Selbstorganisierende Systeme – 4.Übung

Christian Gruber, 0625102 und Johannes Reiter, 0625101

Automatische Analyse von Metro Maps

1. How-To

Wir haben unsere Aufgabe gleich direkt in den Source-Klassen der SOM Toolbox implementieren können, wodurch sich für das Starten des Programms nichts geändert hat.
Es bleibt bei den Standardparametern: *./somtoolbox.sh SOMViewer -u /path/to/file.unit.gz -w /path/to/file.wgt.gz --dw /path/to/file.dwm.gz*

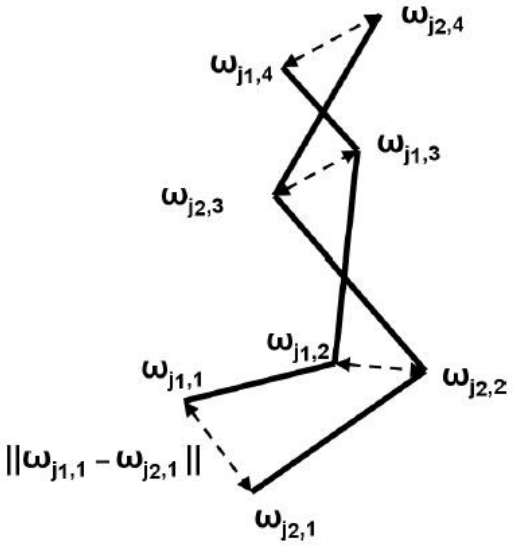
Um sich nun zu unserer automatischen Analyse einer Metro Map zu gelangen, braucht man nur mehr SOMToolBox wie zuvor angegeben starten und unter Visualisierungen die Metro Map auswählen.

2. Dokumentation

Unser Modul zur automatischen Analyse und Interpretation von Metro Map Visualisierungen umfasst das Erkennen von Korrelationen zwischen Attributen, die Ausgabe vom Verlauf von Attributen, die Bestimmung der Extrema von Attributen und ob Attribute eher als ein Rauschen bezeichnet werden können. Diese vier Erweiterungen erlauben es weitere Rückschlüsse auf die Attribute und somit auch auf die zugrundeliegenden Daten zu ziehen.

2.1. Korrelation von Attributen

Für die Berechnung der Korrelation von zwei Attributen haben wir uns an das Paper „The Metro Visualisation of Component Planes for Self-Organising Maps“ von Robert Neumayer, Rudolf Mayer, Georg Pölzlbauer und Andreas Rauber gehalten. Die folgende Grafik, die aus diesem Paper entstammt, zeigt schon gut, nach welchem Prinzip hier vorgegangen wird.



Zuerst berechnet man sich die Summe der Distanzen zwischen den einzelnen Bins, wobei man hierbei an beiden Enden einmal starten muss, um dann das Minimum verwenden zu können. Die erhaltene Distanz der beiden Metro Lines vergleichen wir nun mit zwei verschiedenen Limits (c_{weak}, c_{strong}), die wir nach unserer Evaluierung als sinnvolle Grenze für eine schwache bzw. eine starke Korrelation erachtet haben. Dadurch kann unsere automatische Metro Map Analyse auch den Unterschied von eher schwach oder eher stark korrelierenden Attribute erkennen.

$$c_{weak} = \frac{map_{height} + map_{length}}{2} * |bins| * \frac{1}{3}$$

$$c_{strong} = \frac{map_{height} + map_{length}}{2} * |bins| * \frac{1}{4}$$

$$c_{neg} = \frac{map_{height} + map_{length}}{2} * |bins| * \frac{3}{4}$$

Falls nun die zuvor berechnete Distanz zwischen den beiden Metro Lines kleiner als c_{strong} ist, bezeichnen wir die beiden Attribute als stark korrelierend und falls sie kleiner als c_{weak} ist, als schwach korrelierend. Nachdem in die Berechnung der Limits auch die Map-Größe und die Anzahl der Bins eingehen, sollten sie weitgehend unabhängig von deren Änderungen sein. Auch für die negative Korrelation haben wir uns ein Limit überlegt, das eine sinnvolle Grenze darstellen sollte.

