----------|-----------|-----|-----|-----|
| λ | | ● | 0 | ● |
| a | 0 | 2 | ● | 1 |
| b | ● | ● | 2 | ● |
| c | ● | ● | ● | ● |

Definition ('similarity' scoring of an alignment)

$$\Theta(\alpha : \mathcal{S} \mapsto \mathcal{T}) = \sum_{(i,j) \in \mathcal{M}} C^\ominus(i^\gamma, j^\gamma) - \sum_{i \in \mathcal{D}} C^\ominus(i^\gamma, \lambda) - \sum_{j \in \mathcal{I}} C^\ominus(\lambda, j^\gamma)$$

eg. $\Theta(\alpha) = 9$

Similarity as max score mapping

Definition ('similarity' scoring of a tree pair)

The Tree- or Tai-similarity $\Theta(S, T)$ between two trees S and T :

$$\Theta(S, T) = \max\{\Theta(\alpha : S \mapsto T) \mid \alpha \text{ is a Tai-mapping}\}$$

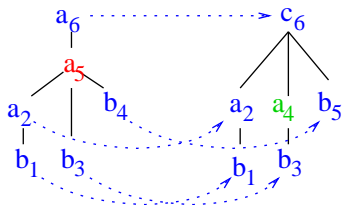
Similarity as max score mapping

Definition ('similarity' scoring of a tree pair)

The Tree- or Tai-similarity $\Theta(S, T)$ between two trees S and T :

$$\Theta(S, T) = \max\{\Theta(\alpha : S \mapsto T) \mid \alpha \text{ is a Tai-mapping}\}$$

the mapping α :



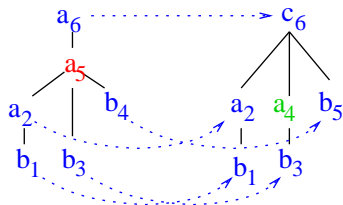
Similarity as max score mapping

Definition ('similarity' scoring of a tree pair)

The Tree- or Tai-similarity $\Theta(S, T)$ between two trees S and T :

$$\Theta(S, T) = \max\{\Theta(\alpha : S \mapsto T) \mid \alpha \text{ is a Tai-mapping}\}$$

the mapping α :



this mapping has maximum score,
for assumed C^\ominus

hence $\Theta(S, T) = 9$

Minimal Constraints on Θ

- ▶ $\Theta(S, T)$ can be computed by simple modifications of the Zhang/Shasha algorithm and its ie. that it finds $\max\{\Theta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^Θ to satisfy any particular properties**

Minimal Constraints on Θ

- ▶ $\Theta(S, T)$ can be computed by simple modifications of the Zhang/Shasha algorithm and its ie. that it finds $\max\{\Theta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^Θ to satisfy any particular properties**
- ▶ but some settings of C^Θ make little sense, in particular

negative deletion/insertion C^Θ -entries \Rightarrow a supertree (or subtree) of S is 'more similar' to S than S itself

Minimal Constraints on Θ

- ▶ $\Theta(S, T)$ can be computed by simple modifications of the Zhang/Shasha algorithm and its ie. that it finds $\max\{\Theta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^\ominus to satisfy any particular properties**
- ▶ but some settings of C^\ominus make little sense, in particular

negative deletion/insertion C^\ominus -entries \Rightarrow a supertree (or subtree) of S is 'more similar' to S than S itself

- ▶ this motivates the nearly universal adopted non-negativity assumption

$$\forall x, y \in \Sigma (C^\ominus(x, \lambda) \geq 0, C^\ominus(\lambda, y) \geq 0) \quad (2)$$

Minimal Constraints on Θ

- ▶ $\Theta(S, T)$ can be computed by simple modifications of the Zhang/Shasha algorithm and its ie. that it finds $\max\{\Theta(\alpha : S \mapsto T)\}$ – **does not require the cost-table C^\ominus to satisfy any particular properties**
- ▶ but some settings of C^\ominus make little sense, in particular

negative deletion/insertion C^\ominus -entries \Rightarrow a supertree (or subtree) of S is 'more similar' to S than S itself

- ▶ this motivates the nearly universal adopted non-negativity assumption

$$\forall x, y \in \Sigma (C^\ominus(x, \lambda) \geq 0, C^\ominus(\lambda, y) \geq 0) \quad (2)$$

- ▶ For the C^\ominus -entries which are not deletions or insertions, it is quite common in biological sequence comparison to have both positive and negative entries

Summary

Tree Distance

$$\Delta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Delta(i^\gamma, j^\gamma) + \sum_{i \in \mathcal{D}} C^\Delta(i^\gamma, \lambda) + \sum_{j \in \mathcal{I}} C^\Delta(\lambda, j^\gamma)$$

Tree Similarity

$$\Theta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Theta(i^\gamma, j^\gamma) - \sum_{i \in \mathcal{D}} C^\Theta(i^\gamma, \lambda) - \sum_{j \in \mathcal{I}} C^\Theta(\lambda, j^\gamma)$$

Summary

Tree Distance

$$\Delta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Delta(i^\gamma, j^\gamma) + \sum_{i \in \mathcal{D}} C^\Delta(i^\gamma, \lambda) + \sum_{j \in \mathcal{I}} C^\Delta(\lambda, j^\gamma)$$

$$\Delta(S, T) = \min(\{\Delta(\alpha : S \mapsto T)\})$$

Tree Similarity

$$\Theta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Theta(i^\gamma, j^\gamma) - \sum_{i \in \mathcal{D}} C^\Theta(i^\gamma, \lambda) - \sum_{j \in \mathcal{I}} C^\Theta(\lambda, j^\gamma)$$

$$\Theta(S, T) = \max(\{\Theta(\alpha : S \mapsto T)\})$$

Summary

Tree Distance

$$\Delta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Delta(i^\gamma, j^\gamma) + \sum_{i \in \mathcal{D}} C^\Delta(i^\gamma, \lambda) + \sum_{j \in \mathcal{I}} C^\Delta(\lambda, j^\gamma)$$

$$\Delta(S, T) = \min(\{\Delta(\alpha : S \mapsto T)\}) \quad \begin{array}{l} C^\Delta(x, y) \geq 0 \\ C^\Delta(x, \lambda) \geq 0, C^\Delta(\lambda, y) \geq 0 \end{array}$$

