

aiNet: ein künstliches Immun Netzwerk zur Datenanalyse

Herzlich Willkommen

aiNet: An Artificial Immune Network Model for Data Analysis

Leandro Nunes de Castro &
Fernando José Von Zuben

`{lnunes,vonzuben}@dca.fee.unicamp.br`

<http://www.dca.fee.unicamp.br/~lnunes>

<ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/lnunes/DMHA.pdf>

Inhalt

- ▶ Grundlegende Ideen und Ziel(e)
- ▶ Das Immunsystem
- ▶ Grundlagen
- ▶ Lernalgorithmus
- ▶ Charakterisierung von aiNet
- ▶ Knowledge Extraction
- ▶ Beispiele
- ▶ Fazit

Grundlegende Idee

Sinnvolle Ansätze aus der informationstheoretischen Sicht ?

- ▶ Immune Network Theory
- ▶ Clonale Selektion
- ▶ affinity maturation

Grundlegendes Ziel

- ▶ Datensätze clustern, filtern und Redundanz reduzieren
- ▶ Datensätze sind durch hochdimensionale Beispiele gegeben
- ▶ ABER: es ist nicht das Ziel, das Immunsystem in irgendeiner Weise nachzubilden !

Das Immunsystem

- ▶ Immune Network Theory
- ▶ Clonale Selektion
- ▶ affinity maturation

aiNet: Grundlagen

- ▶ ShapeSpace S
 - ▶ $S = \mathbb{R}^L$
 - ▶ L Dimensionale Abbildung der Realität
 - ▶ Physio-Chemische Messungen
 - ▶ Alle immunen Ereignisse finden hier statt
- ▶ Antibody, Antigen
 - ▶ L Dimensionaler String (oder Vektor)
 - ▶ Keine Unterscheidung zwischen Oberfläche und Zelle/Molekül

aiNet: Grundlagen

- ▶ Antigene
 - ▶ Abkürzung Ag
 - ▶ Daten dargestellt in ShapeSpace S
- ▶ Antibodys
 - ▶ Abkürzung Ab
 - ▶ Netzknoten
 - ▶ Befinden sich auch im ShapeSpace S
 - ▶ Ziel: gleiche räumliche Verteilung wie Ag

aiNet: Grundlagen

- ▶ Ab-Ab / Ag-Ab Interaktionen
 - ▶ Als Konnektivitätsgraph
 - ▶ Distanzmetrik
 - ▶ Approximiert über Affinität
 - ▶ Distanz ist invers proportional zur Affinität
 - ▶ Je ähnlicher, räumlich näher, sich Ab-Ab oder Ab-Ag sind, desto höher ist die Affinität
 - ▶ Netzwerkunterdrückung
 - ▶ Netzwerkaktivierung

aiNet: Grundlagen

Definition: Das aiNet ist ein ***kantengewichteter Graph***, nicht notwendigerweise vollständig verbunden, bestehend aus einem Satz aus Knoten, genannt ***Antibodys*** und einem Satz aus Knotenpaaren, genannt ***Kanten***, mit einer Zahl, die ***Gewicht*** oder ***Verbindungsstärke*** genannt wird und jeder verbundenen Kante zugeordnet wird.

aiNet: Grundlagen

- ▶ Antibody Antigen Paar (Ab-Ag)
 - ▶ d = Affinität
 - ▶ Invers proportional zur Distanz
 - ▶ Erkennung
 - ▶ Wenn Affinität d hoch genug ist
 - ▶ Netzwerkaktivierung
 - ▶ Zellvermehrung
 - ▶ Clonen, Mutieren...

aiNet: Grundlagen

- ▶ Antibody Antibody Paar (Ab-Ab)
 - ▶ s = Ähnlichkeit (similarity)
 - ▶ Invers proportional zur Distanz
 - ▶ Erkennung
 - ▶ wenn Ähnlichkeit s groß genug ist
 - ▶ über Variable σ_s Suppression Threshold gesteuert
 - ▶ Netzwerkunterdrückung
 - ▶ Zelltod
 - ▶ Entfernen der Zelle

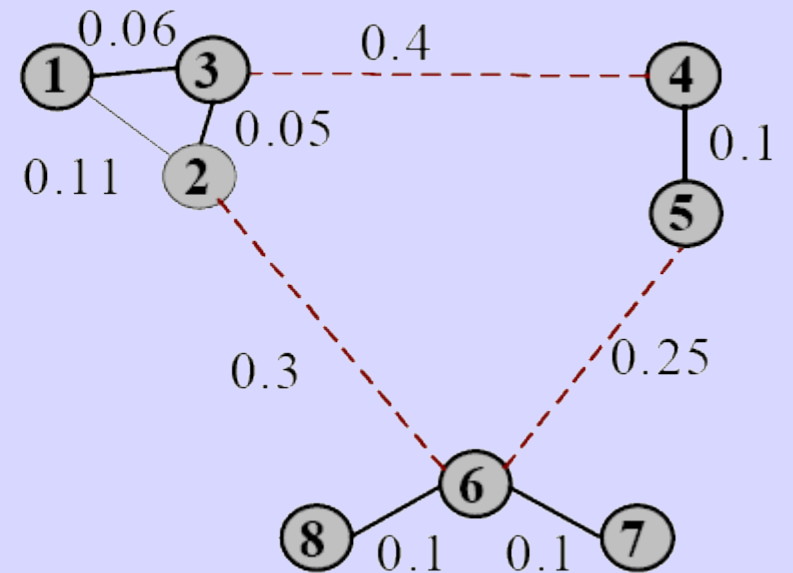
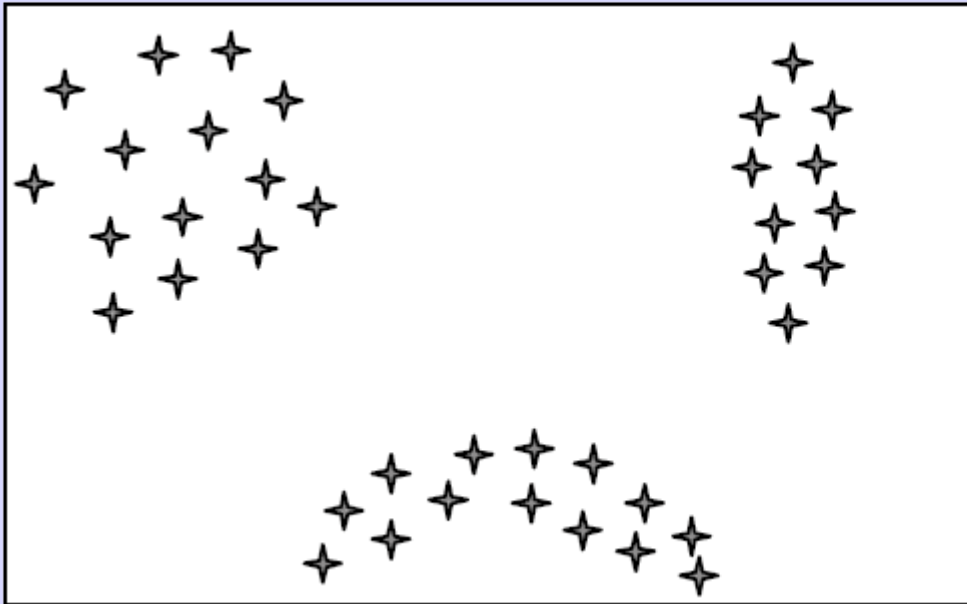
aiNet: Grundlagen

