





Gephi – Der Guide zum [#twitter101dh-Workshop](#)

von Sophie Schneider

 [ORCID](#) ·  [Twitter](#) ·  [Github](#) ·  [Blog](#)

Vorbereitung

“Gephi is the leading visualization and exploration software for all kinds of graphs and networks.”
(<https://gephi.org/>)

Bevor wir mit unserer Netzwerkanalyse beginnen können, müssen wir zunächst Gephi installieren. Neben der Installation werden außerdem Daten benötigt, auf deren Grundlage die Netzwerkanalyse durchgeführt werden soll. Hier macht es Sinn, noch einmal kurz innezuhalten und über die eigene Forschungsfrage sowie die verwendeten Daten zu reflektieren:

- Zielt meine Forschungsfrage darauf ab, Relationen aufzudecken und zu untersuchen?
- Enthalten meine Daten alle hierfür notwendigen Informationen (Entitäten = Knoten und deren Beziehungen = Kanten) oder müssen ggf. noch zusätzliche Daten generiert werden?
- Sind Netzwerke wirklich die beste Option, meine Daten zu visualisieren? Oder sind andere Diagramme wie beispielsweise Scatter Plots oder Box Plots besser geeignet, die wesentlichen Erkenntnisse meiner Arbeit (visuell) zusammenzufassen?

Erste Schritte

Projekt öffnen

Öffnet man Gephi, dann öffnet sich als erstes auch das Fenster aus **Abb. 1**. Hier kann man unter anderem **1** kürzlich geöffnete Dateien erneut öffnen und daran weiterarbeiten. Der aktuelle Arbeitsstand wird in Gephi in sogenannten Projekten (.gephi-Dateien) gespeichert. Wenn wir ganz neu anfangen, empfiehlt es sich, entweder ein komplett **2** neues Projekt zu erstellen oder auf eines der hier verfügbaren **3** Beispiele zurückzugreifen. Da wir Projekte auch danach noch erstellen/öffnen können, lässt sich **4** das automatische Öffnen dieses Fensters beim Start von Gephi bei Bedarf abwählen.

Gephi GUI

Sehen wir uns zunächst einmal den Aufbau des User Interfaces von Gephi etwas genauer an (**Abb. 2**). Ganz oben finden wir eine **1** Menüleiste, die die wichtigsten Funktionen (noch einmal) enthält. Die drei Bereiche **2** Übersicht, Datenlabor und Vorschau werden für uns im Folgenden sehr relevant sein und sie strukturieren auch den Prozess der Datenverarbeitung in Gephi. Darunter befindet sich unser aktueller **3** Arbeitsbereich, d.h. der Bereich, in dem wir jetzt gerade Daten bearbeiten. Innerhalb eines Projektes können wir also parallel in verschiedenen Arbeitsbereichen arbeiten. Der letzte große Bereich **4** verändert sich - je nachdem, ob wir uns gerade im Übersichts-, Datenlabor- oder Vorschaumodus befinden.



Abb. 1: Willkommen zu Gephi

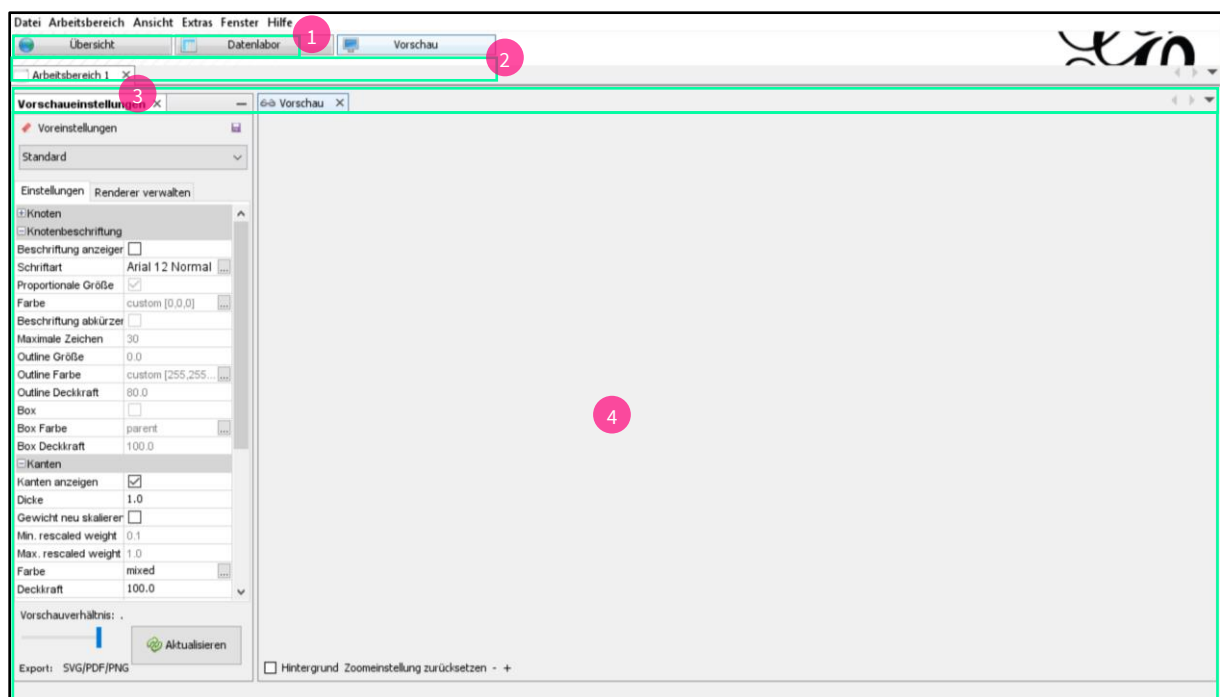


Abb. 2: Übersicht GUI

Menüleiste

Die Menüleiste besteht aus den sechs Reitern: **Datei**, **Arbeitsbereich**, **Ansicht**, **Extras**, **Fenster** und **Hilfe** (Abb. 3). Schauen wir uns diese einmal genauer an.