Tree Similarity

$$\Theta(\alpha : S \mapsto T) = \sum_{(i,j) \in \mathcal{M}} C^\Theta(i^\gamma, j^\gamma) - \sum_{i \in \mathcal{D}} C^\Theta(i^\gamma, \lambda) - \sum_{j \in \mathcal{I}} C^\Theta(\lambda, j^\gamma)$$

$$\Theta(S, T) = \max(\{\Theta(\alpha : S \mapsto T)\}) \quad C^\Theta(x, \lambda) \geq 0, C^\Theta(\lambda, y) \geq 0$$

Outline

Distance and Similarity

Distance

Similarity

Order-equivalence Notions

Alignment Duality

Neighbour and Pair Ordering

Distance to Similarity

Similarity to Distance

Empirical Investigation

a 'distance' Δ scoring of alignments induce orderings three different kinds entities

a 'distance' Δ scoring of alignments induce orderings three different kinds entities

Alignment ordering Given fixed S , and fixed T , rank the possible *alignments*
 $\alpha : S \mapsto T$ by $\Delta(\alpha : S \mapsto T)$

a 'distance' Δ scoring of alignments induce orderings three different kinds entities

Alignment ordering Given fixed S , and fixed T , rank the possible *alignments* $\alpha : S \mapsto T$ by $\Delta(\alpha : S \mapsto T)$

Neighbour ordering Given fixed S , and varying candidate neighbours T_i , rank the *neighbours* T_i by $\Delta(S, T_i)$ – typically used in k-NN classification.

a 'distance' Δ scoring of alignments induce orderings three different kinds entities

Alignment ordering Given fixed S , and fixed T , rank the possible *alignments* $\alpha : S \mapsto T$ by $\Delta(\alpha : S \mapsto T)$

Neighbour ordering Given fixed S , and varying candidate neighbours T_i , rank the *neighbours* T_i by $\Delta(S, T_i)$ – typically used in k-NN classification.

Pair ordering Given varying S_i , and varying T_j , rank the *pairings* $\langle S_i, T_j \rangle$ by $\Delta(S_i, T_j)$ – typically used in hierarchical clustering.

a 'distance' Δ scoring of alignments induce orderings three different kinds entities

Alignment ordering Given fixed S , and fixed T , rank the possible *alignments* $\alpha : S \mapsto T$ by $\Delta(\alpha : S \mapsto T)$

Neighbour ordering Given fixed S , and varying candidate neighbours T_i , rank the *neighbours* T_i by $\Delta(S, T_i)$ – typically used in k-NN classification.

Pair ordering Given varying S_i , and varying T_j , rank the *pairings* $\langle S_i, T_j \rangle$ by $\Delta(S_i, T_j)$ – typically used in hierarchical clustering.

same for a 'similarity' Θ scoring

For each kind of ordering can ask whether an ordering by Δ can be replicated by Θ , and vice-versa

Definition (A-,N- and P-dual)

C^Δ and C^\ominus are **A-duals** if the alignment orderings induced are the reverse of each other

C^Δ and C^\ominus are **N-duals** if the neighbour orderings induced are the reverse of each other

C^Δ and C^\ominus are **P-duals** if the pair orderings induced are the reverse of each other

The natural question is:

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa.

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

A-duality {

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A\text{-}duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \end{array} \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \end{array} \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \end{array} \right.$$

$$\mathbf{P-duality} \left\{ \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \end{array} \right.$$

$$\mathbf{P-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals}) \end{array} \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \end{array} \right.$$

$$\mathbf{P-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are P-duals}) \end{array} \right.$$

The natural question is: whether for *every* choice of C^Δ , there is a choice of C^\ominus which is a A-dual, N-dual or P-dual, and vice-versa. More precisely we have the following

Order-relating Conjectures

$$\mathbf{A-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals}) \end{array} \right.$$

$$\mathbf{N-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals}) \end{array} \right.$$

$$\mathbf{P-duality} \left\{ \begin{array}{l} (i) \quad \forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals}) \\ (ii) \quad \forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are P-duals}) \end{array} \right.$$

if these duality conjectures do not hold, then there are substantive difference, with the outcomes achievable by distances and similarities being distinct.

A-dualizing conversions

Lemma

A-dualizing conversions

Lemma

For any C^Δ , and δ s.t.

$$0 \leq \delta/2 \leq \min(C^\Delta(\cdot, \lambda), C^\Delta(\lambda, \cdot))$$

derive C^\ominus via (i)

A-dualizing conversions

Lemma

For any C^Δ , and δ s.t.

$$0 \leq \delta/2 \leq \min(C^\Delta(\cdot, \lambda), C^\Delta(\lambda, \cdot))$$

derive C^\ominus via (i)

$$(i) \begin{cases} C^\ominus(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\ominus(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\ominus(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

A-dualizing conversions

Lemma

For any C^Δ , and δ s.t.

$$0 \leq \delta/2 \leq \min(C^\Delta(\cdot, \lambda), C^\Delta(\lambda, \cdot))$$

derive C^\ominus via (i)

For any C^\ominus , and δ s.t.

$$0 \leq \delta \leq \max(C^\ominus(\cdot, \cdot))$$

(ii)

$$(i) \begin{cases} C^\ominus(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\ominus(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\ominus(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

A-dualizing conversions

Lemma

For any C^Δ , and δ s.t.

$0 \leq \delta/2 \leq \min(C^\Delta(\cdot, \lambda), C^\Delta(\lambda, \cdot))$
 derive C^\ominus via (i)

$$(i) \begin{cases} C^\ominus(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\ominus(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\ominus(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

For any C^\ominus , and δ s.t.

