

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



# Graphdatenbanken für Historiker Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi

Andreas Kuczera

## Inhaltsverzeichnis

1 Einleitung.....	1
2 Graphdatenbanken für Historiker.....	2
3 Das kumulierte Register der Regesten Kaiser Friedrichs III. als Graphdatenbank.....	4
3.1 Dateneingabe.....	5
3.1.1 Datenaufbereitung.....	5
3.1.2 Die Lemmaliste.....	6
3.1.3 Die Regestenliste.....	8
3.1.4 Die Liste der Edges.....	9
3.1.5 ist_empfaenger.....	9
3.1.6 is_mother_off.....	9
3.1.7 Dateneingabe.....	10
3.2 Datenausgabe.....	12
3.2.1 Neo4j-Webfrontend.....	13
3.2.2 Die Abfragesprache cypher.....	15
3.2.3 Visualisierungsoptionen im neo4j-Webfrontend.....	16
3.2.4 Export von Subgraphen zur Weiterbearbeitung.....	19
4 Visualisierung von Netzwerken in Gephi.....	19
4.1 Das Programm Gephi.....	19
4.2 Import und Aufbereitung der Daten.....	19
4.3 Visualisierungsmöglichkeiten.....	20
4.4 Auswahl des Layout: ForceAtlas2.....	21
4.5 Lemmata und Regesten farbig darstellen.....	22
4.6 Degree Range.....	24
4.7 Timeline.....	26
5 Zusammenfassung.....	28

## 1 Einleitung

Zu Beginn der Digitalisierung von historischen Quellen stand die Imagedigitalisierung im Vordergrund. Sie hatte den Vorteil, dass sie unter den gegebenen technischen Rahmenbedingungen bezahlbar und mit vertretbarem Zeitaufwand durchzuführen war. Neben dem Image, also dem Bild der Quelle, wurden noch die sog. Metadaten erfasst, in denen z.B. Eigenschaften und Hinweise zur inhaltlichen Erschließung des Dokuments computerlesbar abgelegt sind<sup>1</sup>.

<sup>1</sup> Die Imagedigitalisierung einer Edition findet man z.B. bei den dmgh ([www.dmgh.de](http://www.dmgh.de)). Neben den üblichen Metadaten werden bei den dmgh sogar die Koordinaten der einzelnen Worte und Buchstaben auf den Seiten mit erfasst und mit dem ebenso erstellten Volltext verknüpft. Einen Überblick zu Publikationen zu den dmgh bietet die

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



Die nächste Stufe der Digitalisierung stellte die Volltextdigitalisierung da. Exemplarisch kann hier auf die Digitalisierung der *Regesta Imperii* seit 2001 verwiesen werden<sup>2</sup>. Neben den Imagedigitalisaten aller bis 2001 gedruckt vorliegenden Bände sind die Volltexte aller Regesten über die Internetseite [www.regest-imperii.de](http://www.regest-imperii.de) recherchierbar.

Im Verlauf der letzten 10 Jahre hat die Menge an digital verfügbaren Volltexten stark zugenommen. Damit einher geht auch eine Veränderung sowohl der Nutzungsformen digitaler Quellen als auch der Möglichkeiten der historischen Arbeitsweise. Bestand um die Jahrtausendwende noch enger Kontakt zwischen Historiker und Quelle, nimmt dies mit zunehmender Digitalisierung perspektivisch ab. Hat der Forscher früher die für seine Forschungsfragen relevanten Quellen in der Regel alle mindestens einmal gelesen, scheint dies bei den heute recherchierbaren Mengen an digitalen Quellen kaum noch möglich. Ein Hauptproblem ergibt sich hier aus der Schnittstelle zwischen Forscher und den im Netz erreichbaren Quellendatenbanken. Die Suchinterfaces der Datenbanken sind oft für die Nutzung durch Experten des jeweiligen Materials optimiert<sup>3</sup>. Dies ist auf der einen Seite zu begrüßen, da sie den Experten damit besten Zugriff auf das Material gewähren. Daneben sollten aber weitere Zugriffsmöglichkeiten für übergreifende Text-Mining- oder Big-Data-Recherchen bereitgestellt werden, mit denen verschiedene Quellenkorpora parallel im Hinblick auf übergreifende Fragestellungen untersucht werden können.

