

Lerneinheit: Netzwerkanalyse mit Gephi

Mareike Schumacher  ¹

1. Universität Regensburg

forTEXT

Thema:	Netzwerkanalyse	DOI:	10.48694/fortext.3779
Jahrgang:	1	Ausgabe:	6
Erscheinungsdatum:	2024-08-30	Erstveröffentlichung:	2020-12-04 auf fortext.net
Lizenz:			open  access

Allgemeiner Hinweis: Rot dargestellte Begriffe werden im Glossar am Ende des Beitrags erläutert. Alle externen Links sind auch am Ende des Beitrags aufgeführt.

Eckdaten der Lerneinheit

- Anwendungsbezug: Figurenkonstellation in Emilia Galotti
- Methode: Netzwerkanalyse
- Angewendetes Tool: Gephi
- Lernziele: Erstellen und Auswerten eines Figuren-Netzwerks: Anpassen des Netzwerks auf Datenebene und Individualisierung der Visualisierung, Download der Netzwerk-Grafiken
- Dauer der Lerneinheit: ca. 90 Minuten
- Schwierigkeitsgrad des Tools: mittel

Bausteine

- Anwendungsbeispiel: Welcher Primärtext liegt der Analyse zugrunde? Erstellen Sie ein Figuren-Netzwerk und analysieren Sie digital Lessings Drama Emilia Galotti.
- Vorarbeiten: Welche Arbeitsschritte sollten vor der Analyse ausgeführt werden? Vor Beginn der Analyse muss Gephi heruntergeladen bzw. installiert und Netzwerkdaten zu Lessings Drama auf Ihren Rechner heruntergeladen werden.
- Funktionen: Welche Funktionen bietet Ihnen Gephi? Mit Gephi können Sie tabellarische Übersichten erstellen, die in Netzwerkvisualisierungen umgewandelt werden. Die Netzwerkvisualisierungen sind auf vielfältige Weise individualisierbar.
- Lösungen zu den Beispielaufgaben: Haben Sie die Beispielaufgaben richtig gelöst? Hier finden Sie Antworten.

1. Anwendungsbeispiel

In dieser Lerneinheit erstellen und untersuchen Sie zwei Figuren-Netzwerke in Lessings Drama *Emilia Galotti*: Ein quantitativ angelegtes Netzwerk aus Kopräsenzen von Figuren in den einzelnen Szenen und ein Netzwerk, das aufzeigt, welche Figuren A welche anderen Figuren B in welcher Häufigkeit erwähnen. Netzwerkanalyse (Schumacher 2024b) ist ein Verfahren, das seinen Ursprung in den Sozialwissenschaften hat. Der Kerngedanke dieser Methode ist, dass der Forschungsgegenstand als System aus Knoten (Nodes) und Verbindungen (Edges) betrachtet wird. In dieser Lerneinheit verwenden wir dazu Gephi (Schumacher 2024c), ein Open-Source-Tool zur Erstellung von Netzwerkvisualisierungen.

2. Vorarbeiten

Für die Erstellung eines quantitativ ausgelegten Netzwerks der Kopräsenzen von Figuren im Drama *Emilia Galotti* benötigen Sie entweder den Primärtext in analoger oder digitaler Form oder bereits für die Netzwerkanalyse optimierte Daten. Das Projekt DraCor: Drama Corpora Project (Horstmann 2024b) stellt Netzwerkdaten für 510 deutschsprachige Dramen als Download im Gephi-kompatiblen Format bereit, darunter auch die Daten zu *Emilia Galotti*. Einige andere Tools bieten ebenfalls die Möglichkeit, für Gephi optimierte Daten direkt zu exportieren. Dazu gehört z. B. das Stilometrietool Stylo (Horstmann 2024a). Der Einfachheit halber greifen wir bei der Erstellung des Kopräsenz-Netzwerkes auf diese Daten zurück. Für das Netzwerk der Figuren-Erwähnungen stellen wir Ihnen Daten zur Verfügung, die aus einer digital annotierten (vgl. [Annotation](#)) Version des Dramentextes gewonnen wurden.

Exkurs: Um die Netzwerkdaten extrahieren zu können, wurden im digitalen Annotations-Tool CATMA (Schumacher 2024a) alle Figuren, die in der Rede der anderen Figuren erwähnt werden, mit einem für sie spezifischen

Tag (vgl. **Tagset**; z. B. Emilia, Prinz, Marinelli usw.) belegt. Außerdem wurde die gesamte Rede einer Figur mit einem Tag wie „Der Prinz spricht“, „Emilia spricht“ usw. annotiert. Eine anschließende Kollokationsabfrage (vgl. **Kollokation**) hat gezeigt, welche Figur welche anderen Figuren in welcher Häufigkeit erwähnt. Die entsprechend annotierte Fassung von *Emilia Galotti* erhalten Sie über diesen [Link](#) auf Zenodo (forTEXT 2020). Diese können Sie in CATMA hochladen, um sich die Annotationen genauer anzuschauen.

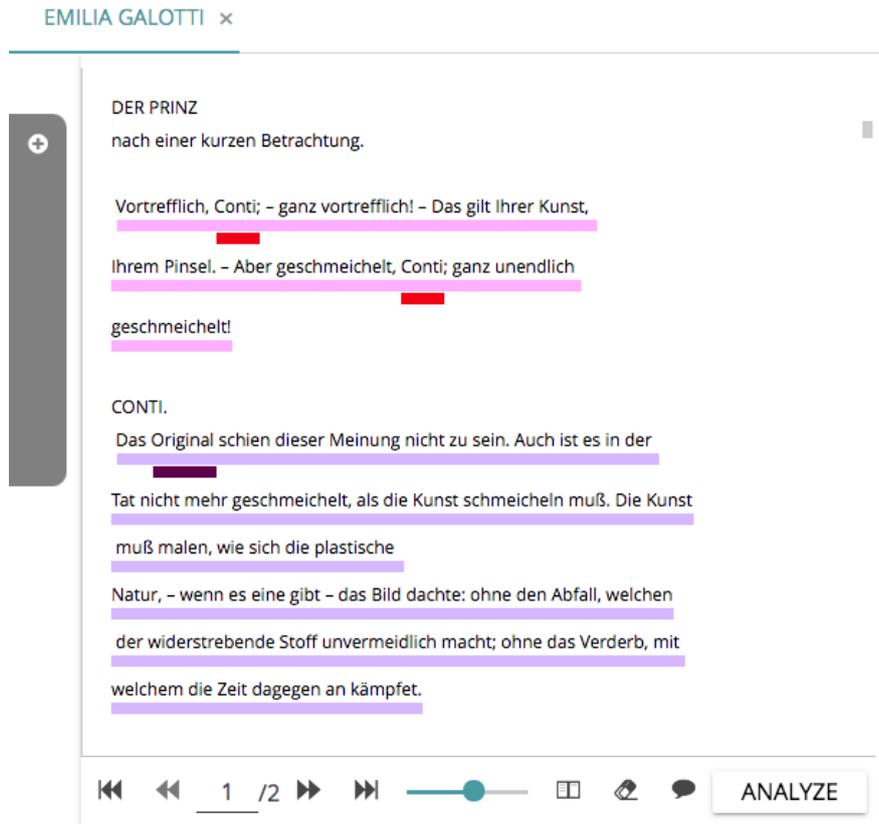


Abb. 1: Vorbereitende Annotation in CATMA

Gehen Sie nun zu DraCor und wählen Sie „GerDraCor - German Drama Corpus“ aus:

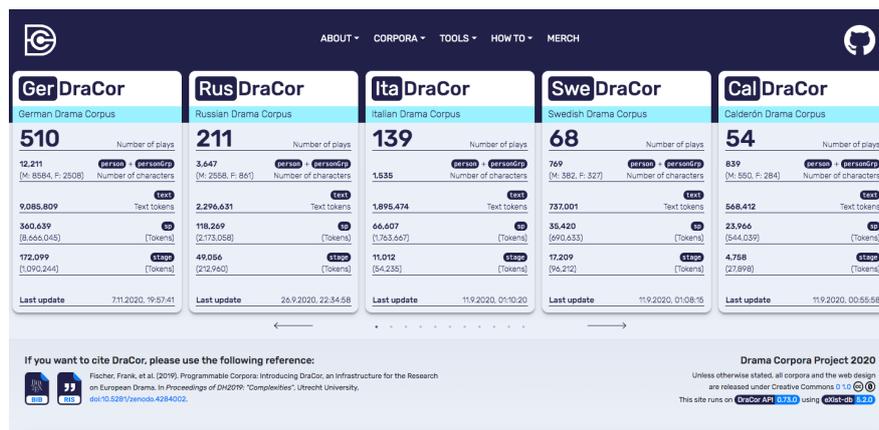


Abb. 2: Startseite des DraCor-Projektes

Wenn Sie in die Suchleiste (vgl. **Query**) „Emilia“ eingeben, werden bereits die relevanten Ergebnisse angezeigt. Klicken Sie auf „Emilia Galotti“.

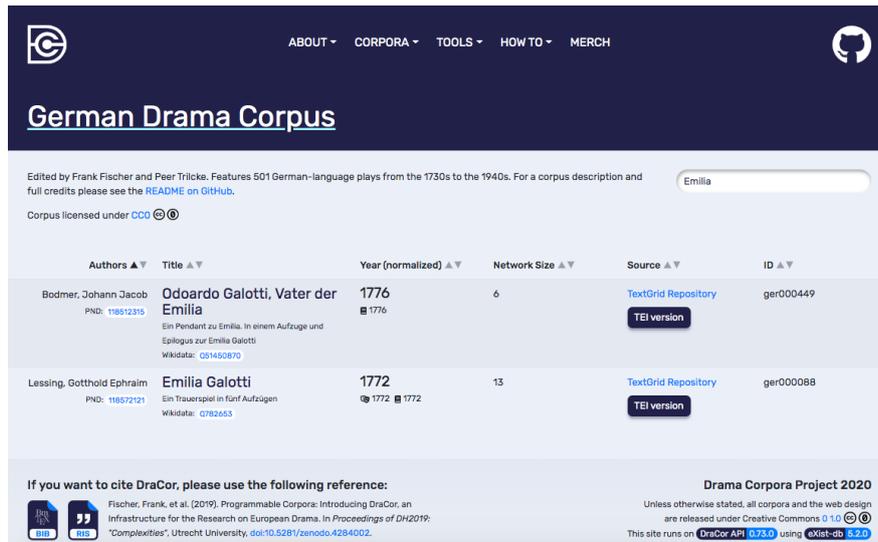


Abb. 3: Ergebnisliste für das Suchwort „Emilia“

Nun kommen Sie auf eine Seite, auf der automatisch ein Netzwerk des Dramas Emilia Galotti mit zufällig ausgewähltem Layout angezeigt wird. Damit Sie mehr darüber erfahren, wie solche Netzwerke erstellt werden und welche Berechnungen dahinter stehen, laden Sie sich nun die für Gephi optimierten Daten herunter. Klicken Sie dafür auf „Downloads“:

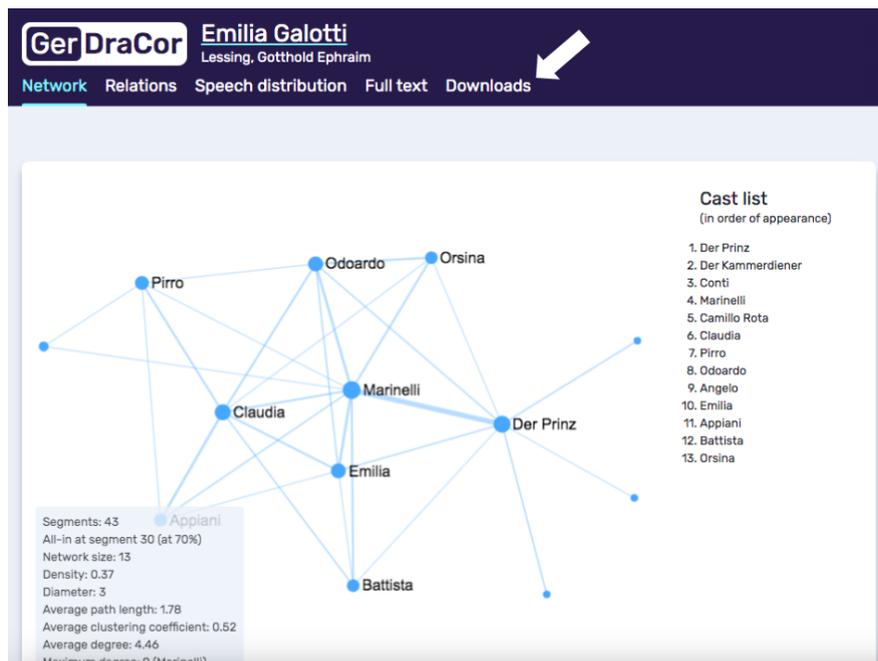


Abb. 4: DraCor-Seite für das Drama „Emilia Galotti“ mit Link zum Downloads-Bereich

Es öffnet sich eine Seite mit Downloads in einer Vielzahl an Formaten. Klicken Sie auf das Gephi-kompatible Format „GEFX“, so startet Ihr Download:

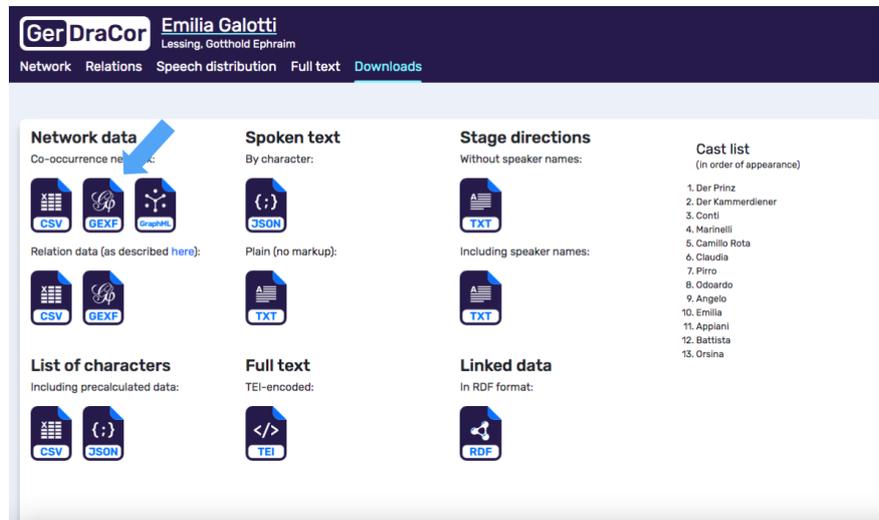


Abb. 5: Downloads-Bereich für das Drama „Emilia Galotti“

Folgen Sie nun [diesem Link](#), um Gephi für Ihr jeweiliges Betriebssystem herunterzuladen (erhältlich für MacOS, Windows und Linux). Diese Lerneinheit arbeitet mit Version 0.9.2. Nachdem Sie die Installationsdatei heruntergeladen haben, öffnen Sie diese per Doppelklick. Am Windows-PC werden Sie anschließend durch einen Installationsassistenten geführt. In MacOS wird automatisch eine ausführbare Datei in einen Ordner namens „Gephi 0.9.2“ geladen. Ist die **Installation** abgeschlossen, können Sie das Programm über einen Doppelklick auf das Gephi-Icon starten. Je nach Sicherheitseinstellungen Ihres Computers müssen Sie evtl. beim ersten Ausführen des Programms bestätigen, dass Sie dies wirklich öffnen möchten, oder eine Sicherheitsausnahme hinzufügen. Wie Sie diesen Arbeitsschritt in Ihrem Betriebssystem durchführen können, erfahren Sie in unserer Video-Anleitung (für Windows siehe forTEXT (2019b); für MacOS siehe forTEXT (2019a)). Die Startseite von Gephi sieht nach dem ersten Öffnen folgendermaßen aus:

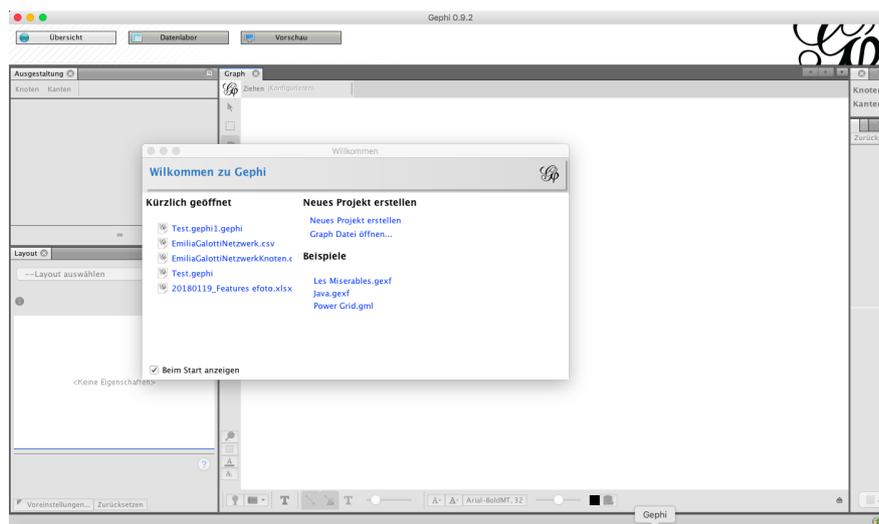


Abb. 6: Startbildschirm von Gephi

Wählen Sie die Option „Graph Datei öffnen“ und dann die gerade heruntergeladene Datei „lessing-emilia-galotti.gefx“ aus und klicken Sie auf „öffnen“. Gehen Sie im sich dann öffnenden Import-Assistenten auf „OK“:

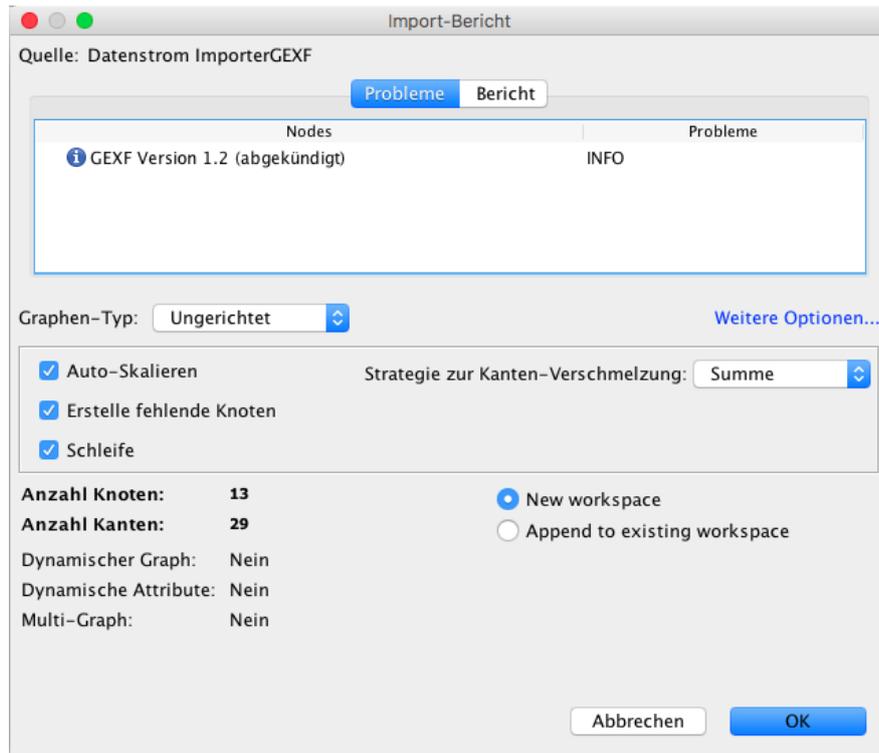


Abb. 7: Import-Assistent in Gephi

Nun landen Sie im Bereich „Übersicht“. Klicken Sie im Menü oberhalb des Arbeitsbereiches auf den Button „Datenlabor“.



Abb. 8: Menü, über das Sie den Arbeitsbereich in Gephi wechseln können.

Es öffnet sich dann eine Tabelle:

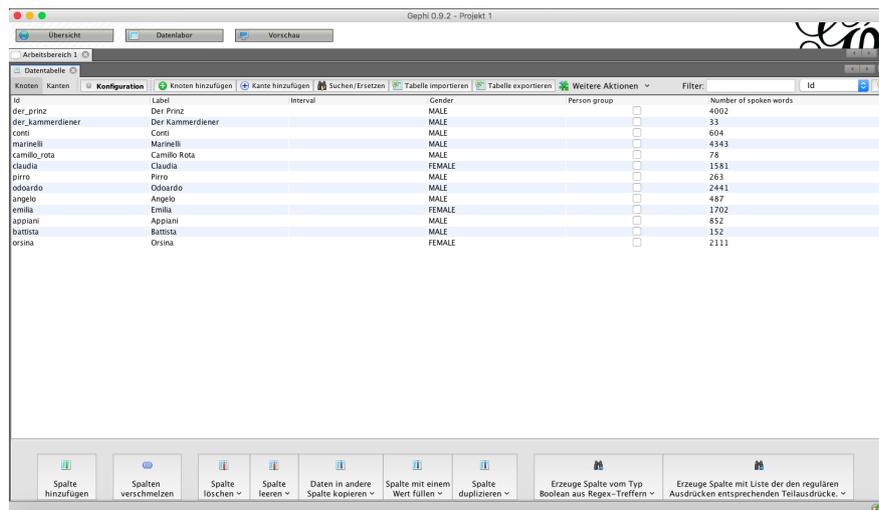


Abb. 9: Ansicht des Datenlabors in Gephi

Es handelt sich hierbei um die Tabelle mit den **Knoten** des Netzwerkes, also den Figuren. Jede Figur hat eine ID und ein Label, das zur späteren Beschriftung dient. In einer weiteren Spalte ist vermerkt, ob es sich um weibliche oder männliche Figuren handelt. In der letzten Spalte steht, wie viele Wörter die jeweilige Figur insgesamt

spricht. Es handelt sich bei den letzten beiden Spalten um Zusatzinformationen, mit denen Sie Ihrem Netzwerk später unterschiedliche Ebenen hinzufügen können.