$0 \leq \delta \leq \max(C^\ominus(\cdot, \cdot))$ derive C^Δ via
 (ii)

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\ominus(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\ominus(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\ominus(x, y) \end{cases}$$

A-dualizing conversions

Lemma

For any C^Δ , and δ s.t.

$0 \leq \delta/2 \leq \min(C^\Delta(\cdot, \lambda), C^\Delta(\lambda, \cdot))$
 derive C^\ominus via (i)

$$(i) \begin{cases} C^\ominus(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\ominus(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\ominus(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

For any C^\ominus , and δ s.t.

$0 \leq \delta \leq \max(C^\ominus(\cdot, \cdot))$ derive C^Δ via
 (ii)

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\ominus(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\ominus(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\ominus(x, y) \end{cases}$$

then in either case, for any $\alpha : S \mapsto T$

$$\Delta(\alpha) + \Theta(\alpha) = \delta/2 \times \left(\sum_{s \in S} (1) + \sum_{t \in T} (1) \right) \quad (3)$$

A-dualizing conversions

Lemma

For any C^Δ , and δ s.t.

$0 \leq \delta/2 \leq \min(C^\Delta(\cdot, \lambda), C^\Delta(\lambda, \cdot))$
 derive C^\ominus via (i)

For any C^\ominus , and δ s.t.

$0 \leq \delta \leq \max(C^\ominus(\cdot, \cdot))$ derive C^Δ via
 (ii)

$$(i) \begin{cases} C^\ominus(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\ominus(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\ominus(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\ominus(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\ominus(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\ominus(x, y) \end{cases}$$

then in either case, for any $\alpha : S \mapsto T$

$$\Delta(\alpha) + \Theta(\alpha) = \delta/2 \times \left(\sum_{s \in S} (1) + \sum_{t \in T} (1) \right) \quad (3)$$

Theorem

A-duality (i) and (ii) hold

- ▶ so distance and similarity are interchangeable?

- ▶ so distance and similarity are interchangeable?
- ▶ the above concerns [alignment duality](#)

- ▶ so distance and similarity are interchangeable?
- ▶ the above concerns **alignment duality**
- ▶ but what about **N duality** ? (k-NN)

- ▶ so distance and similarity are interchangeable?
- ▶ the above concerns **alignment duals**
- ▶ but what about **N duals** ? (k-NN)
- ▶ and what about **P duals** ? (hierarchical clustering)

- └ Neighbour and Pair Ordering
 - └ Distance to Similarity

Outline

Distance and Similarity

Distance

Similarity

Order-equivalence Notions

Alignment Duality

Neighbour and Pair Ordering

Distance to Similarity

Similarity to Distance

Empirical Investigation

Dist to Sim: N- and P-duals

consider N-duality(i) and P-duality(i): can Neighbour- and Pair-orderings by Δ be replicated by Θ ?

Dist to Sim: N- and P-duals

consider N-duality(i) and P-duality(i): can Neighbour- and Pair-orderings by Δ be replicated by Θ ?

Δ -to- Θ conversion for A-duality was

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

Dist to Sim: N- and P-duals

consider N-duality(i) and P-duality(i): can Neighbour- and Pair-orderings by Δ be replicated by Θ ?

Δ -to- Θ conversion for A-duality was

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

and setting $\delta = 0$ gives

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) \\ C^\Theta(x, y) = -C^\Delta(x, y) \end{cases}$$

Dist to Sim: N- and P-duals

consider N-duality(i) and P-duality(i): can Neighbour- and Pair-orderings by Δ be replicated by Θ ?

Δ -to- Θ conversion for A-duality was

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

and setting $\delta = 0$ gives

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) \\ C^\Theta(x, y) = -C^\Delta(x, y) \end{cases}$$

and this implies: $\Theta(S, T) = -1 \times \Delta(S, T)$, hence

Dist to Sim: N- and P-duals

consider N-duality(i) and P-duality(i): can Neighbour- and Pair-orderings by Δ be replicated by Θ ?

Δ -to- Θ conversion for A-duality was

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

and setting $\delta = 0$ gives

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) \\ C^\Theta(x, y) = -C^\Delta(x, y) \end{cases}$$

and this implies: $\Theta(S, T) = -1 \times \Delta(S, T)$, hence

Theorem

N-duality (i) and P-duality(i) hold

- └ Neighbour and Pair Ordering
- └ Similarity to Distance

Outline

Distance and Similarity

Distance

Similarity

Order-equivalence Notions

Alignment Duality

Neighbour and Pair Ordering

Distance to Similarity

Similarity to Distance

Empirical Investigation

On order equivalences between distance and similarity measures on sequences and trees

└ Neighbour and Pair Ordering

└ Similarity to Distance

The Θ -to- Δ conversion of A-duality

was

$$(ii) \begin{cases} C^\Delta(\mathbf{x}, \lambda) = C^\Theta(\mathbf{x}, \lambda) + \delta/2 \\ C^\Delta(\lambda, \mathbf{y}) = C^\Theta(\lambda, \mathbf{y}) + \delta/2 \\ C^\Delta(\mathbf{x}, \mathbf{y}) = \delta - C^\Theta(\mathbf{x}, \mathbf{y}) \end{cases}$$

The Θ -to- Δ conversion of A-duality

was

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

with condition: $0 \leq \delta \leq \max(C^\Theta(\cdot, \cdot))$

The Θ -to- Δ conversion of A-duality

was

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

with condition: $0 \leq \delta \leq \max(C^\Theta(\cdot, \cdot))$

you can only choose $\delta = 0$ if all $C^\Theta(x, y) \leq 0$; for most natural settings of C^Θ that is not true.