- ▶ Antibody ? Paar
 - ▶ Ab ist zu keinen anderen Ab ähnlich
 - ▶ Ab ist zu keinen Ag affin
 - ▶ Zelltod
 - ▶ Überpopulation

aiNet: Grundlagen

- ▶ AiNet Cluster
 - ▶ Bilder der Daten/ Datencluster
 - ▶ Eigenschaften werden durch σ_s Supression Threshold gesteuert
 - ▶ Ab wann erkennen sich Ab oder Ag ?
 - ▶ Wie groß wird das Netzwerk hierdurch ?
 - ▶ Struktur
 - ▶ Weniger Daten (Ag) als Graphenknoten (Ab)
 - ▶ Weniger Knoten (Ab) als Untergraphen / Cluster
 - ▶ Untergraph über Distanz
 - ▶ Datenkompression
 - ▶ muss extrahiert werden

aiNet: Grundlagen



aiNet: Lernalgorithmus

- ▶ 1. Bei jedem Durchlauf / jeder Generation
 - ▶ 1.1 Für jedes Antigen **Ag_j**, $j=1..M$, (**Ag_j** aus **Ag**):
 - ▶ 1.1.1 Berechne die Affinität **f_{ij}** für alle **Ab_i**, $i=1..N$
$$f_{ij} = 1 / D_{ij}$$
$$D_{ij} = || \text{Ab}_i - \text{Ag}_j || \text{ (Distanz)}$$
 - ▶ 1.1.2 Bilde **Ab_{n}** aus den n Antibody mit der höchsten Affinität

aiNet: Lernalgorithmus

- ▶ 1.1.3 Aus $\mathbf{Ab}_{\{n\}}$ wird \mathbf{C} erstellt, indem proportional zu der antigenen Affinität f_{ij} geklont wird:
Je höher die Affinität, desto mehr Clone gibt es jeweils.
- ▶ 1.1.4 Es wird \mathbf{C}^* generiert, indem die Clone \mathbf{C} gezielt mutiert werden (affinity maturation). Die Mutationsrate α_k ist invers proportional zu der Antigenen Affinität f_{ij} :
je höher die Affinität, desto geringer ist die Mutationsrate.
$$\mathbf{C}_k^* = \mathbf{C}_k + \alpha_k (\mathbf{Ag}_j - \mathbf{C}_k) \quad \text{mit } \alpha_k = 1/f_{ij} ; k = 1..N_C ; i=1..N$$

aiNet: Lernalgorithmus

- ▶ 1.1.5 Berechne die Affinität $\mathbf{d}_{kj} = 1/D_{kj}$ für \mathbf{Ag}_j und alle Elemente von \mathbf{C}^*
- ▶ 1.1.6 Bilde \mathbf{M}_j (clonales Gedächtnis) aus den $\zeta\%$ Antibodys mit der höchsten Affinität \mathbf{d}_{kj}
- ▶ 1.1.7 Apoptose: Eliminiere alle Clone aus \mathbf{M}_j deren Affinität $D_{kj} > \sigma_d$ ist.
- ▶ 1.1.8 Berechne die Ähnlichkeit \mathbf{s}_{ik} für die Clone in \mathbf{M}_j
 $\mathbf{s}_{ij} = ||\mathbf{M}_{ji} - \mathbf{M}_{jk}||$ für alle i und k

aiNet: Lernalgorithmus

- ▶ 1.1.9 Clonale Unterdrückung: Eliminiere alle Clone aus \mathbf{M}_j deren $s_{ik} < \sigma_s$ ist.
- ▶ 1.1.10 Füge die verbliebenen Clone in \mathbf{M}_j den Antibodys des aktuellen Durchlaufs $\mathbf{Ab}_{\{m\}}$ hinzu
- ▶ 1.2 Berechne die Ähnlichkeit aller Antibodys aus $\mathbf{Ab}_{\{m\}}$: $s_{ik} = ||\mathbf{Ab}_{\{m\}}^i - \mathbf{Ab}_{\{m\}}^k||$ für alle i und k
- ▶ 1.3 Netzwerkunterdrückung: Eliminiere alle Antibodys aus $\mathbf{Ab}_{\{m\}}$ deren $s_{ik} < \sigma_s$

aiNet: Lernalgorithmus

- ▶ 1.4 Bilde die neue Generation **Ab** aus **Ab**_{m} und **Ab**_{d} (wobei **Ab**_{d} eine Auswahl aus **Ab** sein muss)
- ▶ 2. Überprüfe das Haltekriterium

aiNet: Lernalgorithmus

- ▶ Haltekriterium
 - ▶ Variabel
 - ▶ Flexibel
 - ▶ Anzahl Iterationen
 - ▶ Anzahl an Antibodys
 - ▶ Mittlerer Fehler zwischen Ab-Ag
 - ▶ Wenn der Mittlere Fehler sich nicht mehr verändert oder schwankt

aiNet: Charakterisierung

- ▶ Verbindungsorientiert
- ▶ Konkurrierend
 - ▶ Ag Erkennung
 - ▶ Überleben der Ab
- ▶ Konstruktive
- ▶ Ab Korrespondieren zu Ag
- ▶ Konzentration und Affinität sind der Zustand
 - ▶ Änderung durch Lernalgorithmus
- ▶ Anpassungsfähig an die Aufgabe
 - ▶ Durch Initialen Antigen Satz

aiNet: Charakterisierung

► Unterschied zu Neuralen Netzen:

	aiNet	Neurales Netz
Knoten	Bilder der Daten	Verarbeitende Elemente
Verbindungen	Ähnlichkeit	Spiegeln gelerntes wieder