Momentan werden digitale Quellen meist in XML oder in SQL-Datenbanken abgelegt. XML hat sich als Standard bewährt und findet in vielen Editionsprojekten als Datenformat Verwendung während Websites meist auf SQL-Datenbanken als Daten-Repositories zurückgreifen. XML-Dateien sind meist noch verständlich lesbar, bei SQL-Datenbanken ist die Lesbarkeit ohne Kenntnis der zu Grunde liegenden Datenstrukturen in der Regel nicht mehr gegeben. Hier könnte die Verwendung von Graphdatenbanken ein neuer Ansatz für die Speicherung von *erschließendem Wissen* sein.

## 2 Graphdatenbanken für Historiker

In SQL-Datenbanken sind die Informationen in Tabellen abgelegt, die untereinander verknüpft sind. Graphdatenbanken folgen hier einem völlig anderen Ansatz. In einem Graph gibt es Knoten und Kanten. Vergleicht man die Knoten mit einem Eintrag in einer Tabelle einer SQL-Datenbank, wäre eine Kante eine Verknüpfung zwischen zwei Tabelleneinträgen. Im Unterschied zu SQL-Datenbanken können Knoten und Kanten jeweils Eigenschaften haben.

Seite <http://www.mgh.de/dmgh/pub/> (abgerufen am 4.2.1015).

- 2 Zur Digitalisierung der RI vgl. [Dieter Rübsamen](#), [Andreas Kuczera](#): *Verborgene, vergessene, verlorene? Perspektiven der Quellenserschließung durch die digitalen Regesta Imperii*. In: Rainer Hering, Jürgen Sarnowsky, Christoph Schäfer und Udo Schäfer [Hrsg.]: *Forschung in der digitalen Welt. Sicherung, Erschließung und Aufbereitung von Wissensbeständen. Tagung des Staatsarchivs Hamburg und des Zentrums „Geisteswissenschaften in der digitalen Welt“ an der Universität Hamburg am 10. und 11. April 2006*. Hamburg 2006, S. 109–123 (=Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg, Bd. 20). ([http://hup.sub.uni-hamburg.de/opus/volltexte/2008/77/pdf/HamburgUP\\_HeringEtAl\\_Forschung.pdf](http://hup.sub.uni-hamburg.de/opus/volltexte/2008/77/pdf/HamburgUP_HeringEtAl_Forschung.pdf) PDF; 13,1 MB abgerufen am 4.2.2015).
- 3 Die Problematik wird in diesem Beitrag näher untersucht: Andreas Kuczera: *Digitale Perspektiven mediävistischer Quellenrecherche*, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 18. April 2014, <http://mittelalter.hypotheses.org/3492> (ISSN 2197-6120) (abgerufen am 4.2.2015).

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



Beispielhaft werden hier die Informationen des kumulierten Registers der Regesten Kaiser Friedrichs III. aufbereitet<sup>4</sup>.

Das folgende Beispiel demonstriert an Hand der mit Regest Nr. 189 aus Heft 19 der Regesten Kaiser Friedrichs III. verknüpften Registereinträge einen Graph mit neun Knoten (künftig als *Nodes* bezeichnet) und zehn Kanten (künftig als *Edges* bezeichnet).