The Θ -to- Δ conversion of A-duality

was

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases} \quad \text{with condition: } 0 \leq \delta \leq \max(C^\Theta(\cdot, \cdot))$$

you can only choose $\delta = 0$ if all $C^\Theta(x, y) \leq 0$; for most natural settings of C^Θ that is not true.

so can't show N-duality(ii) and P-duality(ii) this way

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

Δ - \uparrow seq. of equiv. classes

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} =$$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} = \left\{ \begin{array}{l} \Theta\text{-}\downarrow\text{ seq. of equiv. classes} \end{array} \right.$$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} = \left\{ \begin{array}{l} \Theta\text{-}\downarrow\text{ seq. of equiv. classes} \\ \dots > [\Theta]_s > \dots \end{array} \right.$$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} = \left\{ \begin{array}{l} \Theta\text{-}\downarrow\text{ seq. of equiv. classes} \\ \dots > [\Theta]_s > \dots \end{array} \right.$$

Δ must have a *min* class $[\Delta]_{min}$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} = \left\{ \begin{array}{l} \Theta\text{-}\downarrow\text{ seq. of equiv. classes} \\ \dots > [\Theta]_s > \dots \end{array} \right.$$

Δ must have a *min* class $[\Delta]_{min}$

Θ need not have *max* class $[\Theta]_{max}$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} = \left\{ \begin{array}{l} \Theta\text{-}\downarrow\text{ seq. of equiv. classes} \\ \dots > [\Theta]_s > \dots \end{array} \right.$$

Δ must have a *min* class $[\Delta]_{min}$

Θ need not have *max* class $[\Theta]_{max}$

eg. $C^\ominus(a, a) = 1, C^\ominus(a, \lambda) = 0 \Rightarrow$

$\Theta(a, a) = 1 \dots \Theta(a^n, a^n) = n$

Sim-to-Dist: P-duality(ii) fails

P-duality(ii) is stronger than *N-duality(ii)*. We can fairly easily show *P-duality(ii)* does not hold

Theorem

P-duality (ii) does not hold, that is, there are C^\ominus such that there is no C^Δ such that C^\ominus and C^Δ are P-duals.

consider the equiv. classes

$$[\Delta]_d = \{\langle S, T \rangle \mid \Delta(S, T) = d\}, \quad [\Theta]_s = \{\langle S, T \rangle \mid \Theta(S, T) = s\}$$

If C^Δ is a P-dual of C^\ominus then

$$\left. \begin{array}{l} \Delta\text{-}\uparrow\text{ seq. of equiv. classes} \\ \dots < [\Delta]_d < \dots \end{array} \right\} = \left\{ \begin{array}{l} \Theta\text{-}\downarrow\text{ seq. of equiv. classes} \\ \dots > [\Theta]_s > \dots \end{array} \right.$$

Δ must have a *min* class $[\Delta]_{min}$ eg. $C^\ominus(a, a) = 1, C^\ominus(a, \lambda) = 0 \Rightarrow$
 Θ need not have *max* class $[\Theta]_{max}$ $\Theta(a, a) = 1 \dots \Theta(a^n, a^n) = n$

in that case $\Delta\text{-}\uparrow$ sequence cannot be equal to the $\Theta\text{-}\downarrow$ sequence

- └ Neighbour and Pair Ordering
- └ Similarity to Distance

Sim-to-Dist: N-duality(ii) fails

Theorem

There is C^\ominus such that there is no C^Δ with $C^\Delta(x, \lambda) = C^\Delta(\lambda, x)$ such that C^\ominus and C^Δ are N-duals

- └ Neighbour and Pair Ordering
- └ Similarity to Distance

Sim-to-Dist: N-duality(ii) fails

Theorem

There is C^\ominus such that there is no C^Δ with $C^\Delta(x, \lambda) = C^\Delta(\lambda, x)$ such that C^\ominus and C^Δ are N-duals

Proof outline

Let $S = aa$, and set of neighbours be $\{a, aaa\}$

Sim-to-Dist: N-duality(ii) fails

Theorem

There is C^\ominus such that there is no C^Δ with $C^\Delta(x, \lambda) = C^\Delta(\lambda, x)$ such that C^\ominus and C^Δ are N-duals

Proof outline

Let $S = aa$, and set of neighbours be $\{a, aaa\}$

can define C^\ominus such that $[aaa, a] = \Theta \downarrow$ neighbour ordering

Sim-to-Dist: N-duality(ii) fails

Theorem

There is C^\ominus such that there is no C^Δ with $C^\Delta(x, \lambda) = C^\Delta(\lambda, x)$ such that C^\ominus and C^Δ are N-duals

Proof outline

Let $S = aa$, and set of neighbours be $\{a, aaa\}$

can define C^\ominus such that $[aaa, a] = \Theta\text{-}\downarrow$ neighbour ordering

cannot define C^Δ such that $[aaa, a] = \Delta\text{-}\uparrow$ neighbour ordering

further details

Let $C^\ominus(\mathbf{a}, \mathbf{a}) = x > 0$, $C^\ominus(\mathbf{a}, \lambda) = C^\ominus(\lambda, \mathbf{a}) = y > 0$

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

$$\frac{\alpha : aa \mapsto aaa \quad \Theta(\alpha)}{2 \text{ } a\text{-matches} \quad 2x - y} \quad \text{_____}$$