2.2. Verlauf von Attributen

Den Verlauf von Attributen haben wir mittels den Component Planes analysiert. Dabei wurde die Map in neun Abschnitte eingeteilt, wodurch nun ein grober Verlauf eines Attributs textuell beschrieben werden kann. Die Einteilung ist oben-links, oben-mitte, oben-rechts, mitte-links, mitte-mitte, mitte-rechts, unten-links, unten-mitte und unten-rechts.
Beispielhaft kann die Ausgabe der Analyse des Verlaufs der Attribute nun so aussehen:

comp0: min auf Position (0, 7), max auf Position (0, 0)
comp0 verlaeuft von oben-mitte zu oben-links
comp1: min auf Position (0, 19), max auf Position (3, 6)
comp1 verlaeuft von oben-rechts zu oben-links
comp2: min auf Position (11, 0), max auf Position (0, 3)
comp2 verlaeuft von mitte-links zu oben-links
comp3: min auf Position (18, 19), max auf Position (0, 0)
comp3 verlaeuft von unten-rechts zu oben-links
comp4: min auf Position (2, 0), max auf Position (0, 19)
comp4 verlaeuft von oben-links zu oben-rechts
comp5: min auf Position (0, 0), max auf Position (0, 19)
comp5 verlaeuft von oben-links zu oben-rechts
comp6: min auf Position (9, 19), max auf Position (3, 6)
comp6 verlaeuft von mitte-rechts zu oben-links
comp7: min auf Position (0, 10), max auf Position (19, 16)
comp7 verlaeuft von oben-mitte zu unten-rechts
comp8: min auf Position (0, 0), max auf Position (6, 14)
comp8 verlaeuft von oben-links zu oben-rechts

Um die Zuordnung der Komponenten zu vereinfachen, haben wir deren textuelle Beschreibung in der Farbe ihrer Darstellung in der Metro Map ausgegeben.

2.3. Extrema von Attributen

Für die Extrema von den Attributen haben wir uns jeweils das Minimum bzw. das Maximum eines Attributes aus den Component Planes herausgesucht. Wie schon in der Abbildung zuvor zu erkennen, geben wir bei dieser Erweiterung die genaue Position des Extrema für jede Komponente aus.

2.4. Rauschen von Attributen

Um einen Hinweis ob ein Attribut eventuell nur Rauschen bzw. Noise ist, haben wir die Streuung genauer untersucht. Im Prinzip haben wir uns die Streuung für jedes einzelne Bin und deren Units berechnet. Der Mittelwert dieser berechneten Werte gibt uns nun einen Aufschluss darüber, ob dieses Attribut vielleicht nur Rauschen ist. Der Hintergedanke zu dieser Methode ist, dass wenn die Units *u* eines Bins bin_i sehr streuen, als Hinweis für ein Rauschen eines Attributs zu betrachten.

$$d = \frac{\sum_{i=0}^{|bins|} \sum_{\forall u \epsilon bin_i} pos(u) - pos(bin_i)}{|bins|}$$

Diesen Wert *d* kann man sich nun für jedes Attribut berechnen. Das Ergebnis vergleichen wir wieder mit einem Limit *l*, das wir aus der Evaluierung heraus bestimmt haben.