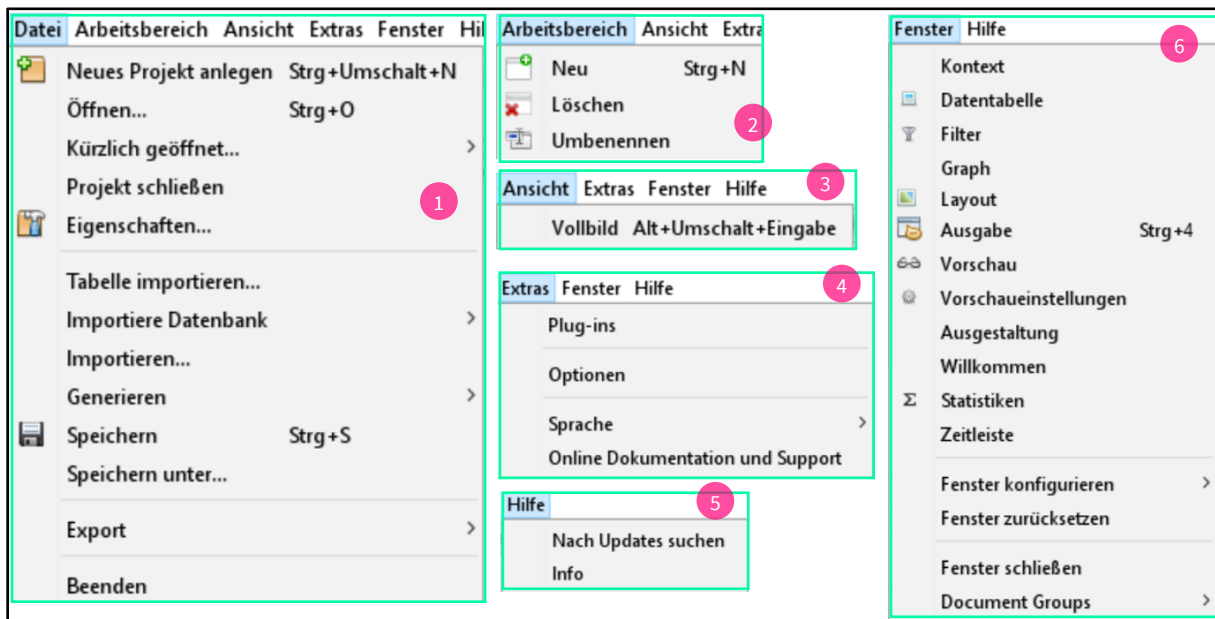


Abb. 3: Menüleiste - Reiter ausgeklappt

Unter **1** Datei können Projekte organisiert, Daten importiert und exportiert sowie das Programm beendet werden. Der Reiter **2** Arbeitsbereich ermöglicht es, einen neuen Arbeitsbereich hinzuzufügen oder den aktuellen zu löschen oder umzubenennen. Unter **3** Ansicht kann man in den Vollbildmodus umschalten. Bei den **4** Extras sind die verschiedenen Plugins aufzufinden, die bei spezifischen Fragestellungen hilfreich sein können oder zusätzliche Funktionen anbieten. Außerdem können hier unter Optionen einige Einstellungen für Gephi und unter Sprache die Sprache geändert werden. Die **5** Hilfe bietet die Möglichkeit, nach Updates zu suchen sowie allgemeine Informationen zur aktuell installierten Version. Unter **6** Fenster können z.B. geschlossene Fenster neu geöffnet werden und Fenster konfiguriert oder geschlossen werden.

Datenlabor

Vorbereitung der Daten

Um unsere Daten in Gephi zu laden, müssen wir sie zunächst in ein für das Programm lesbares Format bringen. Neben einer automatisierten Transformation von Tweets in ein für Gephi geeignetes Format wie .gexf (Graph Exchange XML Format) (siehe z.B. Creating a retweet network for Gephi from a local file with Python) besteht auch die Möglichkeit einer „manuellen“ Transformation der Daten mit Excel (Data Preparation for Gephi) in Knoten und Kanten. Alternativ ist es ebenfalls möglich, eine CSV-Datei in Form einer Adjazenzliste hochzuladen (dies ist aus meiner Sicht weniger umständlich). Im Folgenden wird der Datenimport exemplarisch anhand einer solchen Adjazenzliste (siehe Beispieldatei) gezeigt.

Import der Daten

In **Abb. 4** ist zunächst der Bereich „Datenlabor“ zu sehen. Es wird zwischen **1** Knoten und Kanten unterschieden. Kleine Netzwerke können wir auch direkt in Gephi erstellen, indem wir einzelne **2**

Knoten und Kanten hinzufügen. Größere Netzwerke/Dateien können wir in Form von **3** Tabellen importieren und exportieren. **4** und **5** enthalten weitere Möglichkeiten zur Bearbeitung der Daten.