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ | |
|---------------------------|------------------|-------|
| 2 a-matches | $2x - y$ | _____ |
| 1 a-matches | $x - 3y$ | |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ |
|---------------------------|------------------|
| 2 <i>a</i> -matches | $2x - y$ |
| 1 <i>a</i> -matches | $x - 3y$ |
| 0 <i>a</i> -matches | $-5y$ |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ |
|---------------------------|------------------|
| 2 <i>a</i> -matches | $2x - y$ (max) |
| 1 <i>a</i> -matches | $x - 3y$ |
| 0 <i>a</i> -matches | $-5y$ |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ | $\alpha : aa \mapsto a$ | $\Theta(\alpha)$ |
|---------------------------|------------------|-------------------------|------------------|
| 2 a-matches | $2x - y$ (max) | | |
| 1 a-matches | $x - 3y$ | | |
| 0 a-matches | $-5y$ | | |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ | | $\alpha : aa \mapsto a$ | $\Theta(\alpha)$ |
|---------------------------|------------------|-------|-------------------------|------------------|
| 2 a-matches | $2x - y$ | (max) | 1 a-matches | $x - y$ |
| 1 a-matches | $x - 3y$ | | | |
| 0 a-matches | $-5y$ | | | |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ | | $\alpha : aa \mapsto a$ | $\Theta(\alpha)$ |
|---------------------------|------------------|-------|-------------------------|------------------|
| 2 a-matches | $2x - y$ | (max) | 1 a-matches | $x - y$ |
| 1 a-matches | $x - 3y$ | | 0 a-matches | $-3y$ |
| 0 a-matches | $-5y$ | | | |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ | $\alpha : aa \mapsto a$ | $\Theta(\alpha)$ |
|---------------------------|------------------|-------------------------|------------------|
| 2 a-matches | $2x - y$ (max) | 1 a-matches | $x - y$ (max) |
| 1 a-matches | $x - 3y$ | 0 a-matches | $-3y$ |
| 0 a-matches | $-5y$ | | |

further details

Let $C^\ominus(a, a) = x > 0$, $C^\ominus(a, \lambda) = C^\ominus(\lambda, a) = y > 0$

| $\alpha : aa \mapsto aaa$ | $\Theta(\alpha)$ | $\alpha : aa \mapsto a$ | $\Theta(\alpha)$ |
|---------------------------|------------------|-------------------------|------------------|
| 2 <i>a</i> -matches | $2x - y$ (max) | 1 <i>a</i> -matches | $x - y$ (max) |
| 1 <i>a</i> -matches | $x - 3y$ | 0 <i>a</i> -matches | $-3y$ |
| 0 <i>a</i> -matches | $-5y$ | | |

So $(\Theta(aa, aaa) = 2x - y) > (\Theta(aa, a) = x - y)$

So $[aaa, a] = \Theta\text{-}\downarrow$ neighbour ordering

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\begin{cases} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{cases}$

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\begin{cases} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{cases}$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ |
|---------------------------|------------------|
| 2 a-matches | $2x' + y'$ |
| 1 a-matches | $x' + 3y'$ |
| 0 a-matches | $5y'$ |

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\left\{ \begin{array}{l} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{array} \right.$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) |
|---------------------------|------------------|----------------------------------|------|
| 2 a-matches | $2x' + y'$ | $5y' + 2\epsilon$ | |
| 1 a-matches | $x' + 3y'$ | $5y' + \epsilon$ | |
| 0 a-matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | |

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\left\{ \begin{array}{l} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{array} \right.$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) $y' = x'/2 + \kappa$ |
|---------------------------|------------------|----------------------------------|---|
| 2 a-matches | $2x' + y'$ | $5y' + 2\epsilon$ | $2.5x' + \kappa$ (eq. min = $\Delta(aa, aaa)$) |
| 1 a-matches | $x' + 3y'$ | $5y' + \epsilon$ | $2.5x' + 3\kappa$ |
| 0 a-matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | $2.5x' + 5\kappa$ |

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\left\{ \begin{array}{l} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{array} \right.$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) $y' = x'/2 + \kappa$ |
|---------------------------|------------------|----------------------------------|---|
| 2 a-matches | $2x' + y'$ | $5y' + 2\epsilon$ | $2.5x' + \kappa$ (eq. min = $\Delta(aa, aaa)$) |
| 1 a-matches | $x' + 3y'$ | $5y' + \epsilon$ | $2.5x' + 3\kappa$ |
| 0 a-matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | $2.5x' + 5\kappa$ |
| | | | |
| $\alpha : aa \mapsto a$ | $\Delta(\alpha)$ | | |
| 1 a-matches | $x' + y'$ | | |
| 0 a-matches | $3y'$ | | |

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\left\{ \begin{array}{l} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{array} \right.$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) $y' = x'/2 + \kappa$ |
|---------------------------|------------------|----------------------------------|---|
| 2 a-matches | $2x' + y'$ | $5y' + 2\epsilon$ | $2.5x' + \kappa$ (eq. min = $\Delta(aa, aaa)$) |
| 1 a-matches | $x' + 3y'$ | $5y' + \epsilon$ | $2.5x' + 3\kappa$ |
| 0 a-matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | $2.5x' + 5\kappa$ |

| $\alpha : aa \mapsto a$ | $\Delta(\alpha)$ | $x' = 2y' + \epsilon$ |
|-------------------------|------------------|--------------------------------|
| 1 a-matches | $x' + y'$ | $3y' + \epsilon$ |
| 0 a-matches | $3y'$ | $3y'$ (min = $\Delta(aa, a)$) |

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\begin{cases} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{cases}$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) $y' = x'/2 + \kappa$ |
|---------------------------|------------------|----------------------------------|---|
| 2 a-matches | $2x' + y'$ | $5y' + 2\epsilon$ | $2.5x' + \kappa$ (eq. min = $\Delta(aa, aaa)$) |
| 1 a-matches | $x' + 3y'$ | $5y' + \epsilon$ | $2.5x' + 3\kappa$ |
| 0 a-matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | $2.5x' + 5\kappa$ |
| $\alpha : aa \mapsto a$ | $\Delta(\alpha)$ | $x' = 2y' + \epsilon$ | $y' = x'/2 + \kappa$ |
| 1 a-matches | $x' + y'$ | $3y' + \epsilon$ | $1.5x' + \kappa$ (eq. min = $\Delta(aa, a)$) |
| 0 a-matches | $3y'$ | $3y'$ (min = $\Delta(aa, a)$) | $1.5x' + 3\kappa$ |