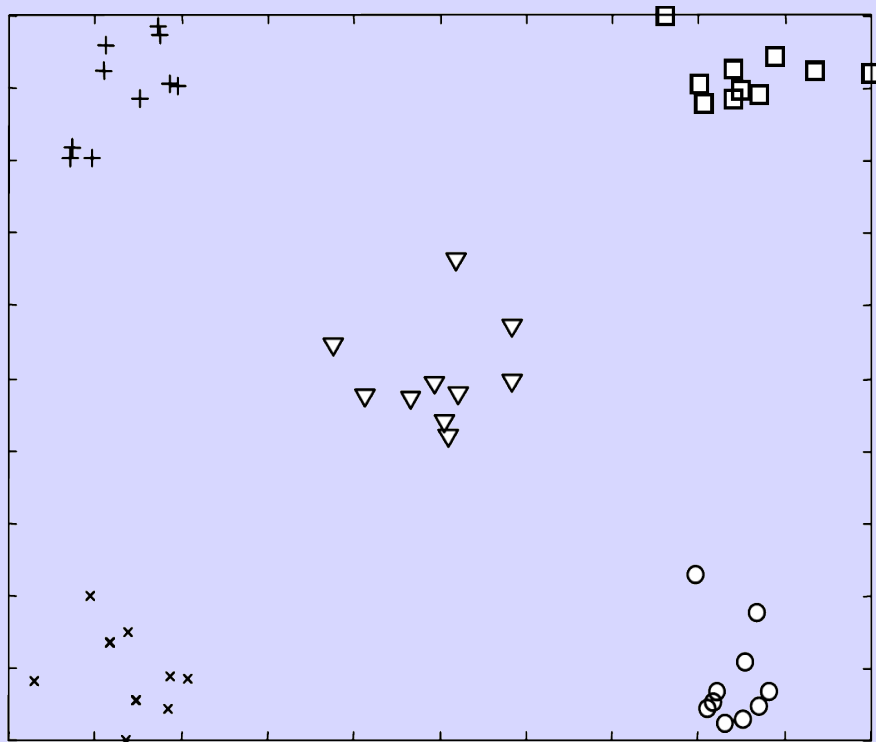
aiNet: Charakterisierung

- ▶ Als evolutionärer Algorithmus
 - ▶ Populations basiert
 - ▶ Anfangspopulation gegeben
 - ▶ Evaluations Funktion
 - ▶ Ähnlichkeitsmessungen müssen definiert werden.
 - ▶ Affinitätsmessungen müssen definiert werden.
 - ▶ Mutation
 - ▶ Clone werden durch Mutation zu Nachkommen
 - ▶ Viele Parameter
 - ▶ Maximale Affinität, Ähnlichkeit
 - ▶ Die Anzahl der zu selektierenden Ab's
 - ▶ Natürlicher Tod, Netzwerkunterdrückung

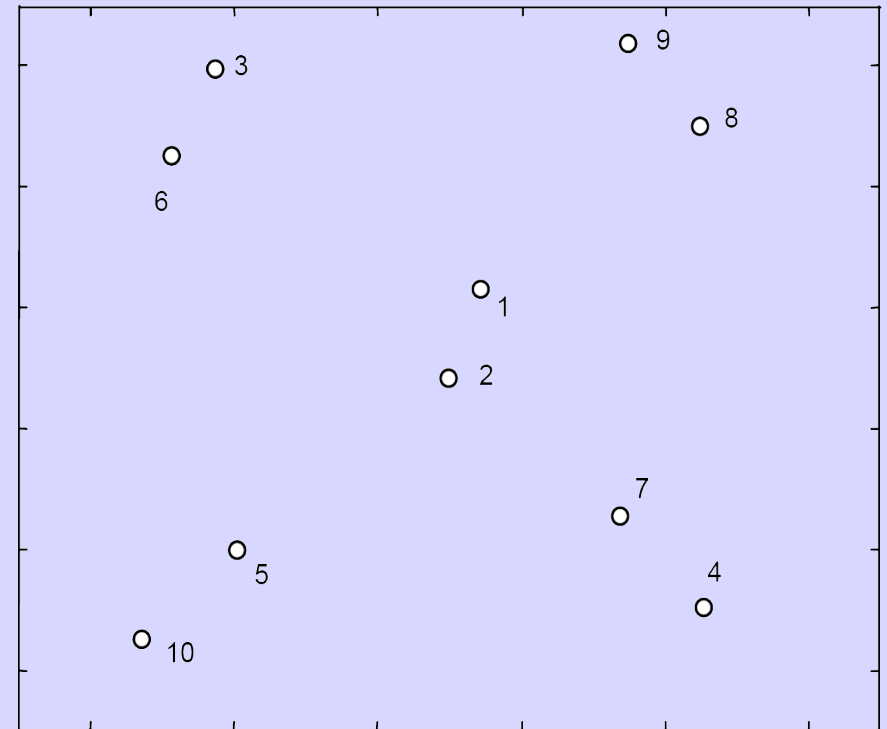
aiNet: Knowledge Extraction

- ▶ Warum Knowledge Extraction ?
 - ▶ Visualisierung des Netzwerks für Antibody/Antigen Dimensionen größer 3 nicht ohne weiteres möglich
 - ▶ Hierarchisches Clustern
 - ▶ Fertige Netzwerkstruktur

aiNet: Knowledge Extraction



(a)



(b)

aiNet illustration. (a) Learning data. (b) Resulting network antibodies.

aiNet: Knowledge Extraction

- ▶ Möglichkeit 1
 - ▶ Alle Verbindungen, die größer als ein bestimmter Wert (Threshold) sind, entfernen.
- ▶ Problem:
 - ▶ Fehlerhafte Interpretationen

aiNet: Knowledge Extraction

- ▶ Anforderungen an KE Algorithmus
 - ▶ Hohe Aussagewahrscheinlichkeit
 - ▶ Anzahl der Cluster
 - ▶ Räumliche Verteilung der Cluster
 - ▶ Antigene den Clustern zuordnen

aiNet: Knowledge Extraction

► Möglichkeit 2

► Dendrogramm

- Definition: Ein Dendrogramm ist ein gewichteter Baum bei dem alle Endknoten die selbe Distanz (Pfadlänge) bis zur Wurzel besitzen.

