

# Glossar

## BorderFlow

ist ein Algorithmus zur Clusterung von Graphen, der ausschließlich lokale Informationen zu jedem Knoten nutzt. Zu einem  $\rightarrow$ Cluster X werden Knoten zusammengefasst, so dass die Flussstärke innerhalb von X maximal, von X nach außen minimal wird.

## Circle-View

ist eine Ansicht eines Graphen, bei der alle Knoten entlang einer Kreisbahn festen Umfangs angeordnet werden, sodass viele Kanten zwischen den Knoten dargestellt werden können.

## Cluster

(engl. Für Bündel, Gruppe, Haufen) eine logische Zusammenfassung von Objekten zu Gruppen bzw. Blöcken nach bestimmten Kriterien (bspw. die Clusterung von Graphen mit dem  $\rightarrow$ BorderFlow-Algorithmus)

## DBpedia

Ist ein Gemeinschaftsprojekt erreichbar unter <http://wiki.dbpedia.org>, das es erlaubt Informationen aus Wikipedia strukturiert abzurufen und mit anderen Webressourcen zu verknüpfen. Weiterhin bietet es Schnittstellen um komplexe Anfragen an Wikipedia zu stellen.

## Graph

In der Mathematik ein Objekt, das aus Punkten / Knoten dargestellt als Kreise und Beziehungen zwischen ihnen den sog. Kanten dargestellt als Linien besteht.

## Graphansichten (Views)

Es gibt verschiedene Ansichten eines Graphen, die für das Produkt bedeutsam sind:  $\rightarrow$ hierarchical,  $\rightarrow$ radial hierarchy,  $\rightarrow$ spring,  $\rightarrow$ simple view,  $\rightarrow$ circle.

## Graph-Browser

ist ein Programm, das dem Anwender erlaubt, sich Daten als Graph visualisieren zu lassen. Zur Funktionalität gehören das Auswählen von Ansichten und das Fokussieren auf bestimmte Aspekte.

## Hierarchical-View

Ansicht eines Graphen, bei der der ausgewählte Knoten als Wurzel oben stehend, damit verbundene Knoten einer Baumstruktur folgend in einer Linie unterhalb davon angeordnet werden. Damit eignet sich H. gut zur Darstellung hierarchisch gegliederter Netzwerke.

## HDN (Human Disease Network)

ist ein Graph, der menschliche Krankheiten und deren Beziehungen und Verbindungen untereinander visualisiert. Die Krankheiten selbst werden als Knoten dargestellt und die Verbindungen dazwischen heißen Kanten. Knoten- und Kantengewichte liefern weitere Informationen im Netzwerk.

## HuDiNe-Datensatz

Datensätze die von <http://hudine.neu.edu/> bezogen werden können. Sie liegen in Tab-separierten Textdateien vor, die pro Zeile (Kante) insgesamt zehn Spalten umfassen: ICD-9 Code 1, ICD-9 Code 2, Prävalenz 1, Prävalenz 2, Kookurenz, RR, 99%iges Konfidenzintervall links, rechts, Phi-Korrelation, t-

Test Wert zu Phi.

### ICD-9 (kurz: International Classification of Diseases)

ist eine internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. Es ist in der Medizin eines der weltweit wichtigsten anerkannten Diagnoseklassifikationssysteme. ICD-Codes folgen einem hierarchischem Aufbau. Länderspezifische Modifikationen werden durch Kürzel hinter ICD-9 angegeben: ICD-9-CM steht für **C**linical **M**odification (der USA) und zum Beispiel -GM für **G**erman **M**odification.

### Information Visualization Reference Model (IVRM)

ist ein 1999 von Ed H. Chi vorgestelltes Design-Muster. Um Informationen zu visualisieren ist eine Vielzahl von Aufgaben zu bewältigen: Datenoperation, visual mapping, Computergraphik und Interaktion. IVRM bietet eine sinnvolle Trennung dieser verschiedenen Aspekte in mehreren Schichten mit den jeweiligen Referenzen zwischen ihnen.

### Kantengewicht

In einem  $\rightarrow$ Graphen können kanten mit Gewichten versehen werden, sie drücken die Stärke des Zusammenhangs zweier Knoten dar. Im Falle eines  $\rightarrow$ HDN wird darunter ein Messwert verstanden, der über die Stärke des Zusammenhanges zweier Krankheiten ( $\rightarrow$ RR oder  $\rightarrow \Phi$ ) Auskunft gibt

### Komorbidität

Als eine Komorbidität werden in der Medizin ein oder mehrere zusätzlich zu einer Grunderkrankung (Indexerkrankung) vorliegende, diagnostisch abgrenzbare Krankheits- oder Störungsbilder bezeichnet (Doppel- oder Mehrfachdiagnose). Komorbiditäten können, müssen aber nicht – im Sinne einer Folgeerkrankung – ursächlich mit der Grunderkrankung zusammenhängen.