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\begin{cases} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{cases}$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) $y' = x'/2 + \kappa$ |
|---------------------------|------------------|----------------------------------|---|
| 2 a-matches | $2x' + y'$ | $5y' + 2\epsilon$ | $2.5x' + \kappa$ (eq. min = $\Delta(aa, aaa)$) |
| 1 a-matches | $x' + 3y'$ | $5y' + \epsilon$ | $2.5x' + 3\kappa$ |
| 0 a-matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | $2.5x' + 5\kappa$ |
| $\alpha : aa \mapsto a$ | $\Delta(\alpha)$ | $x' = 2y' + \epsilon$ | $y' = x'/2 + \kappa$ |
| 1 a-matches | $x' + y'$ | $3y' + \epsilon$ | $1.5x' + \kappa$ (eq. min = $\Delta(aa, a)$) |
| 0 a-matches | $3y'$ | $3y'$ (min = $\Delta(aa, a)$) | $1.5x' + 3\kappa$ |

case (i): $(\Delta(aa, aaa) = 5y') > (\Delta(aa, a) = 3y')$

case (ii) $(\Delta(aa, aaa) = 2.5x' + \kappa) > (\Delta(aa, a) = 1.5x' + \kappa)$

Let $C^\Delta(a, a) = x'$, and $C^\Delta(a, \lambda) = C^\Delta(\lambda, a) = y'$.

two cases: $\begin{cases} \text{(i) in-del} < \text{swap: } 2y' < x', \text{ so } x' = 2y' + \epsilon, \text{ for some } \epsilon > 0 \\ \text{(ii) in-del} \geq \text{swap: } 2y' \geq x', \text{ so } y' = x'/2 + \kappa, \text{ for some } \kappa \geq 0 \end{cases}$

| $\alpha : aa \mapsto aaa$ | $\Delta(\alpha)$ | (i) $x' = 2y' + \epsilon$ | (ii) $y' = x'/2 + \kappa$ |
|---------------------------|------------------|----------------------------------|---|
| 2 a -matches | $2x' + y'$ | $5y' + 2\epsilon$ | $2.5x' + \kappa$ (eq. min = $\Delta(aa, aaa)$) |
| 1 a -matches | $x' + 3y'$ | $5y' + \epsilon$ | $2.5x' + 3\kappa$ |
| 0 a -matches | $5y'$ | $5y'$ (min = $\Delta(aa, aaa)$) | $2.5x' + 5\kappa$ |
| $\alpha : aa \mapsto a$ | $\Delta(\alpha)$ | $x' = 2y' + \epsilon$ | $y' = x'/2 + \kappa$ |
| 1 a -matches | $x' + y'$ | $3y' + \epsilon$ | $1.5x' + \kappa$ (eq. min = $\Delta(aa, a)$) |
| 0 a -matches | $3y'$ | $3y'$ (min = $\Delta(aa, a)$) | $1.5x' + 3\kappa$ |

case (i): $(\Delta(aa, aaa) = 5y') > (\Delta(aa, a) = 3y')$

case (ii) $(\Delta(aa, aaa) = 2.5x' + \kappa) > (\Delta(aa, a) = 1.5x' + \kappa)$

in neither case do we get $\Delta\text{-}\uparrow = [aaa, a]$

- └ Neighbour and Pair Ordering
- └ Similarity to Distance

the Order-relating Conjectures revisited

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|-------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | |

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|-------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | |

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|--------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | <i>FALSE</i> |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | <i>FALSE</i> |

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|--------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are A-duals})$ | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are N-duals})$ | <i>FALSE</i> |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\ominus (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\ominus \exists C^\Delta (C^\Delta \text{ and } C^\ominus \text{ are P-duals})$ | <i>FALSE</i> |

this means

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|--------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\Theta (C^\Delta \text{ and } C^\Theta \text{ are A-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta (C^\Delta \text{ and } C^\Theta \text{ are A-duals})$ | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\Theta (C^\Delta \text{ and } C^\Theta \text{ are N-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta (C^\Delta \text{ and } C^\Theta \text{ are N-duals})$ | <i>FALSE</i> |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\Theta (C^\Delta \text{ and } C^\Theta \text{ are P-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta (C^\Delta \text{ and } C^\Theta \text{ are P-duals})$ | <i>FALSE</i> |

this means

- ▶ any hierarchical clustering outcome achieved via Δ can be replicated via Θ , **but not vice-versa**

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|--------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\Theta$ (C^Δ and C^Θ are A-duals) | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta$ (C^Δ and C^Θ are A-duals) | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\Theta$ (C^Δ and C^Θ are N-duals) | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta$ (C^Δ and C^Θ are N-duals) | <i>FALSE</i> |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\Theta$ (C^Δ and C^Θ are P-duals) | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta$ (C^Δ and C^Θ are P-duals) | <i>FALSE</i> |

this means

- ▶ any hierarchical clustering outcome achieved via Δ can be replicated via Θ , **but not vice-versa**
- ▶ any categorisation outcome using nearest-neighbours achieved via Δ can be replicated via Θ , **but not vice-versa**

the Order-relating Conjectures revisited

| | | | | |
|------------------|---|------|--|--------------|
| A-duality | { | (i) | $\forall C^\Delta \exists C^\Theta (C^\Delta \text{ and } C^\Theta \text{ are A-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta (C^\Delta \text{ and } C^\Theta \text{ are A-duals})$ | <i>TRUE</i> |
| N-duality | { | (i) | $\forall C^\Delta \exists C^\Theta (C^\Delta \text{ and } C^\Theta \text{ are N-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta (C^\Delta \text{ and } C^\Theta \text{ are N-duals})$ | <i>FALSE</i> |
| P-duality | { | (i) | $\forall C^\Delta \exists C^\Theta (C^\Delta \text{ and } C^\Theta \text{ are P-duals})$ | <i>TRUE</i> |
| | | (ii) | $\forall C^\Theta \exists C^\Delta (C^\Delta \text{ and } C^\Theta \text{ are P-duals})$ | <i>FALSE</i> |

this means

- ▶ any hierarchical clustering outcome achieved via Δ can be replicated via Θ , **but not vice-versa**
- ▶ any categorisation outcome using nearest-neighbours achieved via Δ can be replicated via Θ , **but not vice-versa**
- ▶ in this sense 'similarity' and 'distance' comparison measures on sequences and trees **are not interchangeable**.