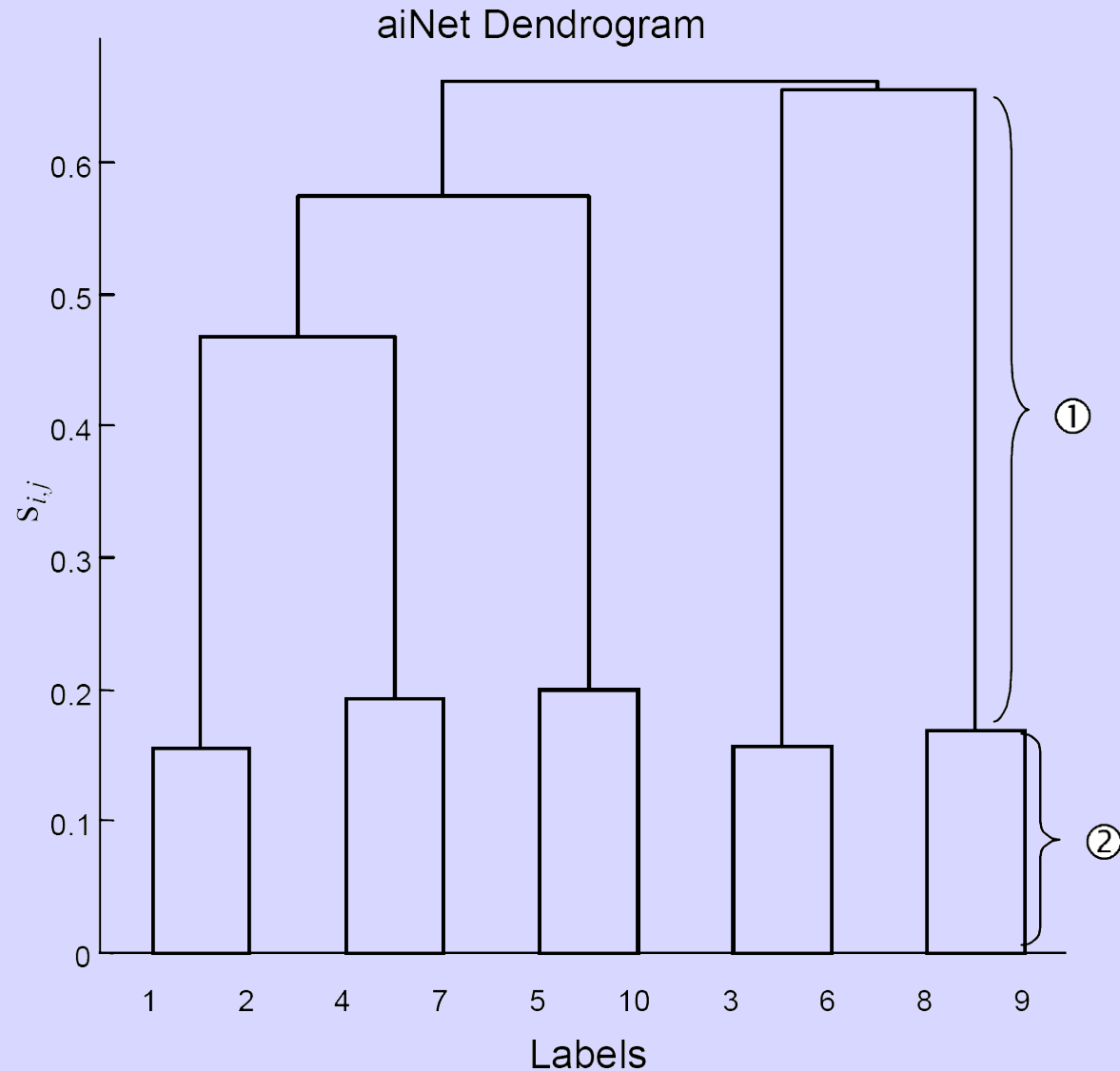
► Konstruktion aus der Ähnlichkeit / Distanz

- Siehe: Hartigan(1967) und Hubert, Arabie, und Meulman (1998)

► Clusterung über die Höhe des Dendrogramms

- Siehe: Milligan and Cooper (1985)

aiNet: Knowledge Extraction



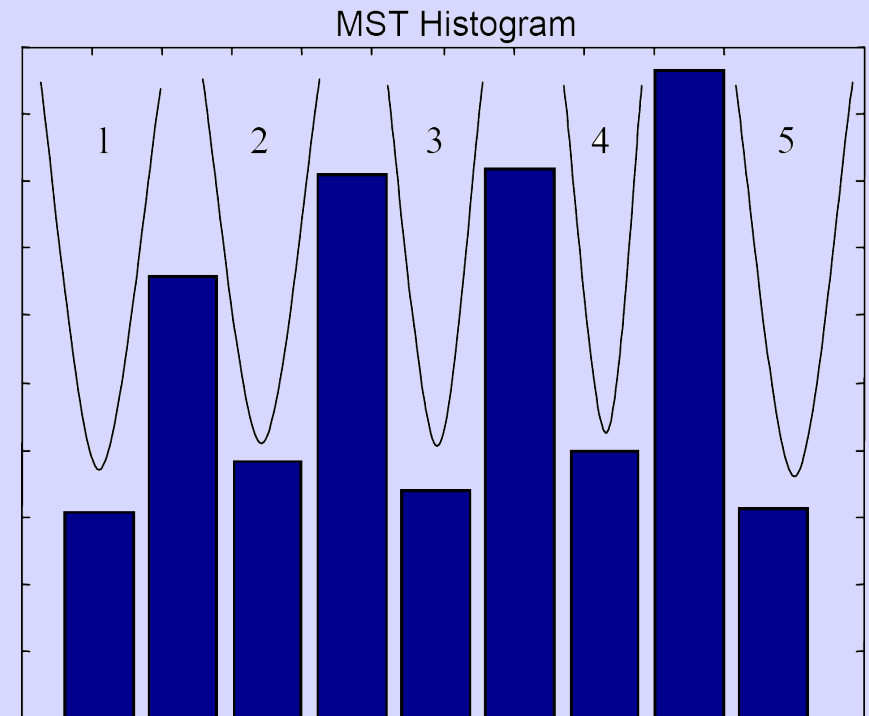
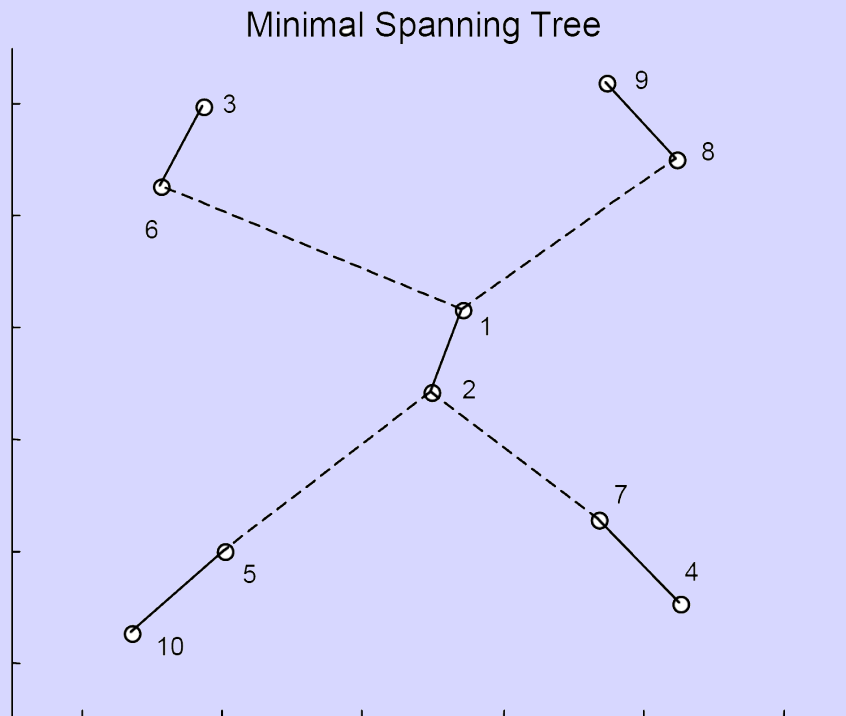
aiNet: Knowledge Extraction