Wechseln Sie nun über die in Abb. 9 markierte Schaltfläche auf „**Kanten**“:

Ursprung	Ziel	Typ	id	Label	Interval	Weight
der_prinz	der_kammerdiener	Ungerichtet	der_prinz/der_kammerdiener			2.0
der_prinz	conti	Ungerichtet	der_prinz/conti			2.0
der_prinz	marinelli	Ungerichtet	der_prinz/marinelli			9.0
der_prinz	camilla_rosa	Ungerichtet	der_prinz/camilla_rosa			1.0
der_prinz	emilia	Ungerichtet	der_prinz/emilia			2.0
der_prinz	battista	Ungerichtet	der_prinz/battista			1.0
der_prinz	orsina	Ungerichtet	der_prinz/orsina			1.0
der_prinz	odaardo	Ungerichtet	der_prinz/odaardo			2.0
marinelli	pirro	Ungerichtet	marinelli/pirro			1.0
marinelli	appiani	Ungerichtet	marinelli/appiani			2.0
marinelli	claudia	Ungerichtet	marinelli/claudia			3.0
marinelli	angelo	Ungerichtet	marinelli/angelo			1.0
marinelli	battista	Ungerichtet	marinelli/battista			3.0
marinelli	emilia	Ungerichtet	marinelli/emilia			4.0
marinelli	orsina	Ungerichtet	marinelli/orsina			3.0
marinelli	odaardo	Ungerichtet	marinelli/odaardo			4.0
claudia	pirro	Ungerichtet	claudia/pirro			3.0
claudia	odaardo	Ungerichtet	claudia/odaardo			3.0
claudia	emilia	Ungerichtet	claudia/emilia			3.0
claudia	appiani	Ungerichtet	claudia/appiani			4.0
claudia	battista	Ungerichtet	claudia/battista			2.0
claudia	orsina	Ungerichtet	claudia/orsina			1.0
pirro	angelo	Ungerichtet	pirro/angelo			1.0
pirro	odaardo	Ungerichtet	pirro/odaardo			1.0
pirro	appiani	Ungerichtet	pirro/appiani			1.0
odaardo	orsina	Ungerichtet	odaardo/orsina			3.0
odaardo	emilia	Ungerichtet	odaardo/emilia			2.0
emilia	appiani	Ungerichtet	emilia/appiani			1.0
emilia	battista	Ungerichtet	emilia/battista			1.0

Abb. 10: Ansicht der Kanten-Tabelle in Gephi

Hier sehen Sie jeweils Ursprung und Ziel einer Relation. Es handelt sich um ein sog. „**ungerichtetes**“ **Netzwerk**, d.h. es ist nicht relevant, welches der Ursprungsknoten und welches der Zielknoten ist (im Gegensatz zu dem später angelegten Netzwerk der Erwähnungen, in dem es von Bedeutung sein wird, welche Figur (Ursprung) welche andere Figur (Ziel) erwähnt). Jede Verbindung hat eine ID, die Ursprung und Ziel beinhaltet und unter „weight“ ist angegeben, wie häufig die Figuren zusammen auftreten.

Sie haben nun Ihre Daten soweit vorbereitet, dass Sie in Gephi ihr Netzwerk konfigurieren können. Außerdem haben Sie das Programm auf Ihrem Rechner installiert, unter dem Menüpunkt „Übersicht“ einen ersten Eindruck von einem vorgefertigten Netzwerk gewonnen und die wichtigsten Bestandteile des Datenlabors kennengelernt. Damit sind Sie nun bereit für die Erstellung Ihrer Netzwerkanalyse mit Gephi.

3. Funktionen

Dadurch, dass Sie vorbereitete Netzwerkdaten nutzen, überspringen Sie das Anlegen der Datentabellen in Gephi. Wenn Sie zu einem späteren Zeitpunkt ein eigenes Netzwerk anlegen möchten, für das es keine vorbereiteten Daten gibt, können Sie im folgenden Video erlernen, die Funktionalitäten des „Datenlabors“ zu nutzen:

Verlassen Sie das Datenlabor und gehen Sie im Menü auf „Übersicht“. Gephi hat Ihnen nun aus den Daten im „Datenlabor“ ein **Netzwerk** errechnet, das aus kleinen schwarzen Punkten und dünnen schwarzen Linien besteht. Die Knoten werden hier alle gleich groß angezeigt, die Kantendicke entspricht dem Wert im Tabellenfeld „weight“, die Länge der Kanten hat in diesem Layout keine Bedeutung. Der nun folgende Schritt der **Layoutanpassung** des Netzwerkes ist ein wichtiger interpretativer Vorgang. Er sollte darum stets im Bewusstsein darüber ausgeführt werden, dass es sich hierbei um eine Interpretation der Daten handelt (vgl. Textvisualisierung (Horstmann und Stange 2024)). Wird eine Visualisierung später verwendet, um ein Argument in einer Literaturanalyse zu unterstützen, so sollte stets eine Bildunterschrift gewählt werden, aus der die Parameter der Visualisierung klar hervorgehen.

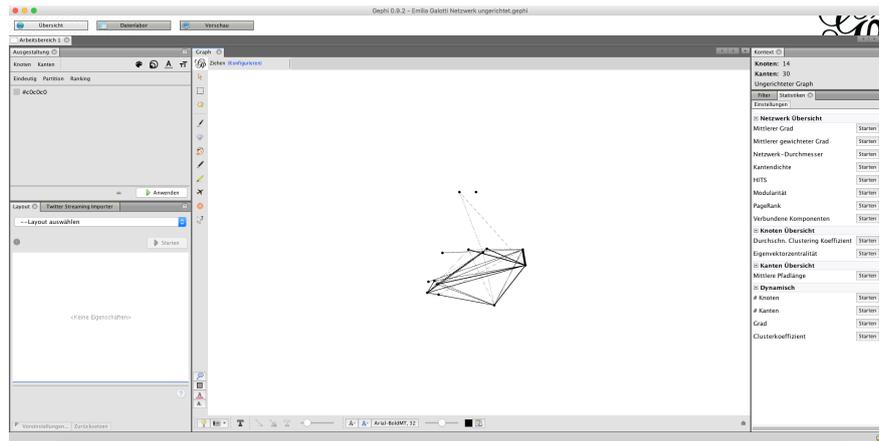


Abb. 11: Arbeitsbereich „Übersicht“, nachdem im „Datenlabor“ ein Netzwerk angelegt wurde

Im „Arbeitsbereich“ auf der linken Seite stehen Ihnen zwei Fenster zur Anpassung Ihrer Visualisierung zur Verfügung. Im oberen Bereich „Ausgestaltung“ lassen sich Größe und Dicke von Knoten und Kanten sowie deren Farben bearbeiten. Im unteren Arbeitsbereich „Layout“ können Sie aus unterschiedlichen Typen von Netzwerken auswählen.

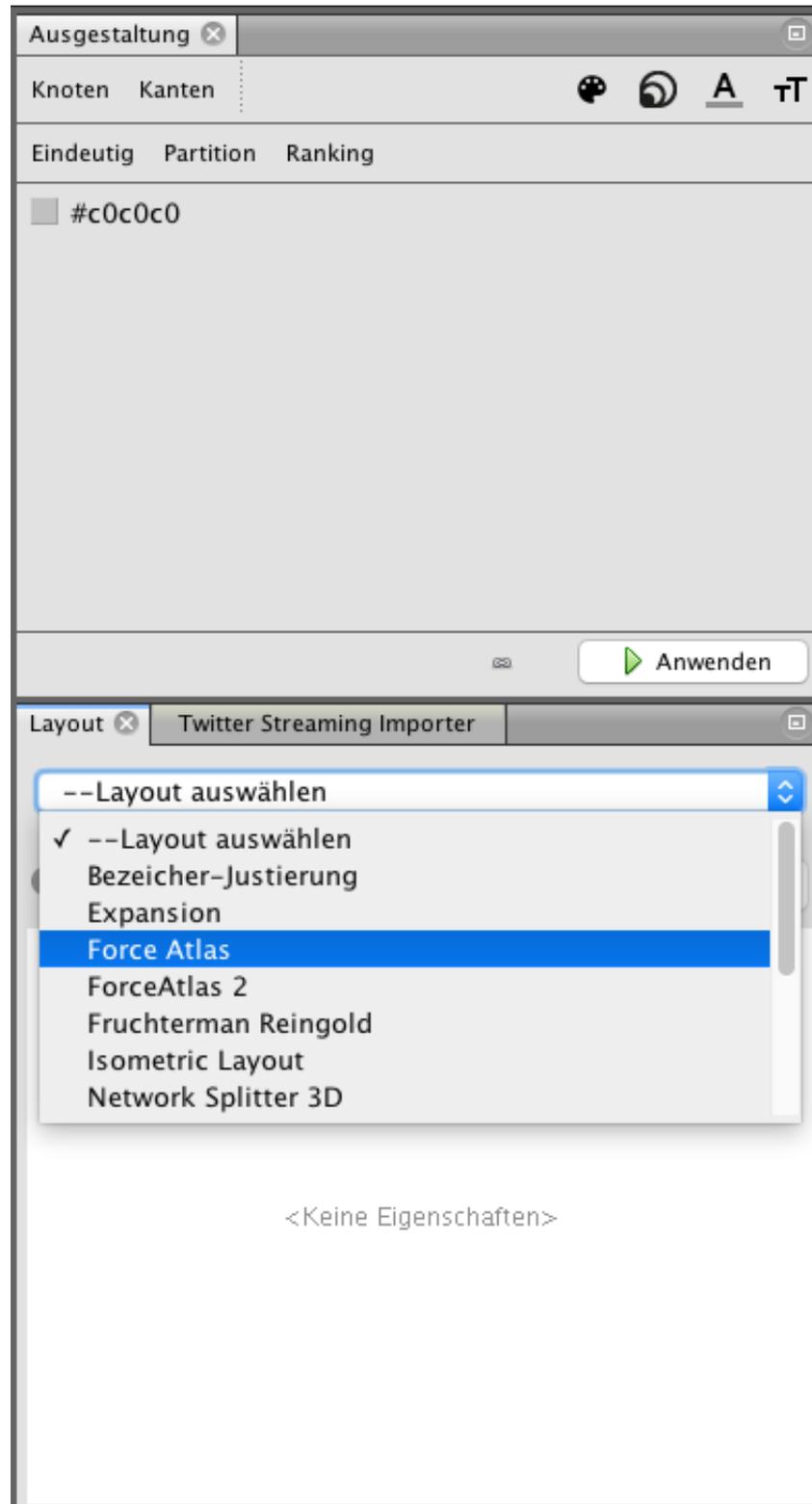


Abb. 12: Drop-Down-Menü zur Auswahl des Netzwerk-Layouts

Aufgabe 1: Wählen Sie im Drop-Down-Menü des unteren Arbeitsbereiches das Layout „Force Atlas“ und klicken Sie auf die Schaltfläche „starten“. Beschreiben Sie Ihre Beobachtungen. Probieren Sie anschließend andere Layouts aus, bis Sie eines gefunden haben, das Ihnen passend für die Visualisierung der Figurenkonstellation erscheint. Achtung: Bei manchen Layouts müssen Sie nicht nur den Start per Schaltfläche auslösen, sondern auch per Klick auf „Stop“ die Layout-Veränderung beenden (der „Start“-Button wird in diesem Fall zum „Stop“-Button). Wenn Sie ein Layout ausgewählt haben, können Sie anschließend noch das Layout „Noverlap“ (einmalig) oder „Expansion“ (einmalig oder mehrmalig) anwenden, um die Lesbarkeit des Netzwerkes zu verbessern. „Noverlap“

sorgt dafür, dass Knoten und Kanten nicht überlappen (tun sie das in Ihrem ausgewählten Layout von Beginn an nicht, so wird „Noverlap“ nichts mehr verändern). Mit „Expansion“ können sie das Netzwerk weiter auseinander ziehen.