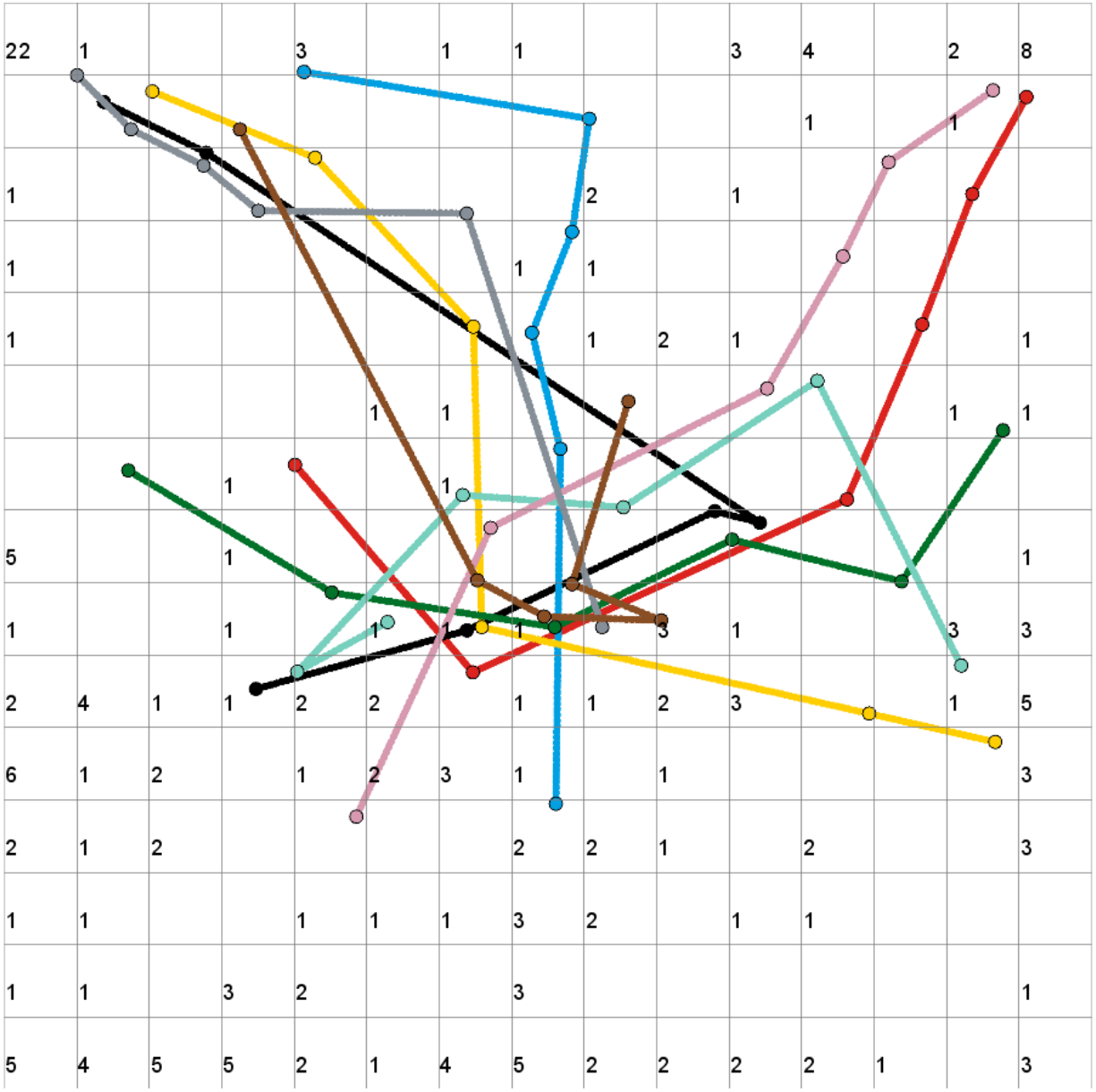
$$l = \frac{map_{height} + map_{length}}{|bins|}$$

3. Evaluierung

Um unsere Erweiterung zu evaluieren und zu testen haben mehrere uns zur Verfügung stehende Datensätze genutzt. Vor allem für die Korrelation von Attributen und der Bestimmung, ob ein Attribut nun rauscht, war ein erhöhter Evaluierungsaufwand nötig, da in beiden Fällen die Veränderung der Größe der Map bzw. der Anzahl der Bins keine Auswirkung auf das Ergebnis der Analyse haben soll.