- ▶ Möglichkeit 3
 - ▶ Minimaler Spannbaum
 - ▶ Ein Baum ist ein Spannbaum eines Graphen wenn er ein Untergraph ist, der alle Knoten enthält und Schleifenfrei ist.
 - ▶ Ein Minimaler Spannbaum eines Graphen ist ein Spannbaum mit minimalen Gewicht.
 - ▶ Das Gewicht eines Baumes ist definiert als die Summe der Gewichte der einzelnen Kanten.

aiNet: Knowledge Extraction

- ▶ Möglichkeit 3
 - ▶ Minimaler Spannbaum
 - ▶ Erstellung über Prim's Algorithmus (Prim, 1957)
 - ▶ Histogramm
 - ▶ Cluster = Anzahl Täler
 - ▶ Inkonsistente Kanten
 - ▶ werden entfernt
 - ▶ gdw. das Gewicht der Kante signifikant größer, als der Durchschnitt der benachbarten Kanten ist.

aiNet: Knowledge Extraction



aiNet: Knowledge Extraction

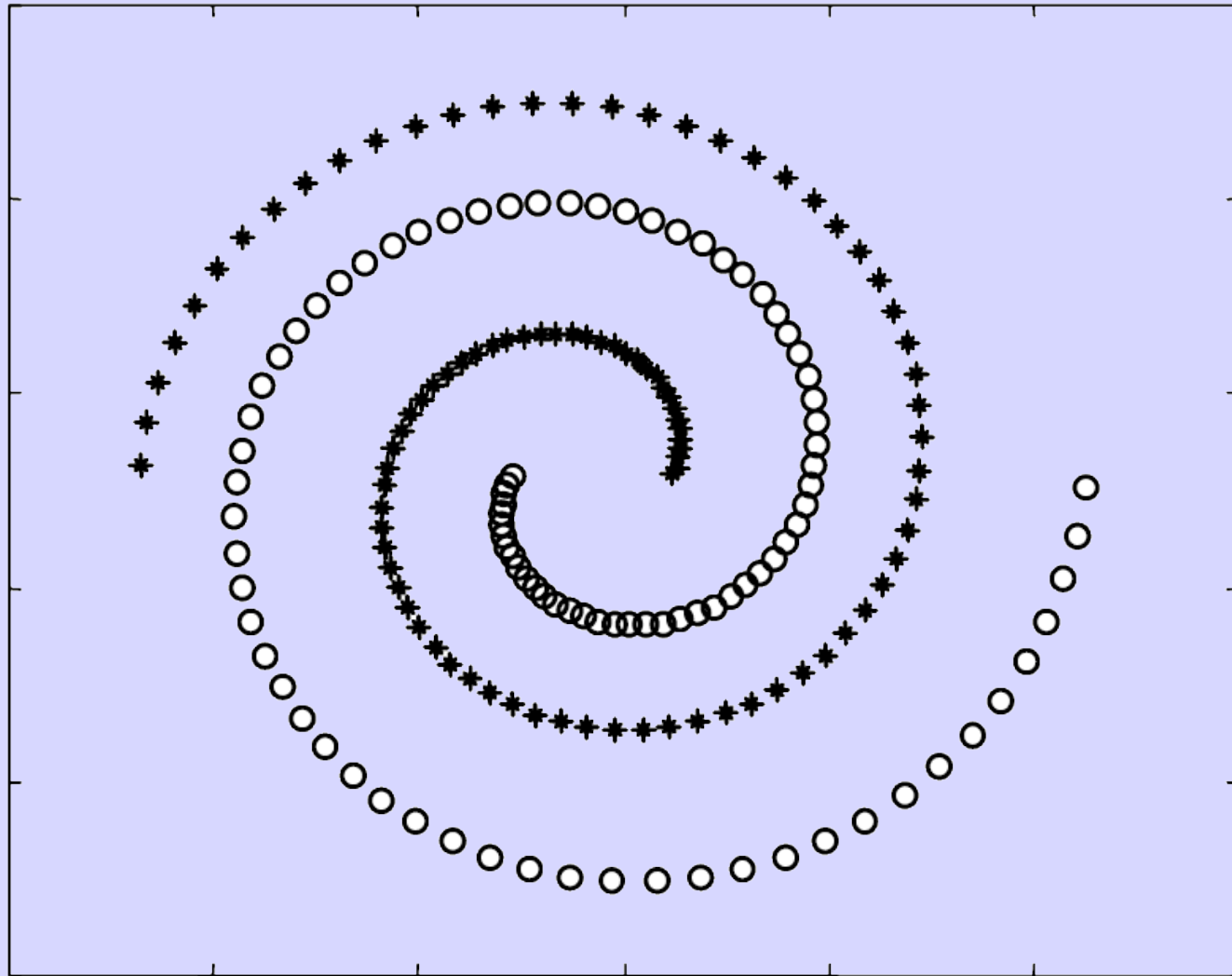
- ▶ Möglichkeit 4
 - ▶ Fuzzy Clustering
 - ▶ Antibody werden nicht fest einem Cluster zugeordnet
 - ▶ membership function
 - ▶ fuzzy k-means algorithm
 - ▶ fuzzy c-means algorithm
 - ▶ Siehe Bezdek und Pal (1992)