Gehen Sie nun zum oberen Arbeitsbereich, um **Farben und Gewichtungen anzupassen**. Die (nicht sichtbare) Vorauswahl zeigt die Menüpunkte „Knoten“, „eindeutig“ und „Farbe“. Sie können nun wählen zwischen „Partition“ - jede Gewichtung bekommt eine eigene Farbe zugewiesen - und „Ranking“ - alle Knoten bekommen dieselbe Farbe, die Sättigung variiert aber nach Gewichtung. Partition bezieht sich hier auf die Möglichkeit, die Knotenpunkte in mittelgroßen bis großen Netzwerken zu gruppieren. Diese Gruppierung ist auch Teil der Berechnungen bestimmter Netzwerk-Layouts (z. B. Network Splitter 3D). Gehen Sie zuerst auf Partition, wählen Sie ein Attribut (hier können Sie zwischen „Gender“ und „Number of spoken words“ wählen). Gehen Sie auf Ranking, so können Sie den „Grad“ (gemeint ist der Vernetzungsgrad, also die Anzahl an Verbindungen, die von einem Knoten ausgehen) farblich visualisieren.

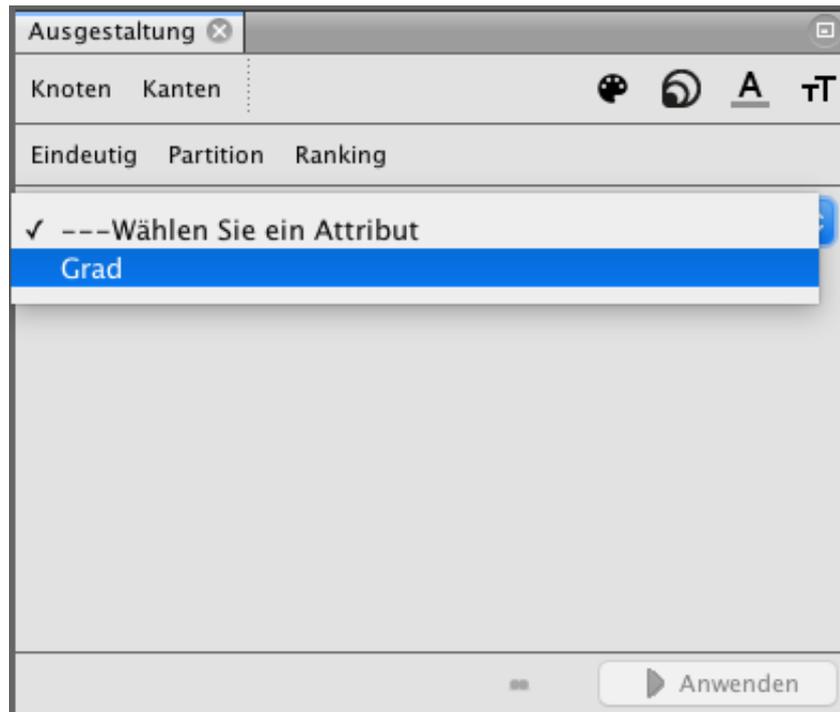


Abb. 13: Dropdown-Menü zur Verknüpfung des Vernetzungsgrads mit der Farbgebung

Übrigens: beim „Ranking“ kommen Sie mit einem Klick auf die kleine Farbpalette rechts neben dem Farbbalken zur Farbauswahl.

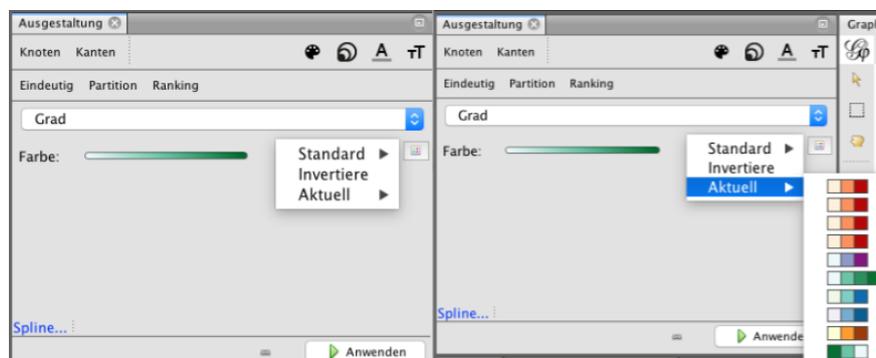


Abb. 14: Anpassung der Farbauswahl

Gehen Sie nun oben rechts im Menü (mit den kleinen Icons) auf das zweite Icon, das für „Größe“ steht.

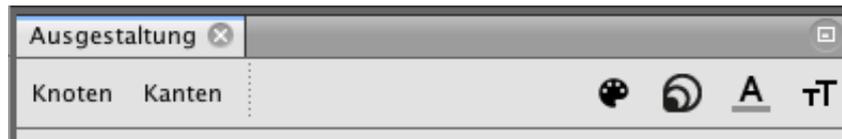


Abb. 15: Die Icons in dieser Menüauswahl stehen für: Farbpalette = Farbe, unterschiedlich große Kreise = Größe, unterstrichenes A = Schriftfarbe und unterschiedlich große Ts = Schriftgröße

Hier können Sie zwischen Eindeutig und Ranking wählen. Die Vorauswahl des Tools ist Eindeutig und Größe 1. Hier können Sie die Größe für alle Knoten auf einen Wert festlegen.

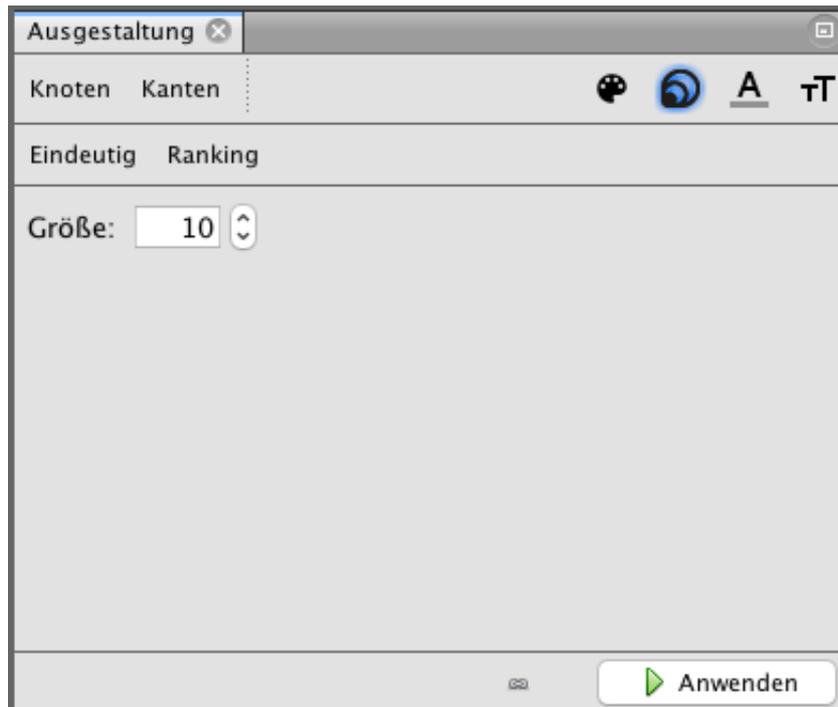


Abb. 16: Standard-Einstellung für Knotengröße

Alternativ können Sie unter „Ranking“ im Drop-Down Menü ein Attribut auswählen (in unserem Beispiel steht hier mindestens „Grad“ und „number of spoken words“ zur Auswahl. Je nachdem, welches Layout Sie gewählt haben, kommen errechnete Größen wie z. B. Modularity Class hinzu), eine Minimal- und eine Maximalgröße eingeben und dann die Knoten über die Schaltfläche „Anwenden“ in variierenden Größen anzeigen lassen. Die Größe visualisiert z. B. die Anzahl der von einem Knoten ausgehenden Verbindungen (Grad), die Anzahl gesprochener Wörter der Figur oder die Zugehörigkeit zu einem Knoten-Cluster (Modularity Class).

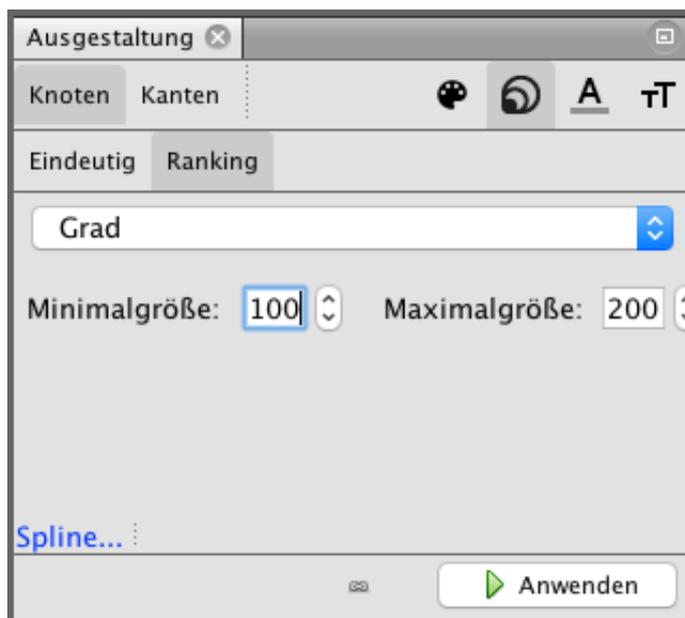


Abb. 17: Auswahlfenster für Minimal- und Maximalgröße der Knoten

Aufgabe 2: Wählen Sie die Farbe der Knoten als genderspezifische Visualisierung aus. Lassen Sie sich die Knotengröße entsprechend der Anzahl gesprochener Wörter anzeigen. Wiederholen Sie die Arbeitsschritte für die Kanten. Legen Sie alle Größen und Farben für Ihr Netzwerk fest. Was passiert, wenn Sie Größen und Farben der Label ändern?

Nachdem wir die automatischen Anpassungen im Arbeitsbereich links angewandt haben, gehen wir nun dazu über, das **Netzwerk manuell zu verändern**. Dazu nutzen wir den Bereich in der Mitte, der mit „Graph“ überschrieben ist. Sie haben in diesem Feld zwei Menüleisten. Die eine befindet sich links neben dem Netzwerk-Graphen, die andere darunter. Die untere Menüleiste ist standardmäßig eingeklappt. Klappen Sie diese aus, indem Sie auf das winzig kleine Icon oben rechts in der Ecke der unteren horizontal angeordneten Menüleiste klicken.