Auf den folgenden Seiten sind nun einige Screenshots einer Metro Map Visualisierung ergänzt mit unserer automatischen Analyse.

Hier ein Screenshot einer beispielhaften, automatischen Analyse der Metro Map des Dataset *dataset_6.generic*:

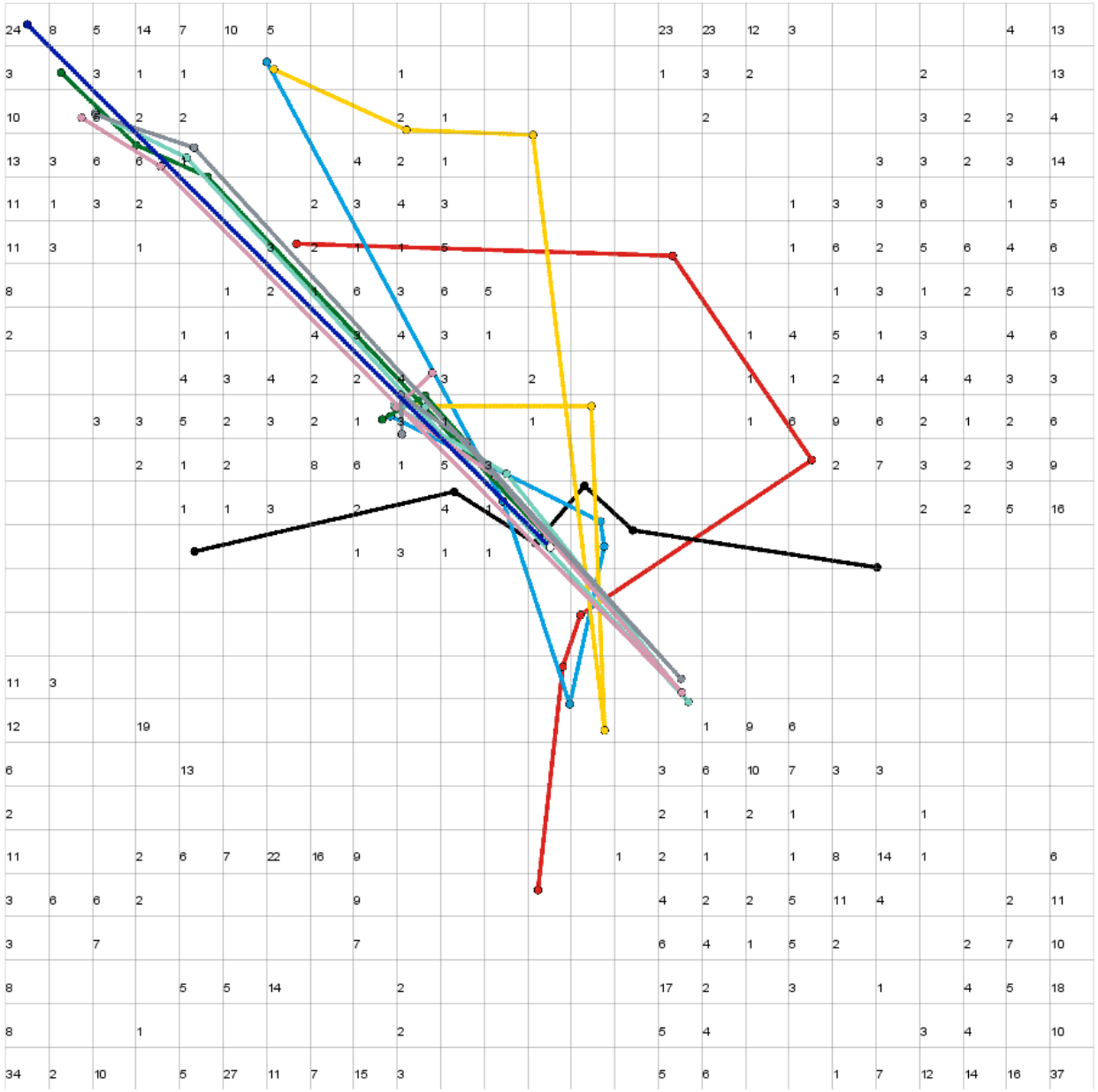


comp0 korreliert stark mit comp6
comp0 korreliert schwach mit comp4, comp5
comp1 korreliert stark mit comp8
comp1 korreliert schwach mit comp2, comp3, comp7
comp2 korreliert schwach mit comp1, comp3, comp6, comp7, comp8
comp3 korreliert schwach mit comp1, comp2, comp4, comp5, comp7
comp4 korreliert stark mit comp5, comp8
comp4 korreliert schwach mit comp0, comp3, comp6
comp5 korreliert stark mit comp4
comp5 korreliert schwach mit comp0, comp3, comp6, comp8
comp6 korreliert stark mit comp0
comp6 korreliert schwach mit comp2, comp4, comp5
comp7 korreliert schwach mit comp1, comp2, comp3
comp8 korreliert stark mit comp1, comp4
comp8 korreliert schwach mit comp2, comp3, comp5

comp0: min auf Position (3, 0), max auf Position (14, 0)
comp0 verlaeuft von oben-links zu unten-links
comp1: min auf Position (0, 9), max auf Position (0, 0)
comp1 verlaeuft von oben-mitte zu oben-links
comp2: min auf Position (0, 0), max auf Position (14, 14)
comp2 verlaeuft von oben-links zu unten-rechts
comp3: min auf Position (14, 9), max auf Position (0, 0)
comp3 verlaeuft von unten-mitte zu oben-links
comp4: min auf Position (14, 5), max auf Position (0, 0)
comp4 verlaeuft von unten-links zu oben-links
comp5: min auf Position (14, 9), max auf Position (4, 0)
comp5 verlaeuft von unten-mitte zu oben-links
comp6: min auf Position (6, 14), max auf Position (14, 0)
comp6 verlaeuft von mitte-rechts zu unten-links
comp7: min auf Position (10, 10), max auf Position (0, 0)
comp7 verlaeuft von mitte-mitte zu oben-links
comp8: min auf Position (0, 0), max auf Position (14, 0)
comp8 verlaeuft von oben-links zu unten-links

comp7 streut stark.