aiNet: Knowledge Extraction

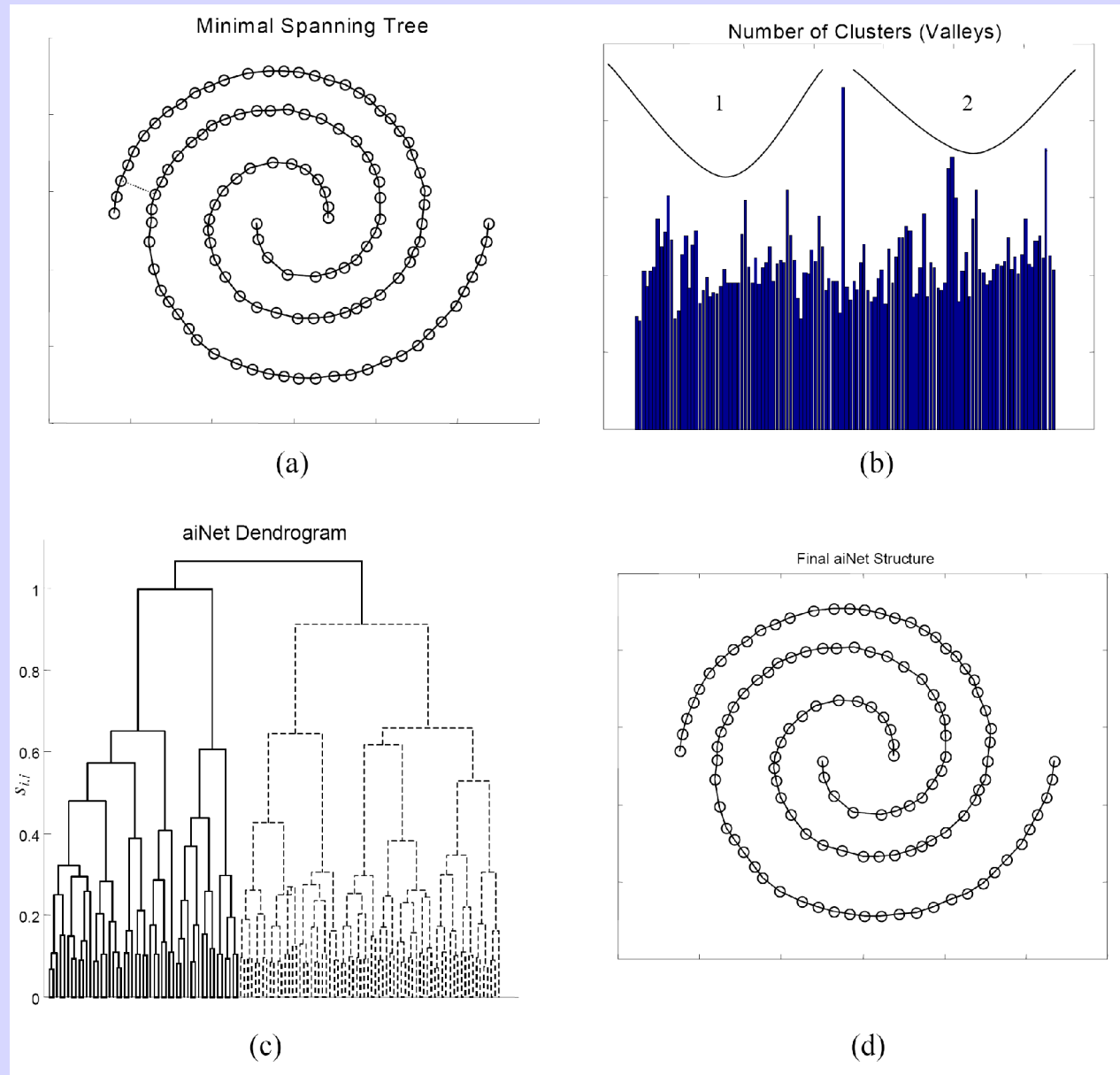
	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}
v_1	1.00	1.00	0.67	0.71	0.76	0.69	0.84	0.75	0.71	0.66
v_2	0.58	0.63	0.50	1.00	0.68	0.50	1.00	0.60	0.56	0.64
v_3	0.67	0.50	0.63	0.56	0.50	0.58	0.57	1.00	1.00	0.50
v_4	0.50	0.55	0.54	0.62	1.00	0.57	0.64	0.50	0.50	1.00
v_5	0.60	0.50	1.00	0.50	0.59	1.00	0.50	0.60	0.56	0.64