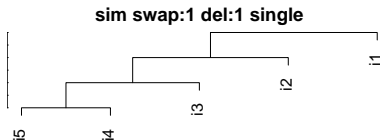
Sim to Dist: unreproducible clustering

single-link clustering of
 $\{a^5, a^4, a^3, a^2, a^1\}$
using $C^\ominus(a, a) = 1, C^\ominus(a, \lambda) = 1$

Sim to Dist: unreproducible clustering

single-link clustering of
 $\{a^5, a^4, a^3, a^2, a^1\}$
 using $C^\ominus(a, a) = 1, C^\ominus(a, \lambda) = 1$

using $C^\Delta(a, a) = 0, C^\Delta(a, \lambda) = 1$
 all on the same level because
 $\Delta(a^m, a^{m+1}) = 1$

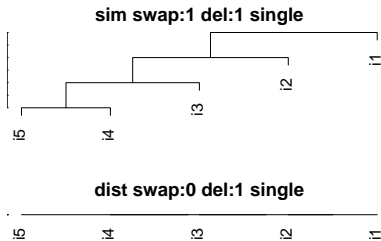


Sim to Dist: unreproducible clustering

single-link clustering of
 $\{a^5, a^4, a^3, a^2, a^1\}$
 using $C^\ominus(a, a) = 1, C^\ominus(a, \lambda) = 1$

using $C^\Delta(a, a) = 0, C^\Delta(a, \lambda) = 1$
 all on the same level because
 $\Delta(a^m, a^{m+1}) = 1$

using $C^\Delta(a, a) = 1$
 and $C^\Delta(a, \lambda)$
 or $0.5 \leq C^\Delta(a, \lambda) \leq 5.5$
 (so $2C^\Delta(a, \lambda) \geq C^\Delta(a, a)$)
 or $0.1 \leq C^\Delta(a, \lambda) \leq 0.4$
 (so $2C^\Delta(a, \lambda) < C^\Delta(a, a)$)

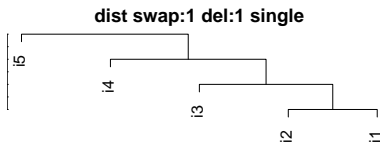
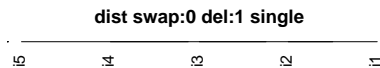
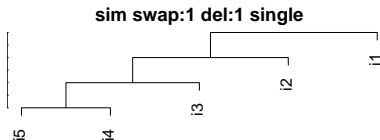


Sim to Dist: unreproducible clustering

single-link clustering of
 $\{a^5, a^4, a^3, a^2, a^1\}$
 using $C^\ominus(a, a) = 1, C^\ominus(a, \lambda) = 1$

using $C^\Delta(a, a) = 0, C^\Delta(a, \lambda) = 1$
 all on the same level because
 $\Delta(a^m, a^{m+1}) = 1$

using $C^\Delta(a, a) = 1$
 and $C^\Delta(a, \lambda)$
 or $0.5 \leq C^\Delta(a, \lambda) \leq 5.5$
 (so $2C^\Delta(a, \lambda) \geq C^\Delta(a, a)$)
 or $0.1 \leq C^\Delta(a, \lambda) \leq 0.4$
 (so $2C^\Delta(a, \lambda) < C^\Delta(a, a)$)



Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

Δ -to- Θ conversion for

A-duality was

$$(i) \quad \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

Δ -to- Θ conversion for
A-duality was

$$(i) \quad \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

eg. derived C^Θ from unit-cost
 C^Δ

| | | A-dual C^Θ for varying δ | | | | | | | |
|----------------|---|--|------|-----|------|------|------|-----|---|
| | | C^Δ | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, λ) | 1 | 0 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 | 1 | |
| (x, x) | 0 | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 | |
| (x, y) | 1 | 1 | 0.5 | 0 | -0.5 | -0.8 | -0.9 | -1 | |

Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

Δ -to- Θ conversion for
A-duality was

(i)

$$\begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

eg. derived C^Θ from unit-cost
 C^Δ

| | | A-dual C^Θ for varying δ | | | | | | | |
|----------------|---|--|------|-----|------|------|------|-----|---|
| | | C^Δ | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, λ) | 1 | 0 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 | 1 | |
| (x, x) | 0 | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 | |
| (x, y) | 1 | 1 | 0.5 | 0 | -0.5 | -0.8 | -0.9 | -1 | |

Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

Δ -to- Θ conversion for
A-duality was

(i)

$$\begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

eg. derived C^Θ from unit-cost
 C^Δ

| | | A-dual C^Θ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|------|------|----|
| | C^Δ | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, λ) | 1 | 0 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 | 1 |
| (x, x) | 0 | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, y) | 1 | 1 | 0.5 | 0 | -0.5 | -0.8 | -0.9 | -1 |

for a S , make Δ - \uparrow neighb.
ordering $N_\Delta(S)$

Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

Δ -to- Θ conversion for
A-duality was

(i)

$$\begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

eg. derived C^Θ from unit-cost
 C^Δ

| | | A-dual C^Θ for varying δ | | | | | | | |
|----------------|---|--|------|-----|------|------|------|-----|---|
| | | C^Δ | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, λ) | 1 | 0 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 | 1 | |
| (x, x) | 0 | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 | |
| (x, y) | 1 | 1 | 0.5 | 0 | -0.5 | -0.8 | -0.9 | -1 | |

for a S , make Δ - \uparrow neighb.

ordering $N_\Delta(S)$

for a S , make Θ - \downarrow neighb.

ordering $N_\Theta(S)$

Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

Δ -to- Θ conversion for
A-duality was

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

eg. derived C^Θ from unit-cost
 C^Δ

| | | A-dual C^Θ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|------|------|----|
| | C^Δ | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, λ) | 1 | 0 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 | 1 |
| (x, x) | 0 | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, y) | 1 | 1 | 0.5 | 0 | -0.5 | -0.8 | -0.9 | -1 |

for a S , make Δ - \uparrow neighb.
ordering $N_\Delta(S)$
for a S , make Θ - \downarrow neighb.
ordering $N_\Theta(S)$
 $\text{tau}(N_\Delta(S), N_\Theta(S)) =$
kendall-tau comparison of
ordering

Dist to Sim: N-duality failures

There are conversions from Dist to Sim which make A-duals: to what degree does this make an N-duals?