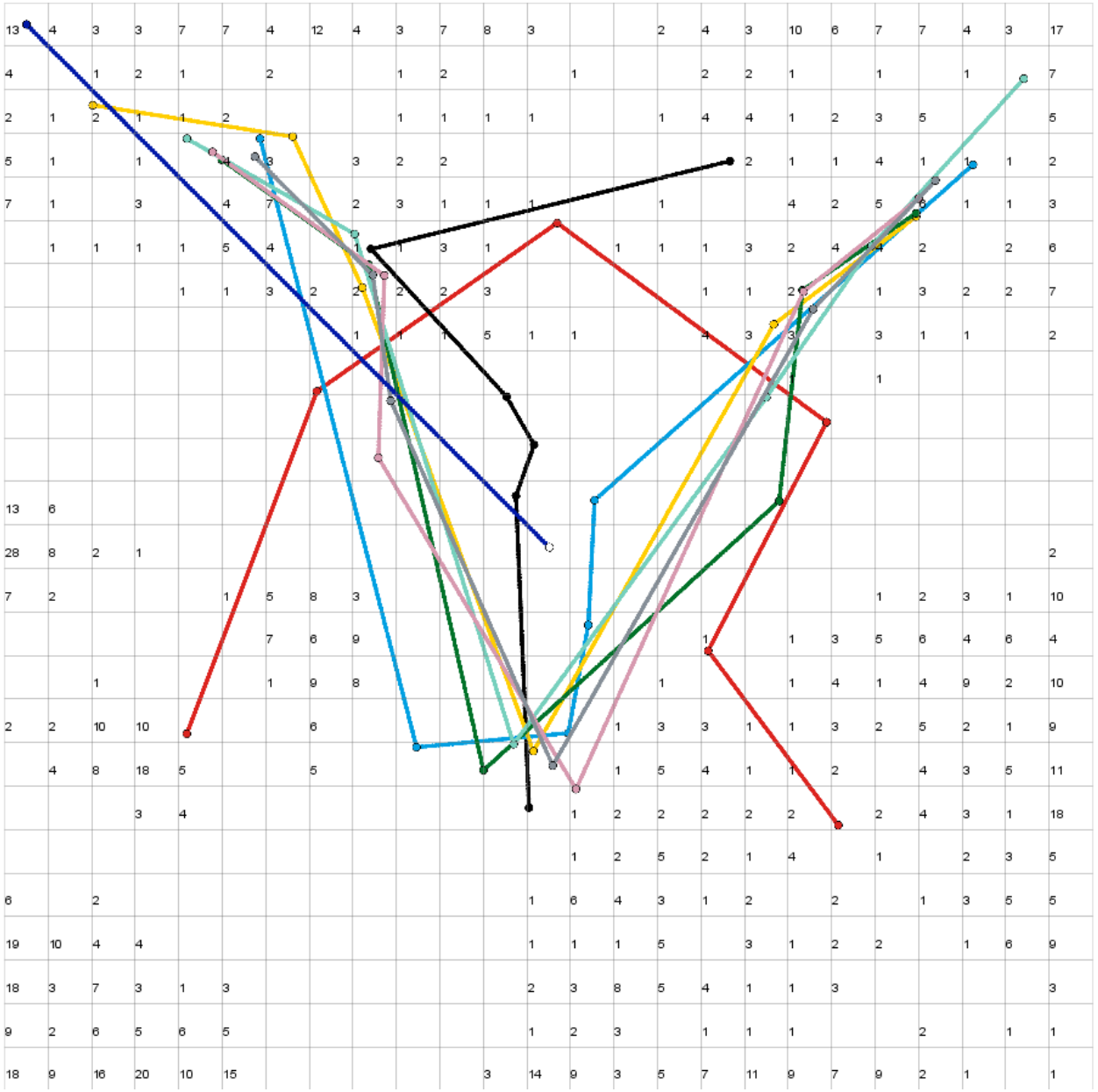
Dataset *dataset_artificial_6*:



comp0 korreliert stark mit comp2, comp8, comp9
comp0 korreliert schwach mit comp1, comp3, comp5, comp6, comp7
comp1 korreliert stark mit comp2, comp8, comp9
comp1 korreliert schwach mit comp0, comp3, comp5, comp6, comp7
comp2 korreliert stark mit comp0, comp1, comp3, comp5, comp6, co
comp2 korreliert schwach mit comp4
comp3 korreliert stark mit comp2, comp4, comp5, comp6, comp7
comp3 korreliert schwach mit comp0, comp1, comp8, comp9
comp4 korreliert stark mit comp3, comp5, comp6, comp7
comp4 korreliert schwach mit comp2, comp8, comp9
comp5 korreliert stark mit comp2, comp3, comp4, comp6, comp7, co
comp5 korreliert schwach mit comp0, comp1
comp6 korreliert stark mit comp2, comp3, comp4, comp5, comp7, co
comp6 korreliert schwach mit comp0, comp1
comp7 korreliert stark mit comp2, comp3, comp4, comp5, comp6, co
comp7 korreliert schwach mit comp0, comp1
comp8 korreliert stark mit comp0, comp1, comp2, comp5, comp6, co
comp8 korreliert schwach mit comp3, comp4
comp9 korreliert stark mit comp0, comp1, comp2, comp5, comp6, co
comp9 korreliert schwach mit comp3, comp4