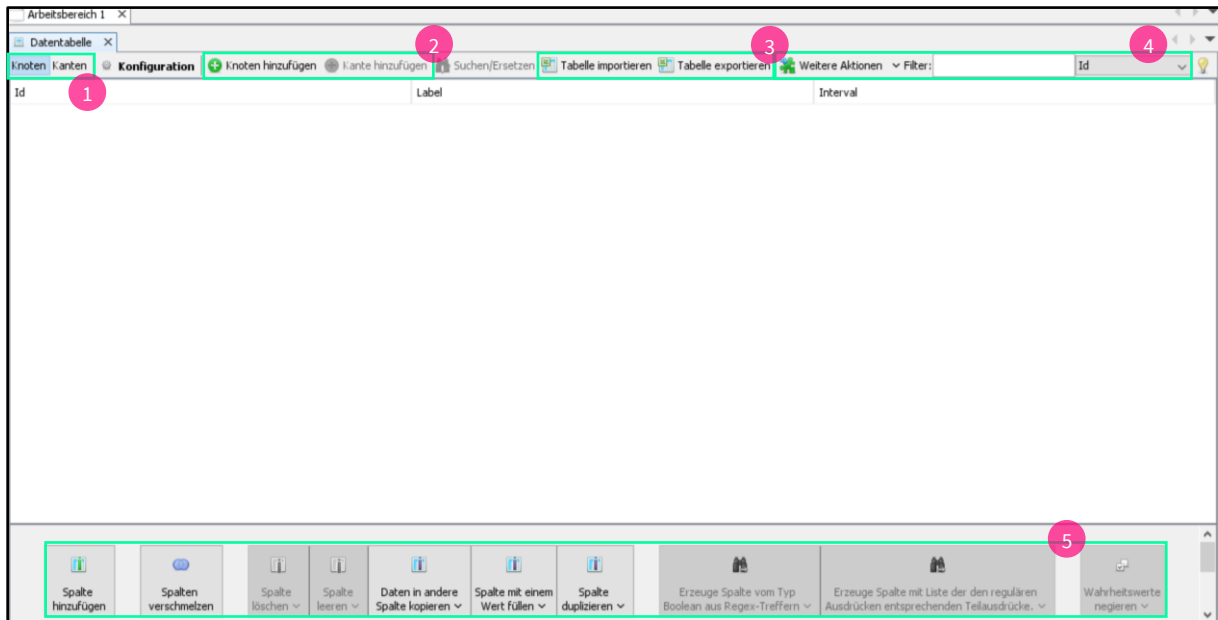


Abb. 4: Datenlabor

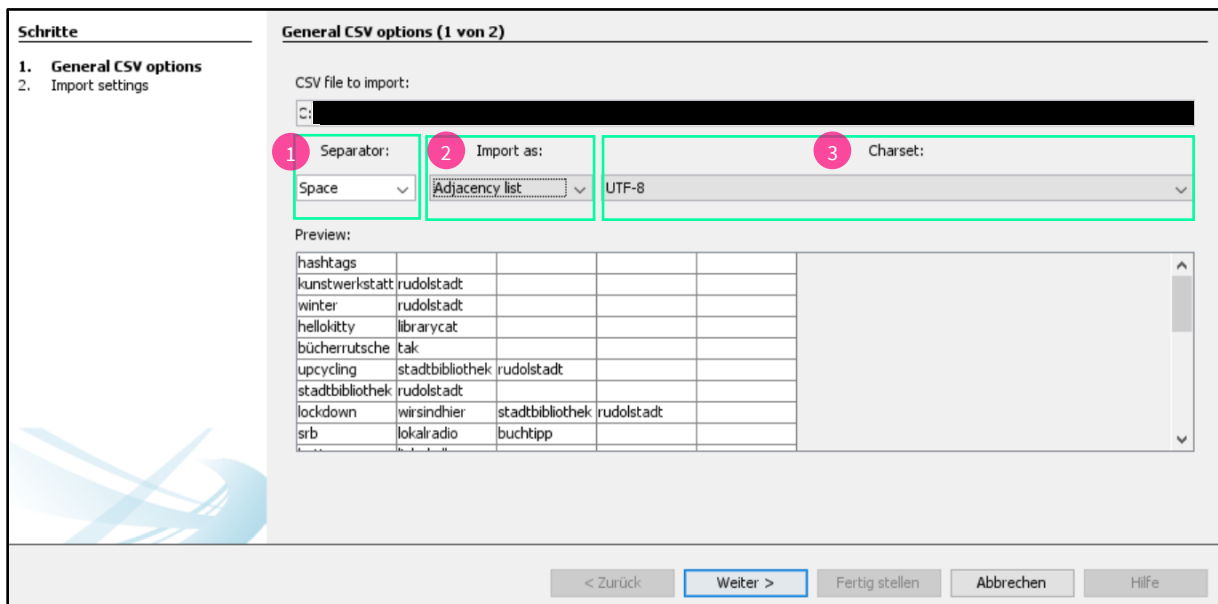


Abb. 5: Import CSV (1/2)

Um die Adjazenzliste in Gephi zu importieren, klicken wir zunächst auf *Kanten* und dann auf *Tabelle importieren*. Es öffnet sich ein Fenster, in welchem wir die entsprechende CSV-Datei auswählen. Anschließend müssen in einem weiteren Fenster (**Abb. 5**) bestimmte Optionen für den Upload angegeben werden: der **1** Separator (Comma, Semicolon, Tab oder Space) und die **3**

Kodierung (hier bzw. häufig: UTF-8) der Datei sowie die Information, dass es sich hierbei um eine **2** „Adjacency list“ handelt. Dann können wir auf *Weiter* klicken.

Im zweiten Schritt des Datenimports sind weitere Import Settings enthalten, die für uns erst einmal nicht weiter relevant sind. Wir klicken auf *Fertig stellen* und kommen dadurch zum Import-Bericht (**Abb. 6**). Hier wird uns zunächst ein **1** „Problem“ angezeigt: „Parallel edges detected, remember to choose a merge strategy“. Diese Meldung war zu erwarten, da in unserem Datensatz vermutlich viele Kanten („edges“) mehr als nur einmal auftreten. Der Default-Wert für die **2** Strategie zur Kanten-Verschmelzung ist *Summe*, d.h. alle mehrfach auftretenden Kanten werden zusammengezählt und beschreiben dann das Gewicht dieser Kante. Da wir in diesem Fall eine Art Kookkurenzanalyse von Hashtags durchführen wollen, ist der **3** Graphen-Typ ungerichtet. Unter **4** sehen wir bereits einige Informationen über unser Netzwerk, u.a. die Anzahl der Knoten und Kanten. Da unser Arbeitsbereich noch leer ist, wählen wir unter **5** am besten *Append to existing workspace* aus. Jetzt können wir auf *Ok* klicken und uns die Daten nochmal genauer ansehen.