aiNet: Beispiele

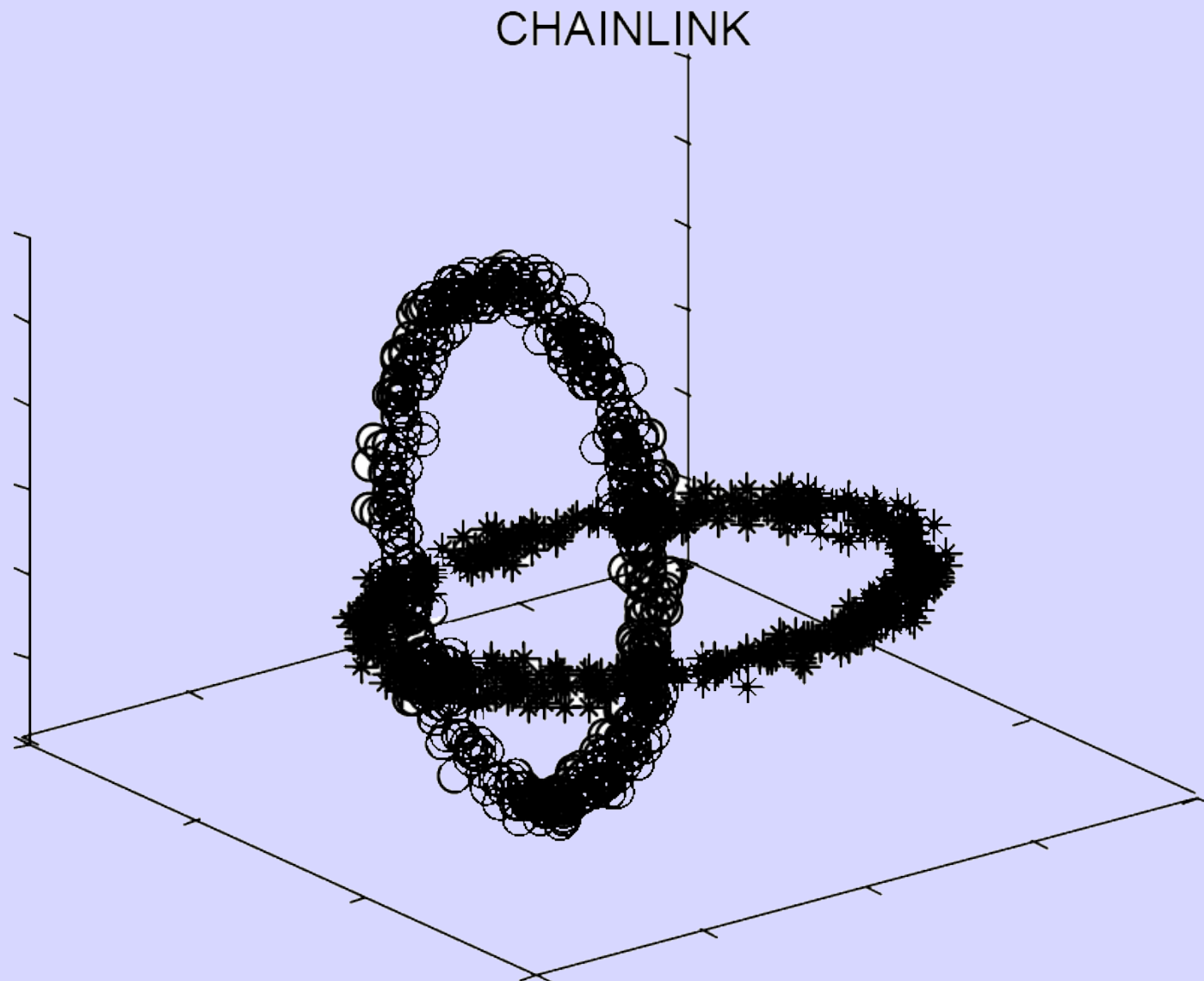
SPIR



aiNet: Beispiele

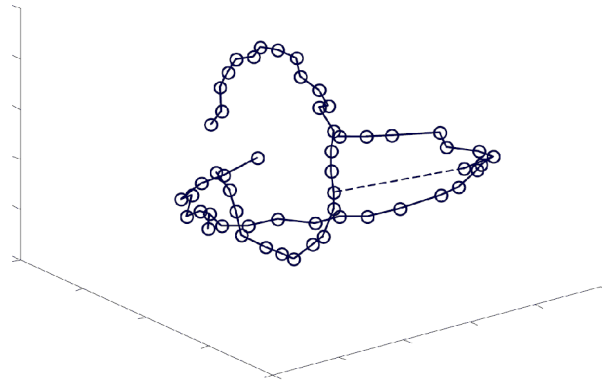


aiNet: Beispiele



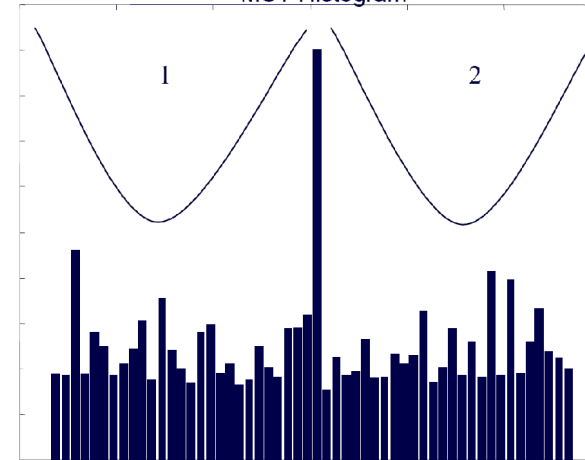
aiNet: Beispiele

Minimal Spanning Tree



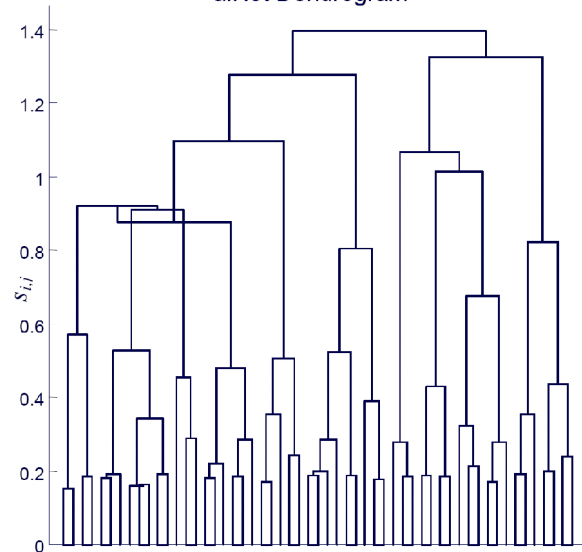
(a)

MST Histogram



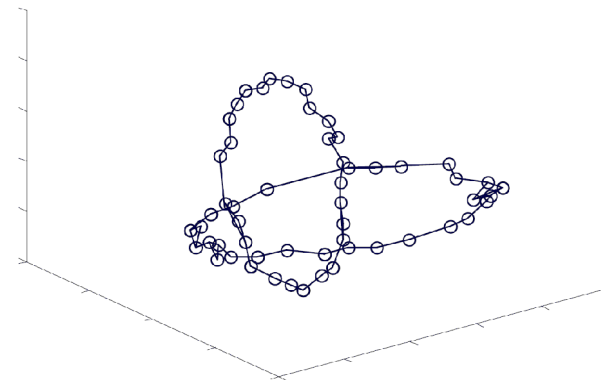
(b)

aiNet Dendrogram



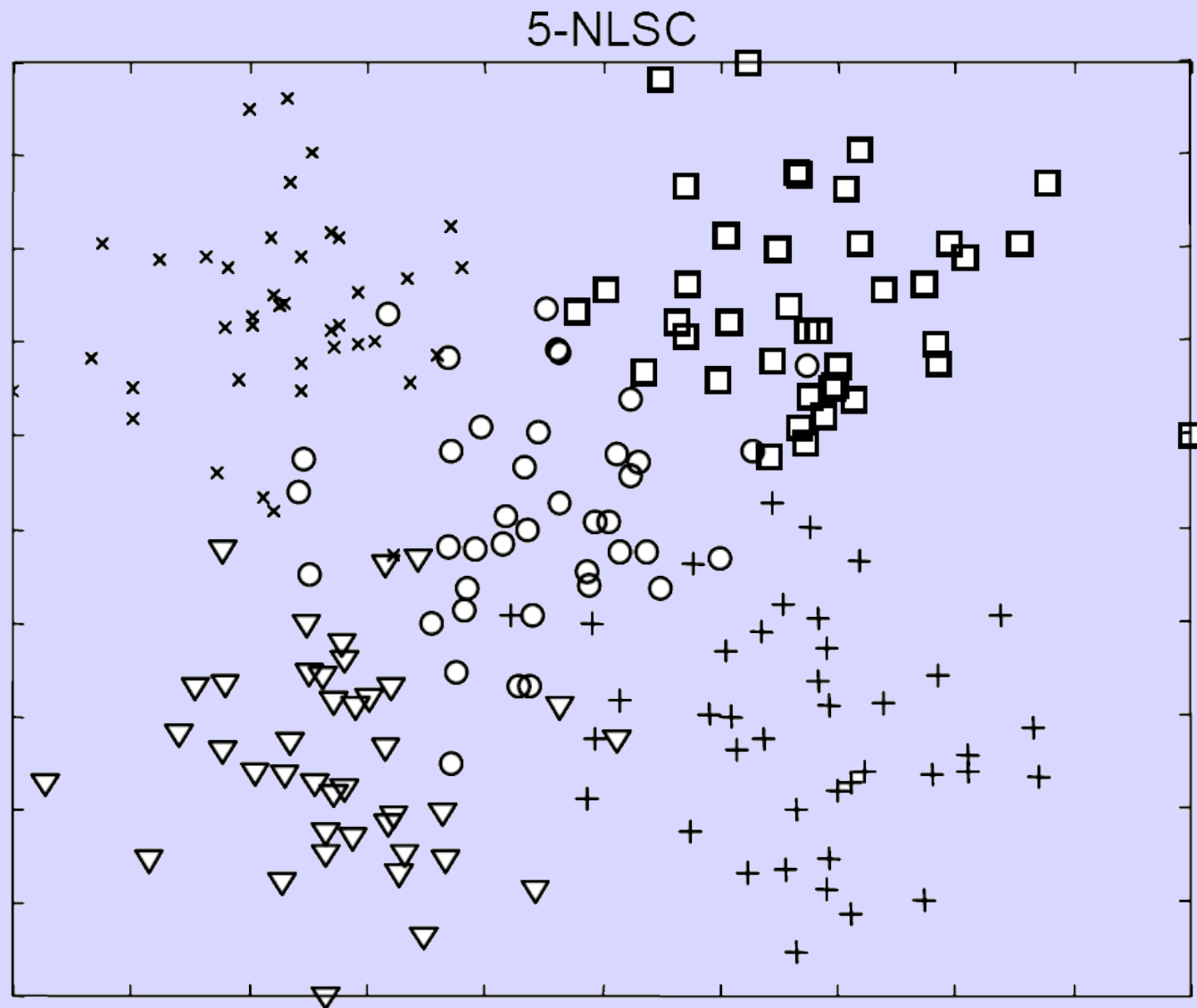
(c)