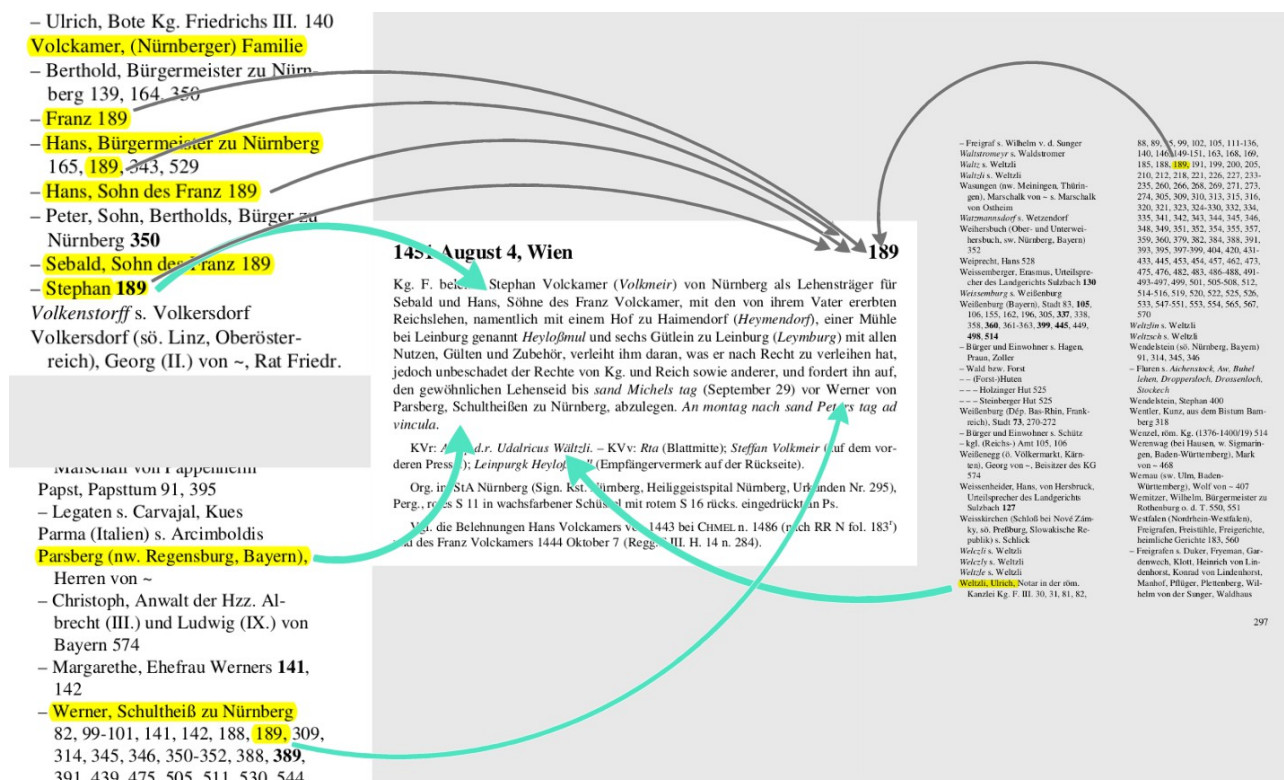


Abbildung 1: Registereinträge, die in Regest Nr. 189 vorkommen, aus: *Regesten Kaiser Friedrichs III. (1440-1493). Nach Archiven und Bibliotheken geordnet. H. 19: Die Urkunden und Briefe aus Archiven und Bibliotheken der Stadt Nürnberg. Teil 2: 1450-1455Rübsamen, Dieter [Bearb.]. - Wien [u.a.] (2004), S. 106, 287, 296 und 297.*

In dem Graphen werden die Registereinträge und das Regest als Nodes dargestellt und die Verknüpfungen als Edges. Vergleicht man diese Datenstrukturen mit jenen einer SQL-Datenbank, fällt vor allem auf, dass sowohl Nodes als auch Edges Eigenschaften (sog. *Properties*) haben können. In der folgenden Abbildung werden also drei verschiedene Typen von Nodes (Registereintrag, Regest und Ort) und zwei verschiedene Typen von Edges (*GENANNT\_IN*, *IST\_BUERGERMEISTER\_ZU* und *IST\_SOHN\_VON*) dargestellt und symbolisieren die verschiedenen Verknüpfungen zwischen den Knoten.

<sup>4</sup> Das kumulierte Register der Regesten Kaiser Friedrichs III. finden sie unter <http://www.regesta-imperii.de/regesten/register.html> (abgerufen am 4.2.2015).

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).