comp0: min auf Position (0, 24), max auf Position (1, 0)
comp0 verlaeuft von oben-rechts zu oben-links
comp1: min auf Position (24, 23), max auf Position (0, 4)
comp1 verlaeuft von unten-rechts zu oben-links
comp2: min auf Position (8, 10), max auf Position (3, 0)
comp2 verlaeuft von oben-mitte zu oben-links
comp3: min auf Position (11, 7), max auf Position (2, 0)
comp3 verlaeuft von mitte-links zu oben-links
comp4: min auf Position (7, 9), max auf Position (0, 0)
comp4 verlaeuft von oben-mitte zu oben-links
comp5: min auf Position (9, 8), max auf Position (0, 0)
comp5 verlaeuft von mitte-links zu oben-links
comp6: min auf Position (10, 7), max auf Position (0, 3)
comp6 verlaeuft von mitte-links zu oben-links
comp7: min auf Position (10, 10), max auf Position (0, 3)
comp7 verlaeuft von mitte-mitte zu oben-links
comp8: min auf Position (10, 10), max auf Position (0, 0)
comp8 verlaeuft von mitte-mitte zu oben-links
comp9: min auf Position (12, 14), max auf Position (0, 0)
comp9 verlaeuft von mitte-mitte zu oben-links

Dataset *dataset_artificial_12*:



comp0 korreliert schwach mit comp1
comp1 korreliert schwach mit comp0, comp2, comp6, comp7, comp8
comp2 korreliert stark mit comp4, comp5, comp6, comp7, comp8, co
comp2 korreliert schwach mit comp1, comp3
comp3 korreliert stark mit comp4, comp5, comp6, comp7
comp3 korreliert schwach mit comp2, comp8, comp9
comp4 korreliert stark mit comp2, comp3, comp5, comp6, comp7
comp4 korreliert schwach mit comp8, comp9
comp5 korreliert stark mit comp2, comp3, comp4, comp6, comp7
comp5 korreliert schwach mit comp8, comp9
comp6 korreliert stark mit comp2, comp3, comp4, comp5, comp7
comp6 korreliert schwach mit comp1, comp8, comp9
comp7 korreliert stark mit comp2, comp3, comp4, comp5, comp6
comp7 korreliert schwach mit comp1, comp8, comp9
comp8 korreliert stark mit comp2, comp9
comp8 korreliert schwach mit comp1, comp3, comp4, comp5, comp6
comp9 korreliert stark mit comp2, comp8
comp9 korreliert schwach mit comp1, comp3, comp4, comp5, comp6

comp0: min auf Position (0, 24), max auf Position (24, 20)
comp0 verlaeuft von oben-rechts zu unten-rechts
comp1: min auf Position (14, 23), max auf Position (24, 0)
comp1 verlaeuft von mitte-rechts zu unten-links
comp2: min auf Position (0, 0), max auf Position (24, 0)
comp2 verlaeuft von oben-links zu unten-links
comp3: min auf Position (0, 0), max auf Position (22, 0)
comp3 verlaeuft von oben-links zu unten-links
comp4: min auf Position (5, 5), max auf Position (21, 4)
comp4 verlaeuft von oben-links zu unten-links
comp5: min auf Position (2, 0), max auf Position (24, 0)
comp5 verlaeuft von oben-links zu unten-links
comp6: min auf Position (2, 1), max auf Position (20, 3)
comp6 verlaeuft von oben-links zu unten-links
comp7: min auf Position (4, 0), max auf Position (24, 0)
comp7 verlaeuft von oben-links zu unten-links
comp8: min auf Position (17, 18), max auf Position (0, 0)
comp8 verlaeuft von unten-rechts zu oben-links
comp9: min auf Position (14, 18), max auf Position (0, 0)
comp9 verlaeuft von mitte-rechts zu oben-links