Final aiNet Structure

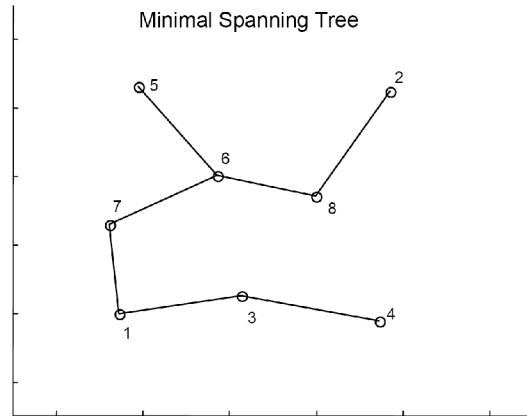


(d)

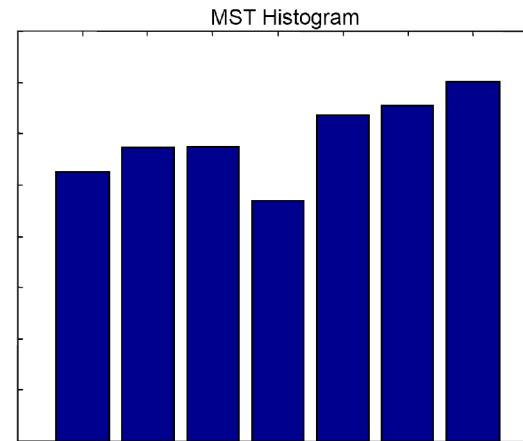
aiNet: Beispiele



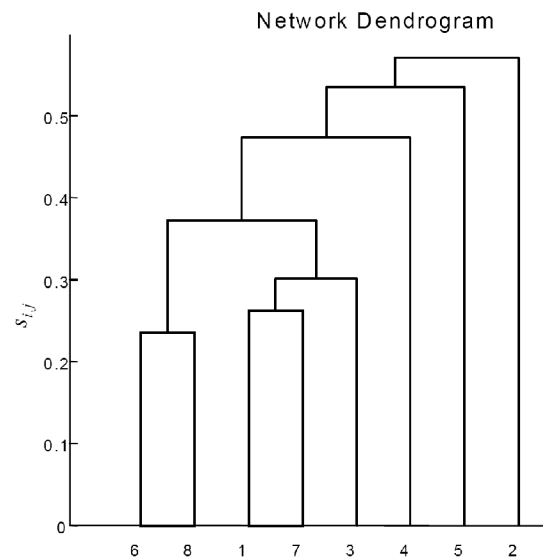
aiNet: Beispiele



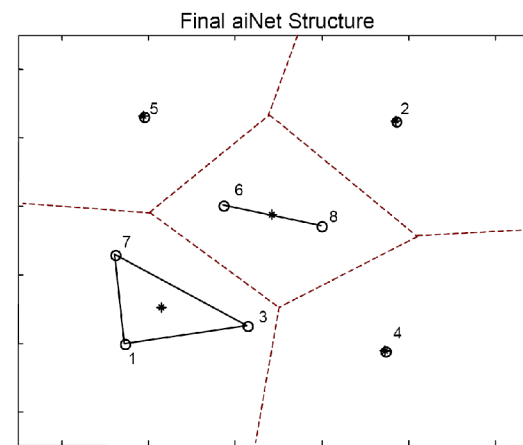
(a)



(b)



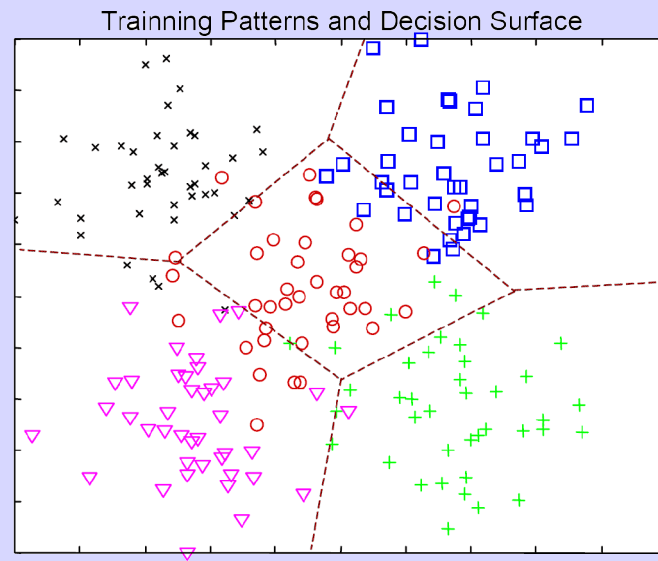
(c)



(d)

aiNet: Beispiele

	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
v_1	1.00	0.50	1.00	0.50	0.50	0.86	1.00	0.63
v_2	0.80	0.50	1.00	0.50	0.50	1.00	0.97	1.00
v_3	0.65	0.50	0.55	0.50	1.00	0.89	0.94	0.50
v_4	0.71	0.50	1.00	1.00	0.50	0.50	0.59	0.72
v_5	0.50	1.00	0.50	0.50	0.50	0.63	0.50	0.80



Fazit

- ▶ Der Algorithmus hat die Laufzeit $O(n^2)$
- ▶ aiNet ist nur als Vorstufe für weitere Clusterung zu sehen.
- ▶ Daten in ShapeSpace zu bringen könnte Problematisch sein.
- ▶ Probleme zu formulieren nicht intuitiv.

Noch Fragen???

Vielen Dank für Eure Aufmerksamkeit.