### Konfidenzintervall

gibt in der Statistik Auskunft über die Präzision der Lageschätzung eines Wertes. Mittels einer festgelegten Wahrscheinlichkeit (Konfidenzniveau) wird ein Intervall (linkes und rechtes K.) angegeben, in dem der wahre Wert mit jener Wahrscheinlichkeit liegt.

### Kontextmenü

Ein über einen Klick der rechten Maustaste kontextsensitives aufrufbares Menü, das dem Nutzer die Möglichkeit bietet mit dem Objekt verknüpfte Funktionalitäten aufzurufen.

### Kookkurrenz

entstammt der Linguistik und bezeichnet das gemeinsame Auftreten zweier Einheiten in einer übergeordneten Einheit. Zwei Einheiten gelten als voneinander abhängig, wenn sie häufig gemeinsam auftreten.

### PDN (Phenotypic-Disease-Network)

ist ein spezielles  $\rightarrow$ HDN, welches den Fokus auf die Phänotypen richtet, also den Erscheinungsformen von Krankheiten und Störungen, die gemeinsam bei einem Individuum auftreten und damit  $\rightarrow$ kookkurrent sind. Als Knotengewichte werden die  $\rightarrow$ Prävalenzen der Krankheiten herangezogen und die Kantengewichte werden durch das  $\rightarrow$ Komorbiditätsmaß zweier in Verbindung stehender Krankheiten bestimmt.

### Pearson-Korrelation ( $\Phi$ )

Maß für die Stärke der Komorbidität zweier Krankheiten. Wertebereich: [-1,1], Mittelpunkt (d.h kein besonderer Zusammenhang) bei 0. Misst zuverlässig den Zusammenhang zwischen ähnlich  $\rightarrow$ prävalenten Krankheiten, unterschätzt aber systematisch den Wert zwischen stark unterschiedlich

prävalenten Krankheiten. →*Signifikanzprüfung* anhand von t-Tests.

### Prävalenz

oder Krankheitshäufigkeit ist eine Kennzahl der Gesundheits- und Krankheitslehre (Epidemiologie) und sagt aus, wie viele Menschen einer bestimmten Gruppe (Population) definierter Größe an einer bestimmten Krankheit erkrankt sind. Dabei wird deren Anzahl durch die Größe der untersuchten Grundgesamtheit geteilt.

### Prefuse

ist ein auf Java basierendes Toolkit, das die Erstellung von interaktiven Applikationen zur Informationsvisualisierung ermöglicht. Es bietet eine Vielzahl von Funktionen zur Datenmodellierung, Visualisierung und Interaktion. Weiterhin bietet P. neben vielerlei Techniken zur Codierung des Layout und der Visualisierung optimierte Datenstrukturen für Tabellen, Graphen und Bäume, eine Datenbankanbindung, dynamische Anfragen und integrierte Suchfunktionen.

### Radial hierarchy

ist eine Ansicht eines Graphen, bei dem ausgewählte Knoten als Mittelpunkt und die mit Kanten verbundenen anderen Knoten entlang eines Kreises um diesen angeordnet werden. Es eignet sich demnach zur Darstellung von →*Komorbiditätsnetzwerken*, bei denen eine Krankheit zu vielen anderen in Verbindung auftritt.

### Relative Risk (RR)

ist ein einfaches Maß um die →*Komorbidität* zweier Krankheiten zu messen. Wertebereich (0,60], Mittelpunkt (also kein besonderer Zusammenhang) bei 1. Neigt zur Unterschätzung von Zusammenhängen zwischen stark →*prävalenten* Krankheiten, hingegen wird die Komorbidität zwischen schwach prävalenten Krankheiten systematisch überschätzt. →*Signifikanzprüfung* erfolgt mittels Konfidenzintervallen.

### Simple View

ist die grundlegende Darstellung eines Graphen, bei der die Bedeutung von Knoten (dargestellt als Kreise) über deren Umfang, die Kantengewichte bzw. Flussstärken über die Dicke der gezeichneten Kanten modelliert wird. Die Anordnung der Knoten unterliegt keiner bestimmten Ordnung bzw. Kräftemodell.

### Signifikanz

Lateinisch für bedeutsam, wesentlich. Werte heißen statistisch signifikant, wenn deren zufälliges Zustandekommen unwahrscheinlich ist. Werte mit einer Irrtumswahrscheinlichkeit  $\leq 0,3\%$  nennt man sehr signifikant, hoch signifikant für Irrtumswahrscheinlichkeiten  $\leq 0,1\%$ .

### Spring

ist eine Ansicht eines Graphen, wobei die Größen der Kantengewichte mittels einer bewegten Ausbalancierung des Graphen visualisiert werden. So bewirken große Kantengewichte eine stärkere Zugkraft bei Bewegung eines Knotens auf den anliegenden Restgraphen aus, als geringere Kantengewichte.

### T-Test

Gruppe von statistischen Tests zur →*Signifikanzmessung*. Der ermittelte t-Wert muss mit t-Testtabellen in Abhängigkeit von der Stichprobengröße auf Signifikanz geprüft werden.