Quelle: Datenstrom Importer Spreadsheet CSV

Probleme Bericht **1**

Nodes	Probleme
i Parallel edges detected, remember to choose a merge strategy	INFO

Graphen-Typ: Ungerichtet **3** Weitere Optionen...

☐ Auto-Skalieren **2** Strategie zur Kanten-Verschmelzung: Summe

☐ Erstelle fehlende Knoten

☐ Schleife

Anzahl Knoten: 3353

Anzahl Kanten: 6606

Dynamischer Graph: Nein

Dynamische Attribute: Nein **4**

Multi-Graph: Nein

☐ New workspace **5**

☒ Append to existing workspace

OK Abbrechen

Abb. 6: Import CSV (2/2)

Übersicht

Der zweite große Bereich „Übersicht“ ist in **Abb. 7** dargestellt. Hier finden wir neben einem Bereich für die **3** manuelle Bearbeitung des Graphen einzelne Fenster für die **1** Ausgestaltung, das **2** Layout sowie **4** Filter und Statistiken.

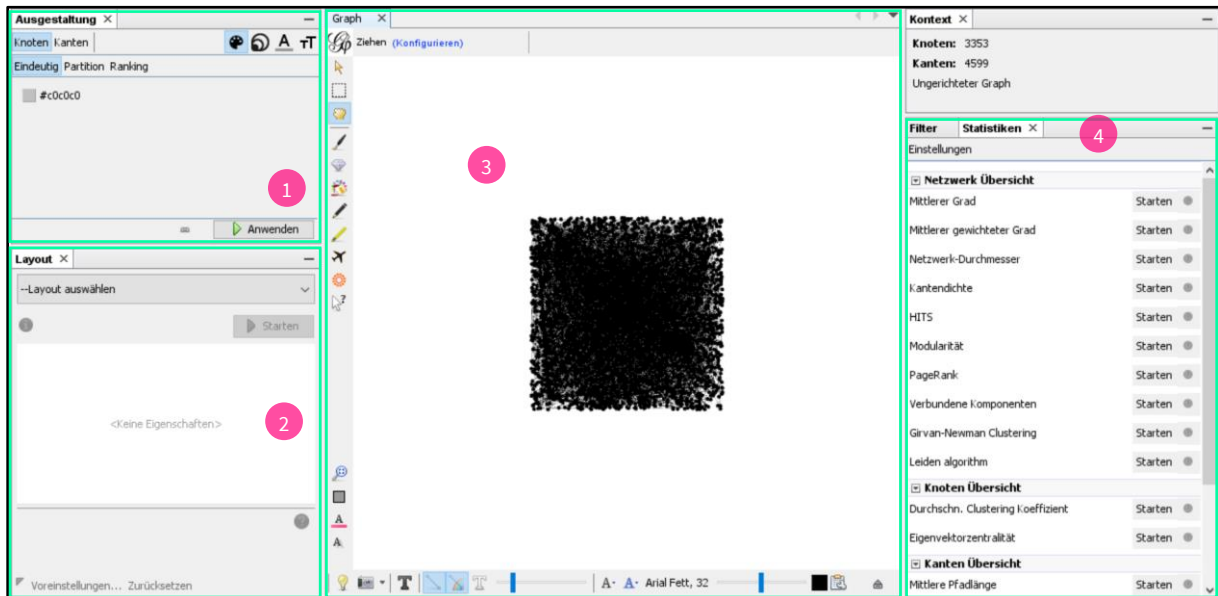


Abb. 7: Übersicht

Manuelle Bearbeitung

Abb. 8 zeigt erneut den Bereich für die manuelle Bearbeitung des **1** Graphen. Bevor wir diesen anpassen, sehen wir zunächst alle Knoten (in Form von Punkten) und Kanten (in Form von Verbindungslinien) in zufälliger Anordnung in einem quadratischen Raum. Via Scrollen können wir in den Graphen hinein- und herauszoomen (und mit **7 Auf Graph zentrieren** wieder auf ihn fokussieren) und mit Klicken und Halten lassen sich einzelne Elemente verschieben. Außerdem kann man **2** Ausschnitte auswählen und diese verschieben/bearbeiten. Hier können auch erneut **4** Knoten und Kanten hinzugefügt und die **6** Knotenattribute bearbeitet werden. Des Weiteren können **3** Farbe und Größe der Knoten geändert bzw. Knoten **5** auf Grundlage der Berechnung des kürzesten Pfades zwischen zwei Knoten oder der Distanz zu den Nachbarschaftsknoten eingefärbt werden. Auch die **9** Hintergrundfarbe kann eingestellt werden. Unter **11** befinden sich weitere Einstellungen hinsichtlich der Kanten sowie Beschriftungen. Einige Farbeinstellungen können **8** zurückgesetzt werden und es möglich, **10** Bildschirmfotos aufzunehmen und **12** weitere, allgemeine Einstellungen hierfür vorzunehmen.

Diese Bearbeitungsoptionen eignen sich vor allem dafür, einen ersten Überblick über die Daten zu bekommen bzw. in kleinen Netzwerken direkt bestimmte Aspekte hervorzuheben. Bei umfangreicheren Netzwerken macht es hingegen häufig Sinn, auf automatisch berechnete Netzwerkmetriken/Statistiken zurückzugreifen und unseren Graphen auch dahingehend zu gestalten.