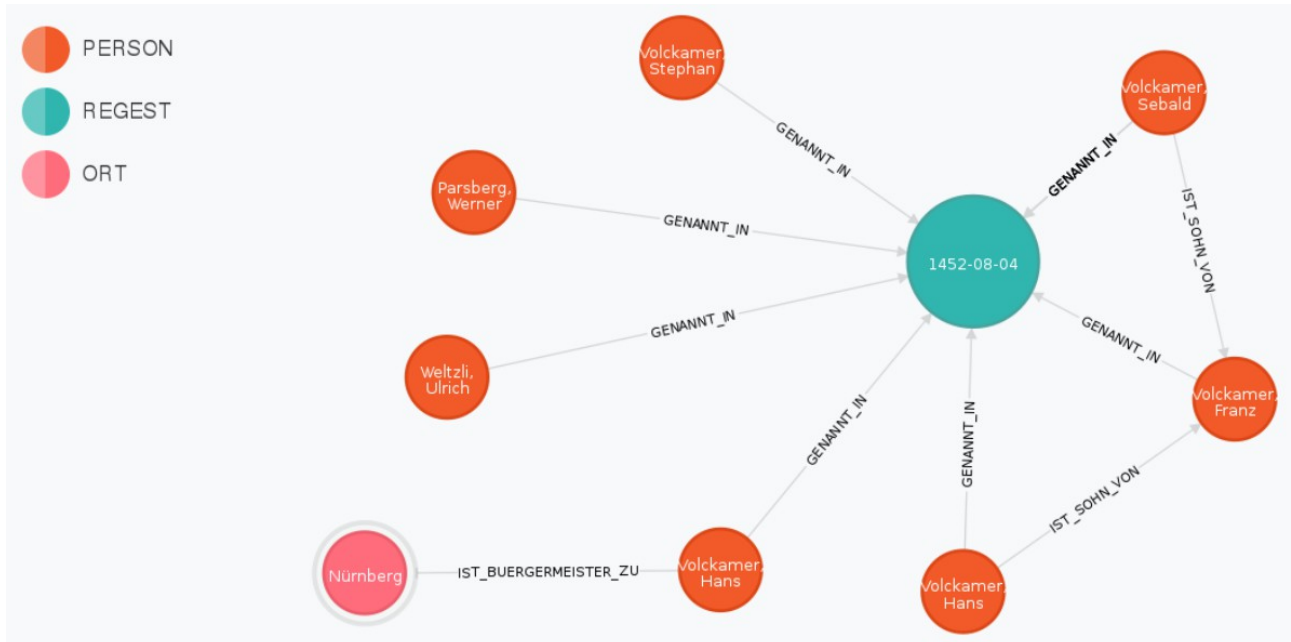


Abbildung 2: Graph zu den Registereinträgen zu Regest Nr. 189.

Neben den offensichtlich aus dem Register übernommenen Verknüpfungen, die sich direkt aus der Hierarchie und den Angaben unter den jeweiligen Registereinträgen ergeben (wie z.B. *GENANNT\_IN*) können auch weitere Verknüpfungen, wie z.B. die Amtszuweisung *IST\_BUERGERMEISTER\_ZU* oder das Verwandtschaftsverhältnis *IST\_SOHN\_VON* dargestellt und computerlesbar abgespeichert werden. Damit bieten Graphdatenbanken sehr flexible Möglichkeiten, z.B. Erschließungsinformationen zu historischen Quellen computerimplementiert zu erfassen. Gleichzeitig bieten Graphdatenbanken durch die dem menschlichen Denken eng verwandte beziehungsorientierte Ablage von Wissen verbunden mit entsprechenden Visualisierungsmöglichkeiten optimale Unterstützung bei der historischen Forschung an digitalen Quellenbeständen.

Im folgenden wird exemplarisch der Import des kumulierten Registers der Regesten Kaiser Friedrichs III. in die Graphdatenbank Neo4j durchgespielt und anschließend einige Anwendungsbeispiele dargestellt.

### 3 Das kumulierte Register der Regesten Kaiser Friedrichs III. als Graphdatenbank

Ausgangspunkt der Arbeiten für diesen Text war die Aufbereitung des kumulierten Registers der Regesten Kaiser Friedrichs III. für die Bereitstellung im Internet. Ein Register bietet auf der einen Seite Registereinträge (Personen und Orte) und auf der anderen Seite Regestnummern, die den jeweiligen Registereinträgen zugeordnet sind. Zunächst gingen die Überlegungen dahin, durch die gemeinsame Nennung von zwei Personen in einem Regest eine direkte Verbindung zu postulieren,

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



der man dann als Eigenschaften noch die jeweiligen Angaben des Regests zuweisen kann<sup>5</sup>. Im Laufe der Arbeiten erwies es sich aber als praktikabler, die Regesten selbst als Knoten mit in die Datenbank aufzunehmen. Im folgenden werden die verschiedenen Schritte zur Datenaufbereitung, zum Einspielen und zum Auswerten des Registers dargestellt.