Abb. 18: Ganz rechts in dieser Menüleiste befindet sich der Button zum Ausklappen

Die Funktionen in den beiden Menüleisten werden angezeigt, wenn Sie mit der Maus über ein Icon hovern. Die wichtigsten Funktionen werden wir nun erkunden.



Abb. 19: Ausgeklapptes unteres Menü

Als erstes schalten Sie die Knoten-Beschriftung an, indem Sie auf das kleine schwarze „T“ klicken, das Sie in der unteren Menüleiste oben links an dritter Stelle finden. Rechts daneben können Sie die Kantenanzeige ein- und ausschalten, die Dicke der Kanten mit einem Schieberegler verändern oder die Schriftarten für die Beschriftung festlegen. Die Größe der Knotenbeschriftungen können Sie über den Schieberegler rechts neben den Schriftarten vergrößern oder verkleinern. Im unteren Bereich finden Sie ausführlichere Bearbeitungsmöglichkeiten für den gesamten Graphen („Global“), die Knoten und die Kanten. Hier können Sie z. B. veranlassen, dass die Benennung der Kanten angezeigt wird, indem Sie unter „Bezeichnung“ auf „Konfigurieren“ gehen und die Einstellungen für Kanten ändern.

Mit den Funktionen in der linken Menüleiste können Sie direkt **in die Gestaltung des Netzwerkes eingreifen**. Vorausgewählt ist die Funktion „ziehen“ (drittes Icon von oben in der linken Menüleiste), wodurch Sie mit

der Maus per Drag and Drop die Knoten an eine beliebige Stelle versetzen können. Die beiden Icons, die über dem mit der kleinen Faust stehen, eignen sich zum Erkunden des Netzwerks. Per direkter Auswahl (Pfeil) oder Bereichsauswahl (Rechteck) lassen sich einzelne Knoten oder bestimmte Bereiche hervorheben. Ohne die Struktur des Netzwerkes zu verändern, können Sie sich hier die einzelnen Verbindungen in ihrem Netzwerkes anzeigen lassen, was vor allem bei der Interpretation komplexer Netzwerke hilfreich sein kann.



Abb. 20: Funktionen zur Netzwerk-Erkundung

Mit den Funktionen unterhalb der kleinen Faust wird das Netzwerk in seiner Gestaltung grundlegend verändert. Hier kann manuell eine bestimmte Farbe für einzelne Knoten oder Kanten ausgewählt werden, die Größe von Knoten lässt sich einzeln manipulieren und Kanten lassen sich direkt von Knoten zu Knoten anlegen bzw. hinzufügen.



Abb. 21: Funktionen zur Anpassung des Layouts bzw. der Datenstruktur

Aufgabe 3: Legen Sie Ihr Netzwerk so an, dass Marinelli im Mittelpunkt steht. Welches ist der kürzeste Pfad zwischen Marinelli und Conti (zwei Figuren, die nie kopräsent sind)? Legen Sie einen Knoten „weitere Bediente“ an. Fügen Sie jeweils eine Verbindung zwischen „weitere Bediente“, Pirro und Battista hinzu. Benennen Sie diese Verbindung als „sind“ (Tipp: Sie müssen dazu zurück in das Datenlabor gehen). Lassen Sie sich die Benennung der Kante im Graphen anzeigen. Wenn Ihnen Ihr Netzwerk gut gefällt, machen Sie einen Screenshot davon. Nutzen Sie dazu nicht die Funktion Ihres Rechners, sondern die Gephi-eigene Screenshot-Funktion. Verlassen Sie nun den Bereich „Übersicht“ und gehen Sie im oberen Menü auf „Vorschau“.



Abb. 22: Oberes Menü, über das der Arbeitsbereich gewechselt wird

Klicken Sie unten links auf den Button „Aktualisieren“.



Abb. 23: „Aktualisieren“-Button zur Anzeige des Netzwerkes in der „Vorschau“

Sie erhalten nun eine Vorschau Ihres Netzwerkes und können die Darstellung abschließend bearbeiten. Alle Anpassungen, die Sie im Bereich „Vorschau“ vornehmen, haben keine Auswirkungen auf Ihre Daten, bzw. hier können Sie nicht mehr am Datenmaterial arbeiten (z. B. können keine zusätzlichen Kanten hinzugefügt werden). Im Dropdownmenü unter „Voreinstellungen“ wählen Sie eine Darstellungsweise aus. Im Gegensatz zur Wahl des Layouts im Bereich „Übersicht“ hat diese Auswahl keine Auswirkungen auf das Netzwerk-Gefüge. Sie können lediglich bestimmen, ob die Kanten gerade oder bogenförmig verlaufen sollen und welche Farbe der Hintergrund haben soll. Achtung: Ihre Vorauswahl wird erst sichtbar, wenn Sie unten auf „Aktualisieren“ klicken.

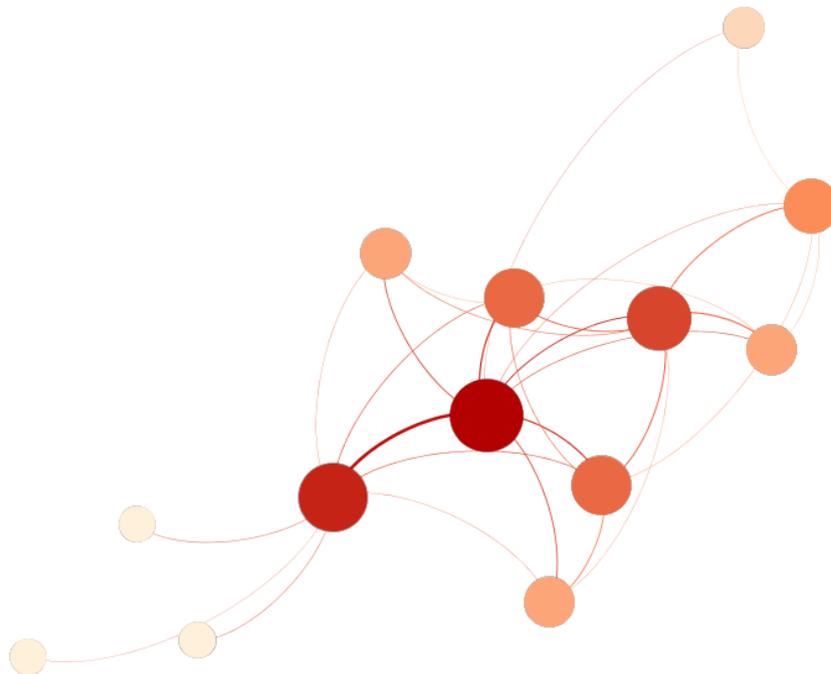


Abb. 24: Vorschau der Netzwerk-Visualisierung

Im Bereich „Einstellungen“ - direkt unter dem zuvor betätigten Dropdownmenü - finden Sie fünf Bereiche, die Sie per Mausklick auf das Pfeilsymbol ein- und ausklappen können. Hier können Sie die Darstellung der Knoten und Knotenbeschriftungen, der Kanten und Kantenbeschriftungen und der Kantenpfeile (die bei einem ungerichteten Netzwerk nicht vorhanden sind) ändern. Auch hier gilt: Jede Änderung wird erst sichtbar, wenn Sie auf die Schaltfläche „Aktualisieren“ klicken.

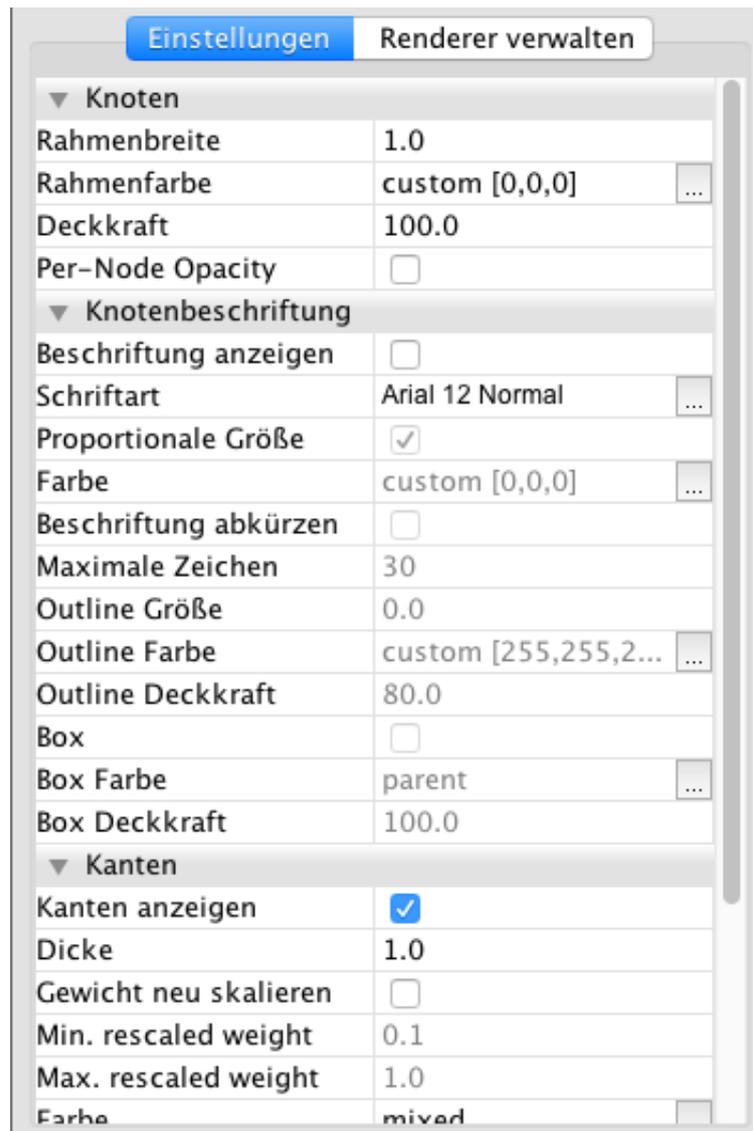


Abb. 25: Fenster zur optischen Feinjustierung des Netzwerkes

Aufgabe 4: Setzen Sie ein Häkchen bei der Knotenbeschriftung unter „Beschriftung anzeigen“. Ändern Sie die Schriftgröße der Knotenbeschriftung auf mindestens 10 Punkt. Legen Sie die Größeneinstellung der dünnsten Kanten auf 6.0 Punkt und die der dicksten Kanten auf 15.0 fest. Sorgen Sie dafür, dass die Beschriftungen der Knoten sich nicht überlappen (Achtung, damit greifen Sie in die Netzwerkdaten ein. Sie müssen also zurück in den Bereich „Übersicht“). Exportieren Sie Ihr Netzwerk über die Gephi-eigene Exportfunktion und vergleichen Sie das Bild mit dem zuvor erstellten Screenshot aus dem Bereich „Übersicht“.

Sie haben nun die wichtigsten Funktionen in Gephi bereits kennen gelernt. Sie haben ein Netzwerk der Kopräsenzen gestaltet, die Daten angepasst und die Visualisierung gestaltet. Laden Sie sich auf Zenodo die Netzwerk-Daten (Knotentabelle, Kantentabelle) für ein Netzwerk herunter, das die Erwähnungen von Figuren in der Figurenrede visualisiert (forTEXT 2020).