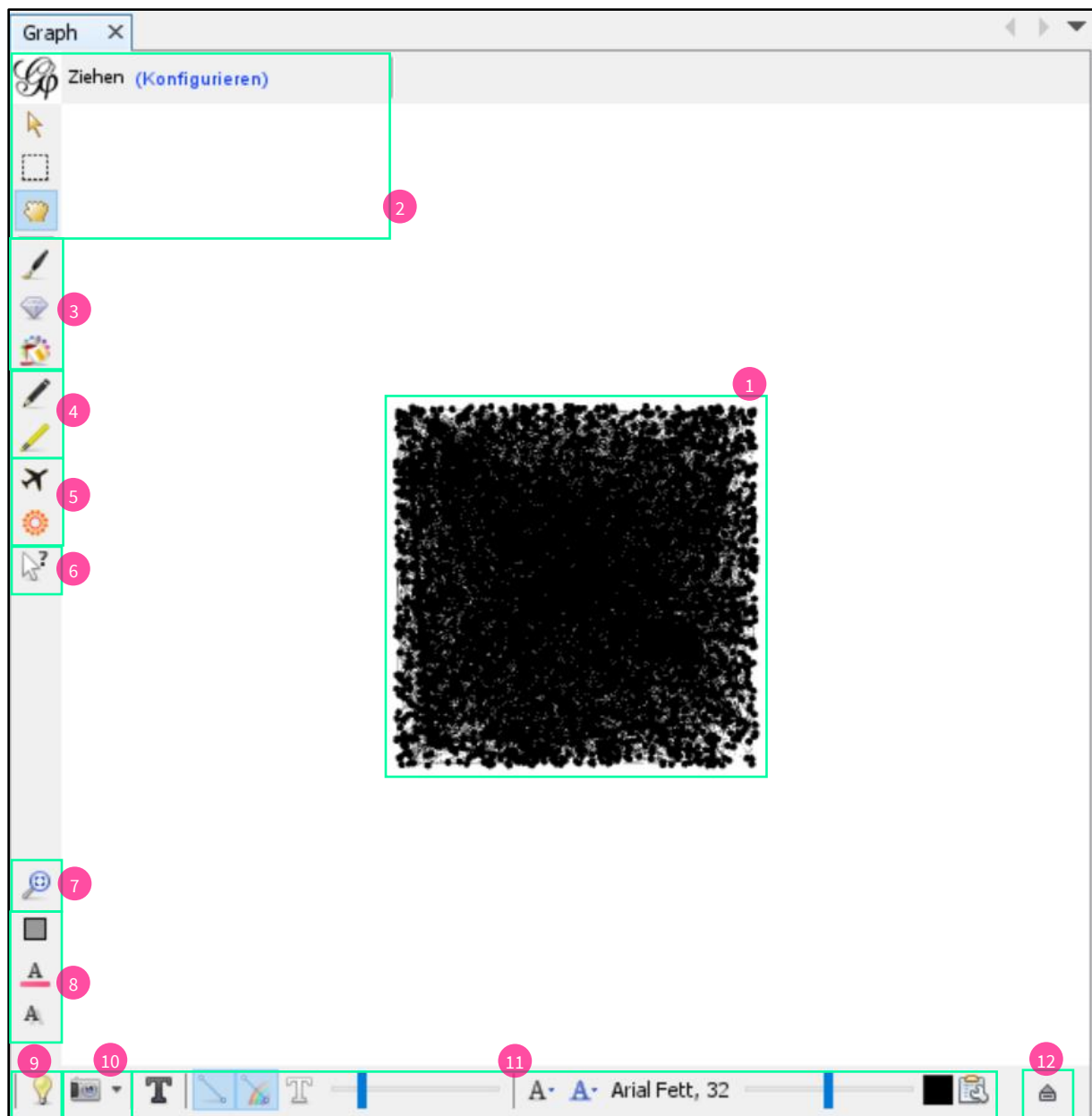


Abb. 8: Übersicht – Manuelle Bearbeitung

Layout

Damit sich die bereits vorhandene Struktur unseres Graphens auch visuell gut abzeichnet und z.B. Muster in Form einzelner Cluster erkennbar werden, müssen wir nun einen Layoutalgorithmus darauf anwenden. [Hier](#) findet sich eine gute Übersicht der in Gephi zur Verfügung stehenden Layouts. Je nach Größe und Art des Netzwerks können unterschiedliche Layouts (**Abb. 9**) besser geeignet sein. Mit einem Klick auf **1** Layout auswählen bekommen wir zunächst weitere Auswahlmöglichkeiten angezeigt. Wir können uns weitere **2** Informationen zu dem ausgewählten Layout anzeigen lassen sowie **4** verschiedene Parameter anpassen (auch hier bekommen wir mit einem Hover über den Parameter eine kurze Beschreibung dessen anzeigt). Mit einem Klick auf **3** *Starten* starten wir schließlich den Algorithmus und die Schaltfläche ändert sich ggf. zu *Stoppen*, da viele der Algorithmen auch manuell wieder beendet werden müssen. Weitere Layouts wie Expansion, Zusammenziehen, Bezeichner-Justierung oder Noverlap verändern das eigentliche

Layout nur bedingt und sorgen eher für mehr Übersichtlichkeit. Diese sollten daher erst im Nachhinein angewendet werden. Generell gilt sowohl bei den Layouts als auch bei deren Einstellungen: im Zweifelsfall einfach ausprobieren. In diesem Beispiel habe ich das Layout ForceAtlas2 in Kombination mit Expansion und Bezeichner-Justierung verwendet.