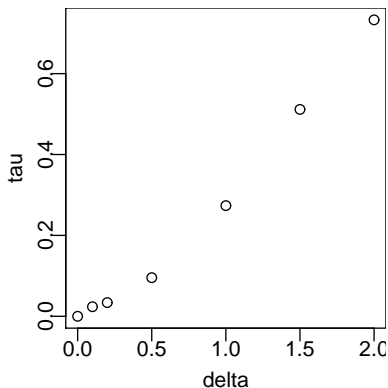
Δ -to- Θ conversion for A-duality was

$$(i) \begin{cases} C^\Theta(x, \lambda) = C^\Delta(x, \lambda) - \delta/2 \\ C^\Theta(\lambda, y) = C^\Delta(\lambda, y) - \delta/2 \\ C^\Theta(x, y) = \delta - C^\Delta(x, y) \end{cases}$$

eg. derived C^Θ from unit-cost C^Δ

| | | A-dual C^Θ for varying δ | | | | | | | |
|----------------|---|--|------|-----|------|------|------|-----|---|
| | | C^Δ | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 |
| (x, λ) | 1 | 0 | 0.25 | 0.5 | 0.75 | 0.9 | 0.95 | 1 | |
| (x, x) | 0 | 2 | 1.5 | 1 | 0.5 | 0.2 | 0.1 | 0 | |
| (x, y) | 1 | 1 | 0.5 | 0 | -0.5 | -0.8 | -0.9 | -1 | |

for a S , make Δ - \uparrow neighb.
ordering $N_\Delta(S)$
for a S , make Θ - \downarrow neighb.
ordering $N_\Theta(S)$
 $\text{tau}(N_\Delta(S), N_\Theta(S)) =$
kendall-tau comparison of
ordering



Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for

A-duality was

(ii)

$$\begin{cases} C^\Delta(\mathbf{x}, \lambda) = C^\Theta(\mathbf{x}, \lambda) + \delta/2 \\ C^\Delta(\lambda, \mathbf{y}) = C^\Theta(\lambda, \mathbf{y}) + \delta/2 \\ C^\Delta(\mathbf{x}, \mathbf{y}) = \delta - C^\Theta(\mathbf{x}, \mathbf{y}) \end{cases}$$

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for
A-duality was

$$(ii) \quad \begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

A C^Θ and several A-dual C^Δ

| | | A-dual C^Δ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|---|------|-----|
| | C^Θ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| (x, λ) | 0.5 | 1 | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 |
| (x, x) | 1 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| (x, y) | 0 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for
A-duality was

$$(ii) \begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

A C^Θ and several A-dual C^Δ

| | | A-dual C^Δ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|---|------|-----|
| | C^Θ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| (x, λ) | 0.5 | 1 | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 |
| (x, x) | 1 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| (x, y) | 0 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for
A-duality was

$$(ii) \quad \begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

A C^Θ and several A-dual C^Δ

| | | A-dual C^Δ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|---|------|-----|
| | C^Θ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| (x, λ) | 0.5 | 1 | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 |
| (x, x) | 1 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| (x, y) | 0 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |

for a S , make Θ - \downarrow neighb.
ordering $N_\Theta(S)$

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for
A-duality was

(ii)

$$\begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

A C^Θ and several A-dual C^Δ

| | | A-dual C^Δ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|---|------|-----|
| | C^Θ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| (x, λ) | 0.5 | 1 | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 |
| (x, x) | 1 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| (x, y) | 0 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |

for a S, make Θ - \downarrow neighb.

ordering $N_\Theta(S)$

for a S, make Δ - \uparrow neighb.

ordering $N_\Delta(S)$

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for
A-duality was

(ii)

$$\begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

A C^Θ and several A-dual C^Δ

| | | A-dual C^Δ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|---|------|-----|
| | C^Θ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| (x, λ) | 0.5 | 1 | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 |
| (x, x) | 1 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| (x, y) | 0 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |

for a S , make Θ - \downarrow neighb.

ordering $N_\Theta(S)$

for a S , make Δ - \uparrow neighb.

ordering $N_\Delta(S)$

$\text{tau}(N_\Theta(S), N_\Delta(S)) =$

kendall-tau comparison of
ordering

Sim-to-Dist: N-duality failures

There are conversions from Sim to Dist which make A-duals: to what degree does this make an N-duals

Θ -to- Δ conversion for A-duality was

(ii)

$$\begin{cases} C^\Delta(x, \lambda) = C^\Theta(x, \lambda) + \delta/2 \\ C^\Delta(\lambda, y) = C^\Theta(\lambda, y) + \delta/2 \\ C^\Delta(x, y) = \delta - C^\Theta(x, y) \end{cases}$$

A C^Θ and several A-dual C^Δ

| | | A-dual C^Δ for varying δ | | | | | | |
|----------------|------------|--|------|-----|------|---|------|-----|
| | C^Θ | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |
| (x, λ) | 0.5 | 1 | 1.25 | 1.5 | 1.75 | 2 | 2.25 | 2.5 |
| (x, x) | 1 | 0 | 0.5 | 1 | 1.5 | 2 | 2.5 | 3 |
| (x, y) | 0 | 1 | 1.5 | 2 | 2.5 | 3 | 3.5 | 4 |

for a S , make Θ - \downarrow neighb.
ordering $N_\Theta(S)$
for a S , make Δ - \uparrow neighb.
ordering $N_\Delta(S)$
 $\text{tau}(N_\Theta(S), N_\Delta(S)) =$
kendall-tau comparison of
ordering