Aufgabe 5: Öffnen Sie die heruntergeladene Datei in Gephi. Schauen Sie sich die Datentabellen an. Welche Unterschiede im Vergleich zum ersten Netzwerk fallen Ihnen auf? Gestalten Sie dann in den Bereichen „Übersicht“ und „Vorschau“ ein Netzwerk, das in Ihren Augen die Erwähnungen in der Figurenrede gut veranschaulicht. Vergleichen Sie Ihre Visualisierung mit der Grafik des ersten Netzwerkes. Welche Ähnlichkeiten und Unterschiede fallen Ihnen auf?

4. Lösungen zu den Beispielaufgaben

Aufgabe 1: Wählen Sie im Drop-Down-Menü des unteren Arbeitsbereiches das Layout „Force Atlas“ und klicken Sie auf die Schaltfläche „starten“. Beschreiben Sie Ihre Beobachtungen. Probieren Sie anschließend andere Layouts aus, bis Sie eines gefunden haben, das Ihnen passend für die Visualisierung der Figurenkonstellation

erscheint. Achtung: Bei manchen Layouts müssen Sie nicht nur den Start per Schaltfläche auslösen, sondern auch per Klick auf „Stop“ die Layout-Veränderung beenden (der „Start“-Button wird in diesem Fall zum „Stop“-Button). Wenn Sie ein Layout ausgewählt haben, können Sie anschließend noch das Layout „Noverlap“ (einmalig) oder „Expansion“ (einmalig oder mehrmalig) anwenden, um die Lesbarkeit des Netzwerkes zu verbessern. „Noverlap“ sorgt dafür, dass Knoten und Kanten nicht überlappen (tun sie das in Ihrem ausgewählten Layout von Beginn an nicht, so wird „Noverlap“ nichts mehr verändern). Mit „Expansion“ können sie das Netzwerk weiter auseinander ziehen.

Wenn Sie das Layout „Force Atlas“ wählen, beginnt Ihr Netzwerk, sich nach und nach zu verändern. Wenn Sie anschließend das Layout „Noverlap“ auswählen, dehnt sich das Netzwerk etwas aus, bleibt aber insgesamt immer noch sehr klein. Wenn Sie zusätzlich das Layout „Expansion“ mehrfach anwenden, nimmt Ihr Netzwerk langsam eine interpretierbare Form an.

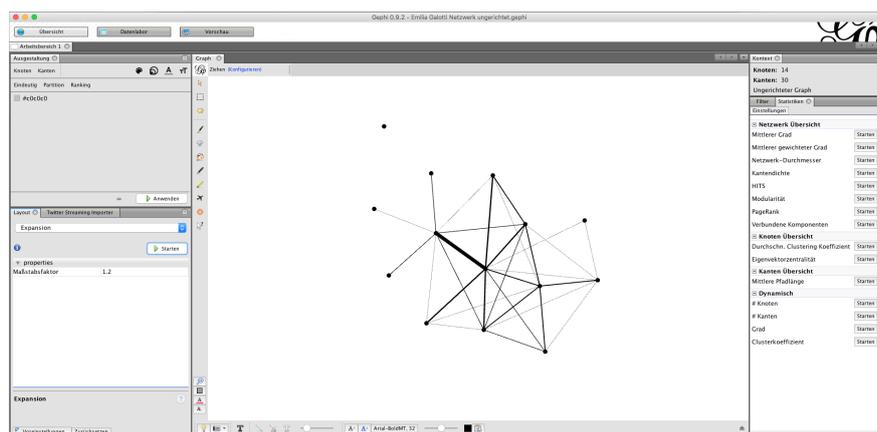


Abb. 26: Ansicht des „Übersicht“-Bereichs nach Anlegen der Datentabellen

Wenn Sie Gephi ohne weitere Plugins installieren, können Sie aus folgenden Netzwerk-Layouts wählen:

- **Bezeichner-Justierung:** Abstände zwischen Knoten werden so lange vergrößert, bis die Bezeichnungen (Label) lesbar sind. Das Layout eignet sich für jede Art von Netzwerk und kann auch zusätzlich zu anderen Layouts verwendet werden.
- **Expansion:** Das Layout zieht das Netzwerk weiter auseinander. Das Layout eignet sich für jede Art von Netzwerk und kann auch zusätzlich zu anderen Layouts verwendet werden.
- **Force Atlas und Force Atlas 2:** Folgen einem zweifach linearen Modell, bei dem Anziehung und Abstoßung der Knoten proportional zu ihrem Abstand berechnet werden. Das Layout kann etwa 10-10.000 Knoten verarbeiten und eignet sich besonders um mittelgroße Netzwerke wie z. B. soziale Netzwerke (Daten aus Facebook oder Twitter) zu visualisieren.
- **Fruchterman Reingold:** Diesem Layout liegt (ähnlich wie dem Force Atlas) die Idee zugrunde, dass zwischen Knoten Anziehungskräfte bestehen (wie bei Magneten) und die Kanten die Knoten auseinander drücken (wie Sprungfedern). Das Layout entwickelt sich so lange weiter, bis ein Äquilibrium zwischen Anziehungs- und Abstoßungskräften entstanden ist. Das Layout kann 1-1000 Knoten verarbeiten und eignet sich darum besonders für kleine Netzwerke (wie das Figuren-Netzwerk aus Emilia Galotti).
- **Isometric Layout:** Das Layout erlaubt es, eine weitere Dimension in die Visualisierung einzubeziehen. Bisher wurden hier zwei Dimensionen betrachtet - Knoten und Kanten, die in einem zweidimensionalen Feld organisiert wurden. Beim isometrischen Layout wird eine sogenannten z-Dimension (wie in einem Achsensystem, in dem es eine x-, eine y- und eine z-Achse gibt) erzeugt. Diese können Sie im „Datenlabor“ auch bearbeiten (automatisch ist für jeden Knoten zunächst der Wert 0 eingetragen). Hier können Sie z. B. unterschiedliche Werte für Haupt- und Nebencharaktere vergeben. In der grafischen Darstellung wird diese dritte Achse dann allerdings wieder in das zweidimensionale Feld eingerechnet. Beim Isometric Layout ist es möglich, das Netzwerk anhand der z-Achse zu spiegeln. Das Layout eignet sich vor allem für Netzwerke, in denen drei Dimensionen gleichwertig oder nahezu gleichwertig nebeneinander stehen. Bei einem Kopräsenz-Netzwerk ist der Status der Figuren aber meist weniger bedeutsam als ihr Name/ihre Funktion und ihre Relation zueinander.
- **Network Splitter 3D:** Ähnlich wie beim isometrischen Layout wird auch hier eine dritte Ebene in die Berechnung des Netzwerkes einbezogen. Dadurch lässt sich das Netzwerk in einzelne Cluster aufsplitten. Wenn ausschließlich dieses Layout genutzt wird, muss die Z-Dimension im Datenlabor definiert und mit Werten versehen werden. Wenn Sie zuerst ein anderes Layout nutzen (z. B. Fruchterman Reingold), erstellt Gephi automatisch aus den Ergebnissen der vorigen Berechnung eine Z-Dimension. Durch die

Kombination mit anderen Layouts können beim Figurennetzwerk z. B. Cluster der Zentralität erstellt werden.

- *Noverlap*: Das Layout sorgt dafür, dass keine Überlappungen von Knoten, Kanten und Labels entstehen. Das Layout kann zusätzlich zu anderen Layouts verwendet werden.
- *OpenOrd*: Das Layout basiert auf der Idee des Fruchterman Reingold Layouts. Allerdings werden hier noch weitere Parameter (z. B. Sieden, Abkühlen) in die Simulation einbezogen. Damit sich eindeutige Cluster bilden, werden sehr lange Kanten abgeschnitten, und in der Visualisierung dann von den anderen entfernt dargestellt. Das Layout eignet sich besonders für mittelgroße bis große Netzwerke, da es eine Datenbasis von 100-1.000.000 Knoten benötigt, um gute Ergebnisse zu liefern.
- *Rotate*: Das Netzwerk rotiert um eine Achse. Das Layout kann zusätzlich zu anderen Layouts verwendet werden.
- *Yifan Hu und Yifan Hu Proportional*: Diese Layouts kombinieren die Methodik der forcierten Ausrichtung mit einer Graph-Vereinfachungs-Technik, um die Komplexität des Netzwerkes zu reduzieren. Dadurch werden Cluster besonders deutlich. Dieses Layout eignet sich besonders für Netzwerke von 100-100.000 Knoten, wobei es besonders bei größeren Netzwerken sinnvoll einsetzbar ist.

Außerdem können Sie in Gephi ein Zufallslayout wählen oder über die Funktion „Zusammenziehen“ das Netzwerk verdichten. Weitere Netzwerk-Layouts können über Plugins aus der Gephi-Entwickler-Community installiert werden.

Aufgabe 2: Wählen Sie die Farbe der Knoten als genderspezifische Visualisierung aus. Lassen Sie sich die Knotengröße entsprechend der Anzahl gesprochener Wörter anzeigen. Wiederholen Sie die Arbeitsschritte für die Kanten. Legen Sie alle Größen und Farben für Ihr Netzwerk fest. Was passiert, wenn Sie Größen und Farben der Label ändern?

Mit der Anpassung der Größe und Farbe der Labels verändert sich der im Netzwerk gezeigte Text. Ihre Knoten haben in diesem Netzwerk Label. Das bedeutet: Der Text verändert sich, sobald Sie die Knoten-Label bearbeiten. Allerdings ist die Anzeige der Labels per **Default** ausgeschaltet, wenn Sie die Einstellungen zum jetzigen Zeitpunkt ändern, passiert erst einmal gar nichts. Ihre Kanten haben in diesem Netzwerk keine Label. Wenn Sie die Einstellungen für die Kanten-Label ändern, verändert sich das Netzwerk gar nicht.

Ihr Netzwerk könnte nach der Farb- und Größenanpassung etwa so aussehen:

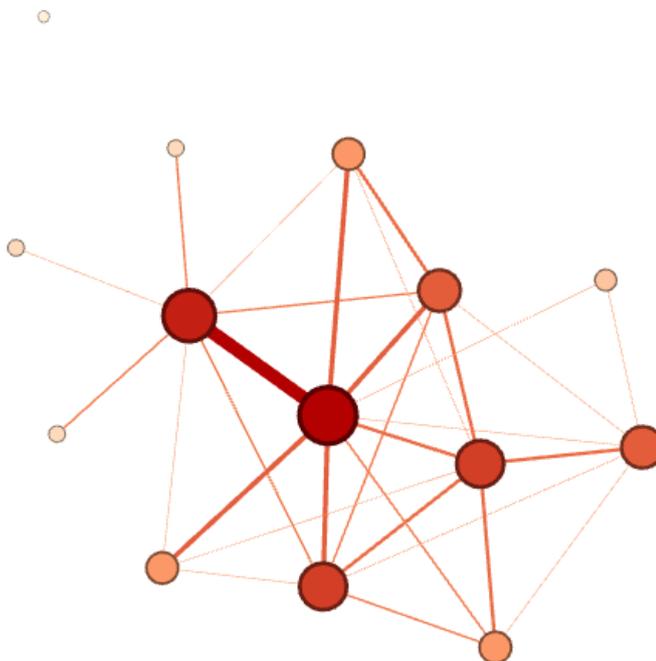


Abb. 27: Ein in Farbe und Größe nach Grad ausgerichtetes Netzwerk

Aufgabe 3: Legen Sie Ihr Netzwerk so an, dass Marinelli im Mittelpunkt steht. Welches ist der kürzeste Pfad zwischen Marinelli und Conti (zwei Figuren, die nie kopräsent sind)? Legen Sie einen Knoten „weitere Bediente“ an. Fügen Sie jeweils eine Verbindung zwischen „weitere Bediente“, Pirro und Battista hinzu. Benennen Sie diese Verbindung als „sind“ (Tipp: Sie müssen dazu zurück in das Datenlabor gehen). Lassen Sie sich die Benennung der Kante im Graphen anzeigen. Wenn Ihnen Ihr Netzwerk gut gefällt, machen Sie einen Screenshot davon.