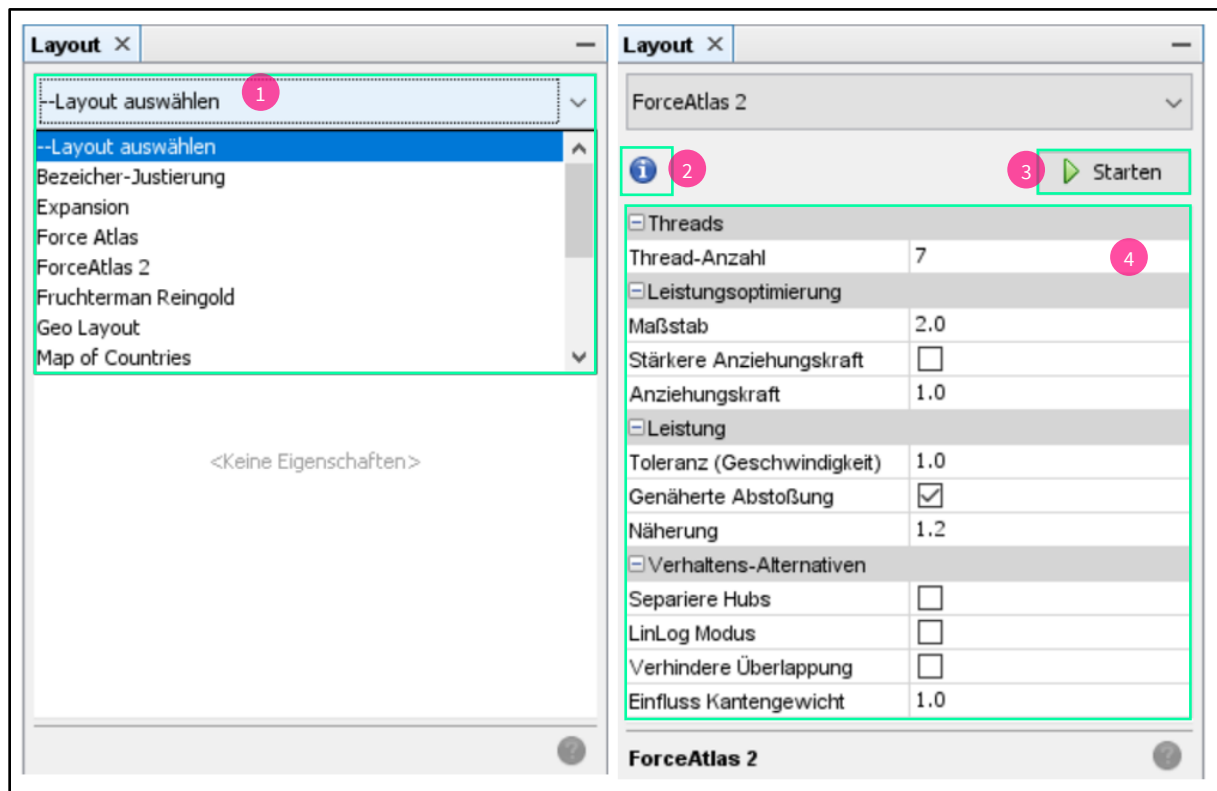


Abb. 9: Layout

Filter und Statistiken

Wichtige Funktionen in Gephi stellen auch die enthaltenen Filter und Statistiken (**Abb. 10**) dar. Um das Netzwerk zu filtern, muss zunächst **1** einer der Filter aus der Liste ausgewählt werden. Anschließend wird weiter unten ein **2** Bereich sichtbar, in welchem die konkreten Parameter für den Filter angegeben werden können. So kann man beispielsweise den Wertebereich eingrenzen, wobei die kleine Grafik die Verteilung der Werte visualisiert. Der Anteil an verbleibenden Knoten und Kanten wird jetzt weiter oben im Fenster „Kontext“ angezeigt. Anschließend kann man die Werte entweder **3** hervorheben oder **4** tatsächlich filtern. Nachdem man auf *Filter* geklickt hat, ändert sich die Schaltfläche zu *Stop*. Klickt man nun auf *Stop*, wird der Filter wieder aufgehoben. Weitere Informationen zur Verwendung von Filtern finden sich [hier](#).

Das Fenster **5** Statistiken enthält verschiedene Metriken, die in Bezug auf das gesamte Netzwerk bzw. die enthaltenen Knoten und Kanten berechnet werden können. Um eine solche Statistik zu erhalten, muss man auf **6** *Starten* klicken. Im Anschluss daran öffnet sich ein Report (**Abb. 11**), der die Ergebnisse dieser Statistik enthält. Der Report kann sofort gespeichert werden, ist aber auch nach dem Schließen noch über das kleine Fragezeichen rechts neben dem Starten-Button abrufbar. Für den nächsten Schritt rufen wir beispielhaft die Metriken Netzwerk-Durchmesser und Modularität ab.