### 3.1 Dateneingabe

#### 3.1.1 Datenaufbereitung

In der folgenden Abbildung wird ein Ausschnitt des kumulierten XML-Registers der Regesten Kaiser Friedrichs III. zum Eintrag der Nürnberger Familie Volckamer dargestellt. Zur Bearbeitung wurde der XML-Editor Oxygen mit einer für die Bearbeitung des Registers angepassten Oberfläche verwendet.

Volckamer, (Nürnberger) Familie

- Agnes, Frau des Hans #0-5172, #0-6841
- Apollonia, Tochter des Niklas #0-8654, #0-8689
- Berthold, Bürgermeister zu Nürnberg #0-535, #0-3039, #CA-56, #7-58, #14-91, #14-92, #14-107, #14-109, #14-123, #14-130, #14-196, #14-197, #14-198, #14-199, #14-200, #14-209, #14-211, #14-212, #14-213, #14-214, #14-215, #14-216, #14-217, #14-218, #14-332, #14-333, #14-490, #19-139, #19-164, #19-350
- Franz, Bürger zu Nürnberg, Vater von Hans und Sebald #14-284, #19-189, #28-384
- Hans, Bürgermeister zu Nürnberg #0-5172, #0-6841, #0-6850, #14-284, #19-165, #19-189, #19-343, #19-529; seine Frau s. [Agnes](#)
  
- Hans #0-1486
- Hans, Bruder Leonhards #0-5123, #0-6455, #0-6715
- Hans, Sohn des Franz, Bruder Sebalds #19-189, #28-138, #28-139, #28-140, #28-141, #28-163, #28-165, #28-179, #28-183, #28-384
- Kunigunde, Witwe des Franz #28-388
- Leonhard #0-5123, #0-6715
- Niklas #0-8654, #0-8689
- Paul #0-6840
- Peter, Sohn Bertholds, Bürger zu Nürnberg #0-3039, #0-7957, #0-8654, #0-8656, #0-8689, #19-350
- Sebald, Sohn des Franz, Bruder des Hans #0-5123, #0-6715, #19-189, #28-384, #28-388
- Stephan #0-555, #19-189
- Veronika, Tochter des Niklas #0-8654, #0-8689

Abbildung 3: Ausschnitt des kumulierten XML-Registers der Regesten Kaiser Friedrichs III. zum Eintrag der Nürnberger Familie Volckamer

Um die Informationen aus dem Register in die Graphdatenbank neo4j einspielen zu können, müssen sie zunächst in einzelne CSV-Dateien (Comma-Separated-Values) transformiert werden. Das Ziel sind einzelne CSV-Dateien, die jeweils Wertepaare enthalten.

5 Die Entstehung der Idee zur Visualisierung von Registernetzwerken beschreibt auch der Artikel Andreas Kuczera: Digitale Farbenspiele oder nützliches Werkzeug – Visualisierung von Netzwerken aus den Registern von Editions- und Regestenwerken, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 8. Januar 2015, <http://mittelalter.hypotheses.org/5089> (ISSN 2197-6120) (abgerufen am 5.2.2015).

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



```
<Stufe0 id="V00000774">
  <Inhalt>Volckamer, (Nürnberger) Familie</Inhalt>
  <Stufel id="V00000775">
    <Inhalt>Agnes, Frau des Hans <Regestnummer>#0-5172, #0-6841</Regestnummer>
  </Stufel>
  <Stufel id="V00000776">
    <Inhalt>Apollonia, Tochter des Niklas <Regestnummer>#0-8654, #0-8689</Regestnummer>
  </Stufel>
  <Stufel id="V00000777">
    <Inhalt>Berthold, Bürgermeister zu Nürnberg <Regestnummer>#0-535, #0-3039, #CA-56, #7-58,
      <RegestnummerFett>#14-107,</RegestnummerFett>
  </Stufel>
</Stufe0>
```

Abbildung 4: Ausschnitt des Registereintrages der Familie Volckamer im XML-Quelltext

Wie in der Abbildung zum XML-Quelltext zu sehen, besitzt jeder Eintrag eine eindeutige ID sowie eine Angabe zur Stufe des Registereintrages, zum Inhalt und den zugehörigen Regestnummern. Die Stufe gibt an, ob es sich um einen Untereintrag oder um ein Hauptlemma handelt.