Nutzen Sie dazu nicht die Funktion Ihres Rechners, sondern die Gephi-eigene Screenshot-Funktion. Der kürzeste Pfad zwischen Marinelli und Conti verläuft über den Prinzen.

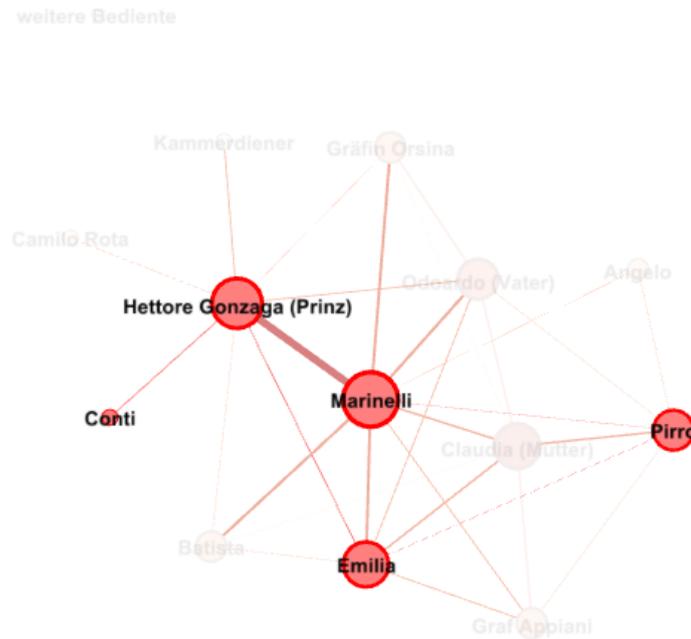


Abb. 28: Anzeige des kürzesten Pfades zwischen Marinelli und Conti

Wenn Sie einen Knoten „weitere Bediente“ hinzufügen, so erscheint dieser auch in Ihrer Datentabelle im „Datenlabor“. Fügen Sie dazu noch Verbindungen zwischen „weitere Bediente“ und Pirro und Battista im Bereich „Übersicht“ hinzu, so erscheinen zwei neue Zeilen in Ihrer Kanten-Tabelle. Hier können Sie in der Spalte Label die Bezeichnung „sind“ eingeben. Wenn Sie zurück in den Bereich „Übersicht“ und dort auf die Konfiguration der Bezeichnung gehen, können Sie für Kanten „Label“ auswählen.

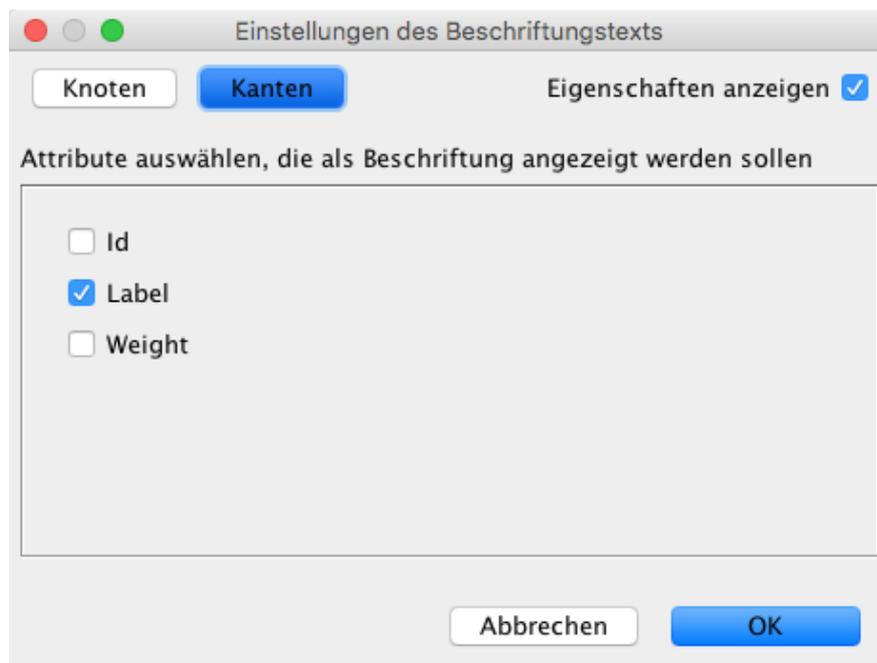


Abb. 29: Fenster zur Auswahl der Kanten-Beschriftung

Nun wird Ihnen in Ihrem Netzwerk der Wert „sind“ neben der entsprechenden Kante angezeigt (vgl. Abb. 36). Sollte dies nicht der Fall sein, müssen Sie im mittleren Menübereich bei „Kanten“ noch das blaue Häkchen setzen:



Abb. 30: Nur, wenn das blaue Häkchen bei „Kanten“ gesetzt ist, wird eine Beschriftung angezeigt

Ihr Netzwerk könnte inzwischen etwa so aussehen:

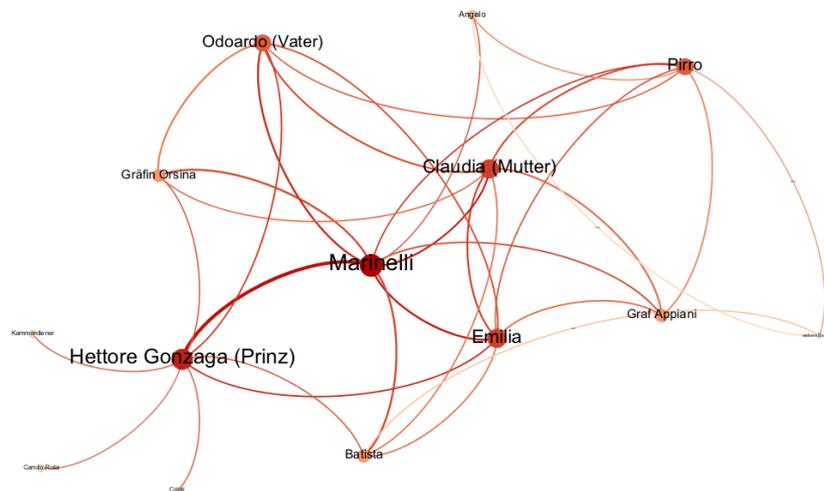


Abb. 31: Netzwerk mit Knoten- und Kantenbeschriftungen

Aufgabe 4: Setzen Sie ein Häkchen bei der Knotenbeschriftung unter „Beschriftung anzeigen“. Ändern Sie die Schriftgröße der Knotenbeschriftung auf mindestens 10 Punkt. Legen Sie die Größeneinstellung der dünnsten Kanten auf 6.0 Punkt und die der dicksten Kanten auf 15.0 fest. Sorgen Sie dafür, dass die Beschriftungen der Knoten sich nicht überlappen (Achtung, damit greifen Sie in die Netzwerkdaten ein. Sie müssen also zurück in den Bereich „Übersicht“). Exportieren Sie Ihr Netzwerk über die Gephi-eigene Exportfunktion und vergleichen Sie das Bild mit dem zuvor erstellten Screenshot aus dem Bereich „Übersicht“.

Der Screenshot Ihres Netzwerkes wirkt sehr technisch und wie eine Rohversion des Netzwerkes. Die Qualität der Auflösung ist nicht zufriedenstellend. Im heruntergeladenen Bild ist die Schrift besser lesbar und die Kanten wirken klarer definiert. Die Auflösung ist höher. Dafür sind die Ausdehnungen des heruntergeladenen Bildes nach den äußeren Knoten (nicht nach deren Beschriftung) berechnet. Haben Sie sehr lange Knoten-Beschriftungen, kann es sein, dass diese am Rand abgeschnitten werden. Es kann darum besser sein, über die Funktion Ihres eigenen PCs einen Screenshot der Netzwerkvisualisierung im Bereich „Vorschau“ zu machen und diese Grafik für Ihre literarische Textanalyse weiter zu verwenden. Hier sehen Sie den Screenshot über die Screenshot-Funktion im Bereich „Vorschau“:

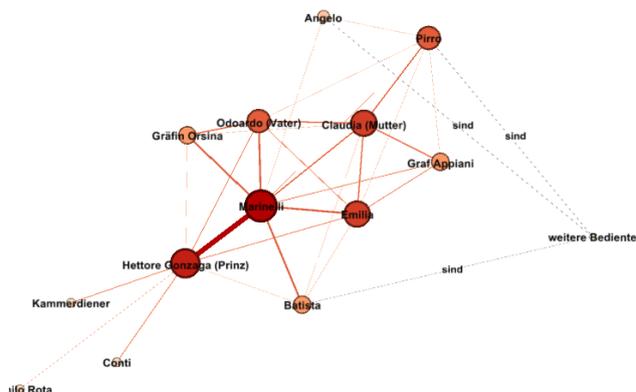


Abb. 32: Download des Netzwerkes im Bereich „Vorschau“ (kein Screenshot)

Aufgabe 5: Öffnen Sie die heruntergeladene Datei in Gephi. Schauen Sie sich die Datentabellen an. Welche Unterschiede im Vergleich zum ersten Netzwerk fallen Ihnen auf? Gestalten Sie dann in den Bereichen „Übersicht“ und „Vorschau“ ein Netzwerk, das in Ihren Augen die Erwähnungen in der Figurenrede gut veranschaulicht. Vergleichen Sie Ihre Visualisierung mit der Grafik des ersten Netzwerkes. Welche Ähnlichkeiten und Unterschiede fallen Ihnen auf?

Die heruntergeladene Datei verzeichnet für die Knotentabelle lediglich eine automatisch vergebene ID in Form einer Zahl für jede Figur und das Label. Es sind keine zusätzlichen Angaben wie Geschlecht oder Anzahl gesprochener Wörter enthalten. Es handelt sich um eine in Gephi selbst angelegte Tabelle, die nur die absolut notwendigen Daten zur Netzwerkerstellung enthält. Auf der Kantentabelle sind Ursprung und Ziel jeweils in Form der Zahlen-IDs angegeben. Es handelt sich um ein gerichtetes Netzwerk, dessen Kanten das Attribut „spricht von“ zugewiesen wurde. Ebenso gut hätte dieses Attribut mit „erwähnt“ bezeichnet werden können. Die Zahlenwerte in der „weight“-Spalte sind wesentlich höher als diejenigen im Kopräsenz-Netzwerk.