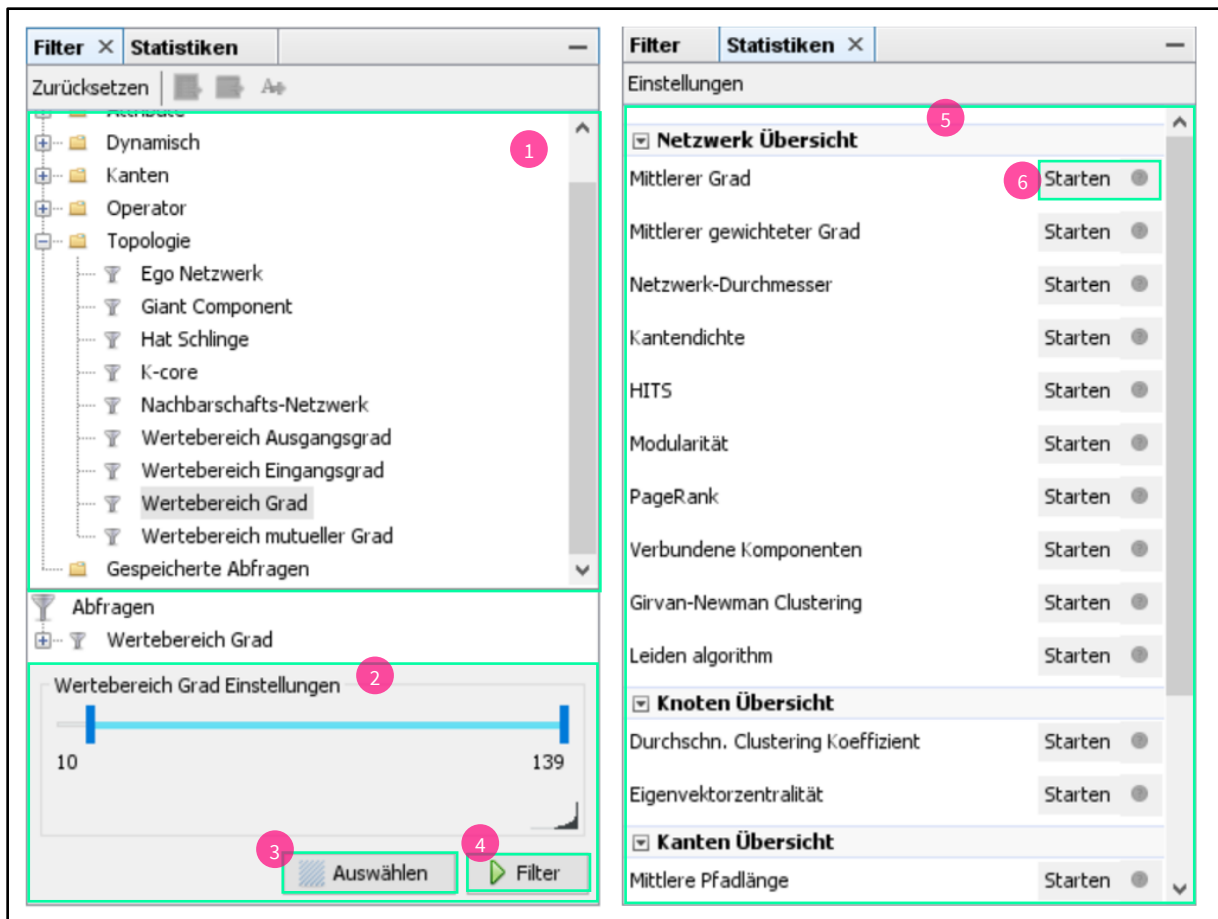


Abb. 10: Filter und Statistiken

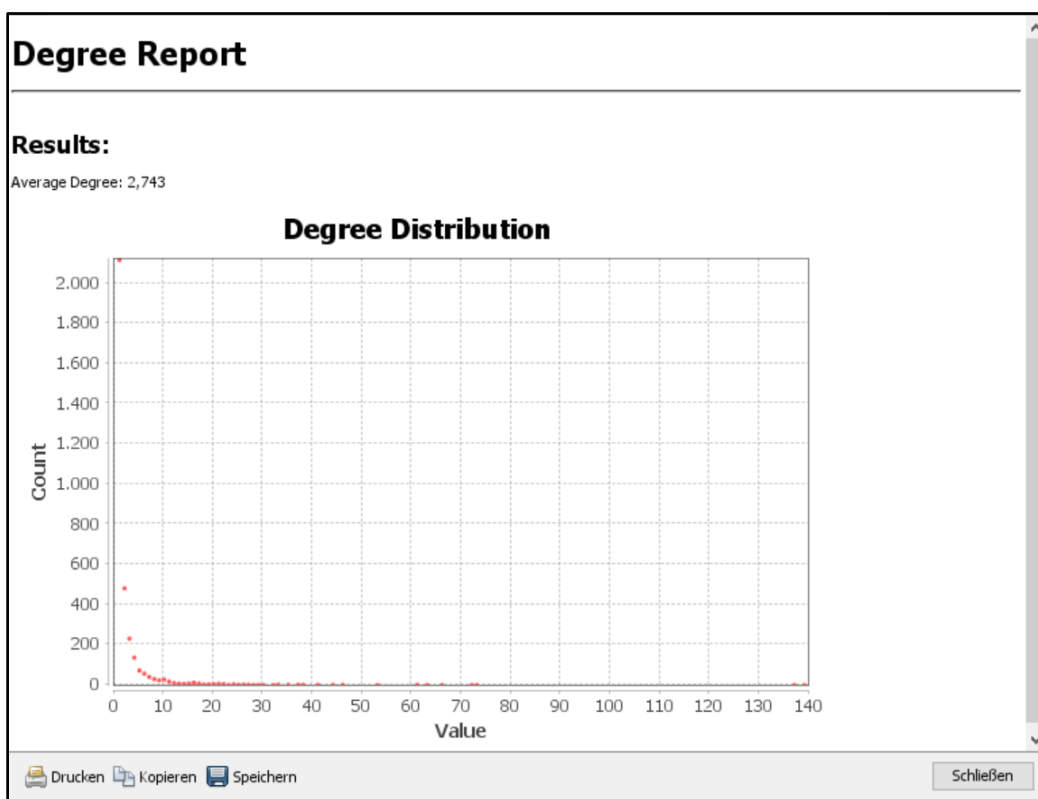


Abb. 11: Report

Ausgestaltung

Das letzte Fenster im Bereich Übersicht ist die Ausgestaltung (**Abb. 12**). Auch hier kann man wieder zwischen **1** Knoten und Kanten unterscheiden. Es stehen die Optionen **2** Farbe, Größe (nur bei Knoten), Beschriftungsfarbe und Beschriftungsgröße zur Verfügung. Neben einer eindeutigen Zuordnung ist teilweise auch die Zuordnung nach Partition bzw Ranking möglich (**3**). Bei der Partitionierung werden beispielsweise alle Knoten in Gruppen eingeteilt und nach ihrer Gruppe eingefärbt, während beim Ranking der Farbton für jeden Wert einzeln nach seiner Position auf einer Skala zugeordnet wird. Für die Gestaltung nach Partition oder Ranking können die zuvor berechneten Statistiken angewandt werden. Hierfür wird die jeweilige Metrik aus einer **4** Liste ausgewählt und ggf. noch **5** weitere Parameter eingestellt. Dann kann die Einstellung auf den Graphen **6** angewendet werden. Nun können wir beispielsweise die Größe der Knoten nach ihrem Grad (=Ranking) und die Farbe nach ihrer Modularity Class (=Partition) gestalten.