Eine mögliche fertige Netzwerkvisualisierung der Erwähnungen von Figuren in der Figurenrede könnte so aussehen:

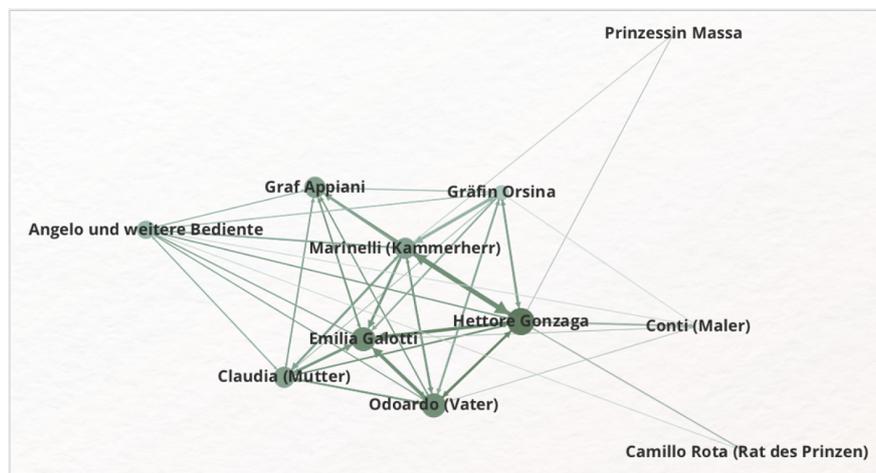


Abb. 33: Netzwerkvisualisierung der in der Figurenrede erwähnten Figuren mit folgenden Parametern: Knotengröße: Anzahl der Erwähnungen in der Figurenrede anderer Figuren (Eingangsgrad), Intensität der Knotenfarbe: Eingangsgrad und Anzahl der Erwähnungen anderer Figuren (Ausgangsgrad), Dicke und Farbintensität der Kanten: Gewicht, Layout: Force Atlas 2

Hier fällt zunächst auf, dass Marinelli weniger oft erwähnt oder angesprochen wird als andere Figuren, obwohl er im Verlauf des Dramas mit sehr vielen Figuren kopräsent ist. Emilia hat einen höheren Eingangsgrad, der jedoch geringer ist als der des Prinzen. Am häufigsten wird Hettore Gonzaga in der Figurenrede erwähnt. Die Bedienten spielen in der Figurenrede keine Rolle, sprechen selbst aber andere Figuren an oder sprechen über sie. Die Prinzessin Massa, die niemals mit einer der Figuren kopräsent ist, wird erwähnt (wenn auch nicht sehr häufig). Das Force Atlas Layout, das in dieser Visualisierung kaum manuell nachbearbeitet wurde, zeigt hier das enge Kräfte-Gefüge zwischen Emilia und ihrer Familie auf der einen und dem Prinzen und seinem Kammerdiener auf der anderen Seite. Figuren, die nicht so stark in dieses Gefüge integriert sind, werden vom Kern des Netzwerks entfernt dargestellt. Dazu gehören die Bedienten, Camillo Rota, Conti und Prinzessin Massa. Insgesamt sind die Abweichungen zwischen den beiden Netzwerken aber nicht so stark, wie man es vielleicht vermutet hätte. Trotzdem wird im Vergleich der beiden Netzwerke deutlich, dass sie jeweils einen interpretativen Blick auf das Drama ermöglichen. Dieser ist bedingt durch die Aufbereitung der Daten auf der einen Seite und die Anpassung des Netzwerkes auf der anderen Seite. Beide Netzwerke zeigen einen Überblick über die Figurenkonstellation und illustrieren bestimmte, vorher klar definierte Aspekte eines Werkes. Damit deutlich wird, dass eine Netzwerk-Visualisierung stets in einem interpretativen Vorgang der Datenaufbereitung entstanden ist, ist es sehr wichtig, bei jeder Netzwerk-Grafik die Parameter (z.B. Bedeutung von Knoten und Kanten, ihrer Farbgebung und Größe) mit anzugeben.

Externe und weiterführende Links

- Materialien zu dieser Lerneinheit (Emilia Galotti Annotationen, Knoten und Kanten): <https://web.archive.org/https://zenodo.org/records/10519600> (Letzter Zugriff: 14.08.2024)
- DraCor: <https://web.archive.org/save/https://dracor.org/> (Letzter Zugriff: 14.08.2024)
- Gephi: <https://web.archive.org/save/https://gephi.org/users/download/> (Letzter Zugriff: 14.08.2024)

Bibliographie

- forTEXT. 2019a. Tutorial: Sicherheitsaufnahme für Internetprogramme Hinzufügen (Mac). 19. Januar. <https://doi.org/10.5281/zenodo.11074232>.
- . 2019b. Tutorial: Sicherheitsausnahme für Internetprogramme Hinzufügen (Windows). 25. Januar. <https://doi.org/10.5281/zenodo.11074222>.
- forTEXT. 2020. Netzwerkanalyse mit Gephi. 4. Dezember. doi: 10.5281/zenodo.10519599, <https://zenodo.org/records/10519600>.
- Horstmann, Jan. 2024a. Toolbeitrag: Stylo. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 1. Stilometrie (26. Februar). doi: 10.48694/fortext.3770, <https://fortext.net/tools/tools/stylo>.
- . 2024b. Ressourcenbeitrag: DraCor - Drama Corpora Project. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3785, <https://fortext.net/ressourcen/textsammlungen/dracor-drama-corpora-project>.
- Horstmann, Jan und Jan-Erik Stange. 2024. Methodenbeitrag: Textvisualisierung. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 5. Textvisualisierung (7. August). doi: 10.48694/fortext.3772, <https://fortext.net/routinen/methoden/textvisualisierung>.
- Schumacher, Mareike. 2024a. Toolbeitrag: CATMA. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 4. Manuelle Annotation (7. August). doi: 10.48694/fortext.3761, <https://fortext.net/tools/tools/catma>.
- . 2024b. Methodenbeitrag: Netzwerkanalyse. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3759, <https://fortext.net/routinen/methoden/netzwerkanalyse>.
- . 2024c. Toolbeitrag: Gephi. Hg. von Evelyn Gius. *forTEXT* 1, Nr. 6. Netzwerkanalyse (30. August). doi: 10.48694/fortext.3783, <https://fortext.net/tools/tools/gephi>.

Glossar

Annotation Annotation beschreibt die manuelle oder automatische Hinzufügung von Zusatzinformationen zu einem Text. Die manuelle Annotation wird händisch durchgeführt, während die (teil-)automatisierte Annotation durch **Machine-Learning-Verfahren** durchgeführt wird. Ein klassisches Beispiel ist das automatisierte **PoS-Tagging** (Part-of-Speech-Tagging), welches oftmals als Grundlage (**Preprocessing**) für weitere Analysen wie Named Entity Recognition (NER) nötig ist. Annotationen können zudem deskriptiv oder analytisch sein.

Default Das/der Default (engl. für Voreinstellung oder Standardwert) bezeichnet den Wert einer Softwareeinstellung oder einer Eingabevariable, der verwendet wird, falls Nutzer*innen selbst keinen Wert oder keine Einstellungen vornehmen. Es handelt sich also um die standardmäßig festgelegten Einstellungen eines Tools oder Programms. Per Default festgelegte Parameter lassen sich i. d. R. manuell umstellen.

- Kollokation** Als Kollokation bezeichnet man das häufige, gemeinsame Auftreten von Wörtern oder Wortpaaren in einem vordefinierten Textabschnitt.
- Lemmatisieren** Die Lemmatisierung von Textdaten gehört zu den wichtigen **Preprocessing**-Schritten in der Textverarbeitung. Dabei werden alle Wörter (**Token**) eines Textes auf ihre Grundform zurückgeführt. So werden beispielsweise Flexionsformen wie „schneller“ und „schnelle“ dem Lemma „schnell“ zugeordnet.
- Machine Learning** Machine Learning, bzw. maschinelles Lernen im Deutschen, ist ein Teilbereich der künstlichen Intelligenz. Auf Grundlage möglichst vieler (Text-)Daten erkennt und erlernt ein Computer die häufig sehr komplexen Muster und Gesetzmäßigkeiten bestimmter Phänomene. Daraufhin können die aus den Daten gewonnen Erkenntnisse verallgemeinert werden und für neue Problemlösungen oder für die Analyse von bisher unbekanntem Daten verwendet werden.
- Named Entities** Eine Named Entity (NE) ist eine Entität, oft ein Eigenname, die meist in Form einer Nominalphrase zu identifizieren ist. Named Entities können beispielsweise Personen wie „Nils Holgerson“, Organisationen wie „WHO“ oder Orte wie „New York“ sein. Named Entities können durch das Verfahren der Named Entity Recognition (NER) automatisiert ermittelt werden.
- POS** PoS steht für *Part of Speech*, oder „Wortart“ auf Deutsch. Das PoS- **Tagging** beschreibt die (automatische) Erfassung und Kennzeichnung von Wortarten in einem Text und ist ein wichtiger **Preprocessing**-Schritt, beispielsweise für die Analyse von **Named Entities**.
- Preprocessing** Für viele digitale Methoden müssen die zu analysierenden Texte vorab „bereinigt“ oder „vorbereitet“ werden. Für statistische Zwecke werden Texte bspw. häufig in gleich große Segmente unterteilt (*chunking*), Großbuchstaben werden in Kleinbuchstaben verwandelt oder Wörter werden **lemmatisiert**.
- Query** *Query* bedeutet „Abfrage“ oder „Frage“ und bezeichnet eine computergestützte Abfrage zur Analyse eines Textes. Um Datenbestände zu durchsuchen, werden Abfragesprachen eingesetzt, die *Queries* (Anfragen) an den Datenbestand senden. So bilden alle möglichen *Queries* zusammen die *Query Language* eines Tools.
- Tagset** Ein Tagset definiert die Taxonomie, anhand derer **Annotationen** in einem Projekt erstellt werden. Ein Tagset beinhaltet immer mehrere Tags und ggf. auch Subtags. Ähnlich der **Type/Token**-Differenz in der Linguistik sind Tags deskriptive Kategorien, wohingegen Annotationen die einzelnen Vorkommnisse dieser Kategorien im Text sind.
- Type/Token** Das Begriffspaar „Type/Token“ wird grundsätzlich zur Unterscheidung von einzelnen Vorkommnissen (Token) und Typen (Types) von Wörtern oder Äußerungen in Texten genutzt. Ein Token ist also ein konkretes Exemplar eines bestimmten Typs, während ein Typ eine im Prinzip unbegrenzte Menge von Exemplaren (Token) umfasst.
Es gibt allerdings etwas divergierende Definitionen zur Type-Token-Unterscheidung. Eine präzise Definition ist daher immer erstrebenswert. Der Satz „Ein Bär ist ein Bär.“ beinhaltet beispielsweise fünf Worttoken („Ein“, „Bär“, „ist“, „ein“, „Bär“) und drei Types, nämlich: „ein“, „Bär“, „ist“. Allerdings könnten auch vier Types, „Ein“, „ein“, „Bär“ und „ist“, als solche identifiziert werden, wenn Großbuchstaben beachtet werden.