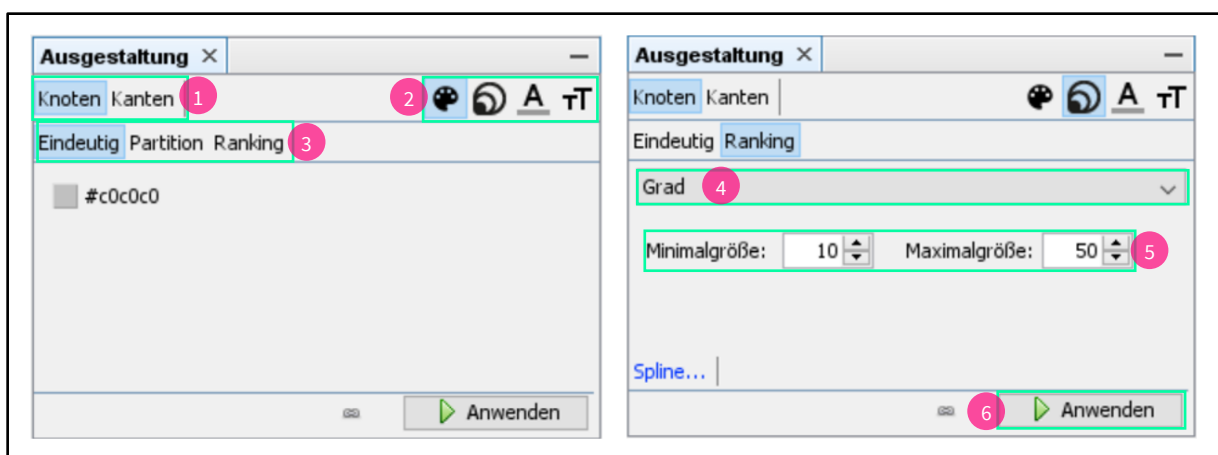


Abb. 12: Ausgestaltung

Vorschau

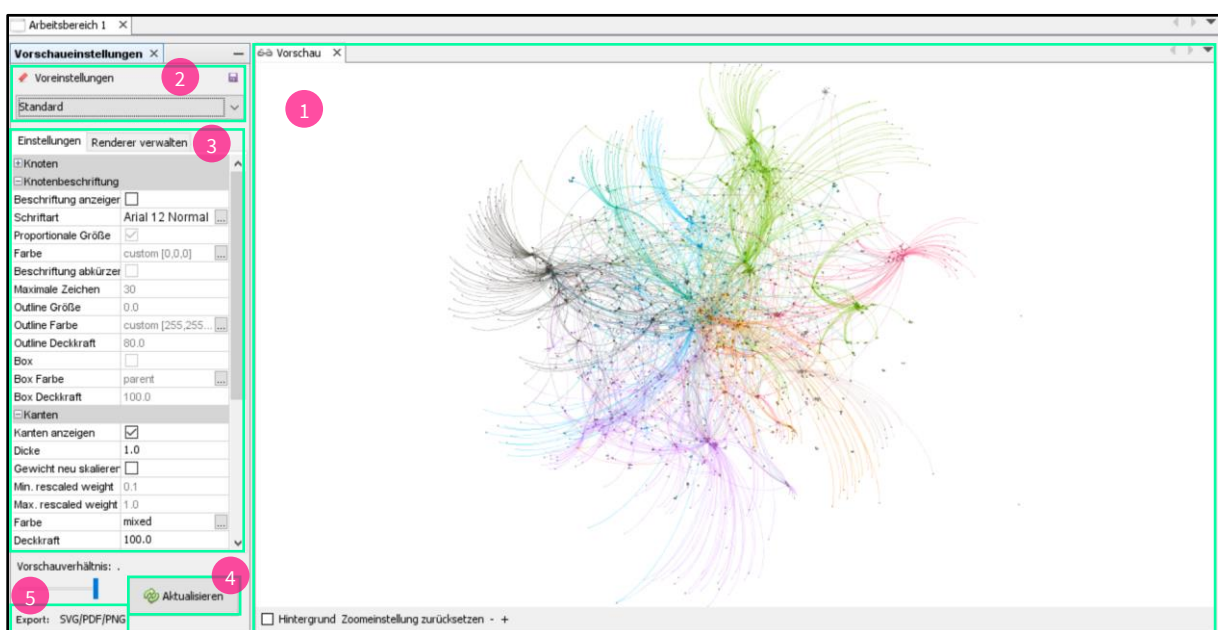


Abb. 13: Vorschau

Kommen wir zur Vorschau (**Abb. 13**), dem letzten Haupt-Bereich in Gephi. Hier sehen wir den **1** Graphen, so wie er am Ende nach dem Export aussehen wird. Um das Aussehen zu verändern, können wir zunächst auf eine der **2** Voreinstellungen zurückgreifen. Ansonsten gibt es weitere **3** Einstellungen, die wir hier vornehmen können. Da die Beschriftung nicht aus der Übersicht übernommen wird, können wir uns hier z.B. unter *Knotenbeschriftung* > *Beschriftung anzeigen* die Labels wieder anzeigen lassen. Dann setze ich noch die Deckkraft der Kanten auf 50% und die Rahmenbreite der Knoten auf 0. Hier können verschiedene Designs ausprobiert und mit **4** Aktualisieren in der Vorschau betrachtet werden. Zum Schluss kann die Visualisierung als SVG-, PDF- oder PNG-Datei **5** exportiert werden (**Abb. 14**).



Abb. 14: Ergebnis Netzwerkvisualisierung