### 3.1.2 Die Lemmaliste

In einem ersten Schritt werden alle einzelnen Registereinträge exportiert und mit einer zusätzlichen ID versehen. Da das Register hierarchisch aufgebaut ist, müssen zum Verständnis der einzelnen Registereinträge auch die ggf. vorhandenen übergeordneten Einträge hinzugenommen werden. Würde man in unserem o.a. Beispiel das erste Unterlemma unter dem Hauptlemma *Volckamer*, nämlich *Agnes, Frau des Hans* ohne weitere Informationen in die Liste übernehmen, wäre es schwierig die Person zu identifizieren. Daher werden die Informationen bei der Transformation in drei Spalten aufgeteilt. In der Spalte *name1* steht das Lemma, in der Spalte *name2* werden ggf. vorhandene Zusatzinformationen, die nach dem eigentlichen Lemma in Klammern angegeben sind, vermerkt. Dabei kann es sich z.B. um Schreibvarianten oder Lokalisierungen von Ortsnamen handeln. Schließlich werden in der Spalte *name3* alle Angaben des Lemmata selbst und der übergeordneten Lemmata, jeweils durch zwei Schrägstriche (//) getrennt, aufgelistet. In der Spalte *name3* sind also alle für ein Lemma relevanten Informationen zusammengefasst.

In der ersten Spalte erhält jedes Lemma eine eindeutige ID, auf die dann später die *Edges* verweisen. Spalte2 enthält die jeweilige XML-ID aus dem XML-Register. Hiermit können später wieder Verknüpfungen ins XML-Register erstellt werden. In Spalte 4 ist die ID des jeweiligen Muttereintrages im Register vermerkt. Über diese Angaben lässt sich das Register später auch wieder in seiner hierarchischen Struktur rekonstruieren. Anschließend folgen die Spalten zu den Namensangaben und schließlich noch eine Spalte mit dem Typ des Eintrages, hier also immer *lemma*.

### Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



<u>nodeID</u>	<u>xmlID</u>	<u>empfaenger-ID</u>	<u>top-neo4j-node-ID</u>	<u>name1</u>	<u>name2</u>	<u>name3</u>	<u>type</u>
1	A00000001	0		Aa. Johann von ~		Aa. Johann von ~	lemma
2	A00000002	1	1	Sophie von ~, Tochter Joh <sup>h</sup> anns, Bürgerin zu Köln		Aa. Johann von ~ // Sophie von ~	lemma
3	A00000003	0		<u>Aach</u>	(Fluß durch <u>Aach</u> , n. Singen, Baden-Württemberg)	<u>Aach</u> (Fluß durch <u>Aach</u> , n. Singen, Baden-Württemberg)	lemma
4	A00000004	0		<u>Aach</u>	(n. Singen, Baden-Württemberg)	<u>Aach</u> (n. Singen, Baden-Württemberg)	lemma
6	A00000006	0		Aachen	(Aache; Nordrhein-Westfalen), <sup>S</sup>	Aachen (Aache; Nordrhein-Westfalen)	lemma
8	A00000008	1	6	Fischmarkt	(Parwisch)	Aachen (Aache; Nordrhein-Westfalen)	lemma
9	A00000009	1	6	Gerichte		Aachen (Aache; Nordrhein-Westfalen)	lemma
10	A00000010	2	6	Kurgericht		Aachen (Aache; Nordrhein-Westfalen)	lemma
11	A00000011	2	6	<u>Schöffentuhl</u> , kgl.		Aachen (Aache; Nordrhein-Westfalen)	lemma
12	A00000012	3	6	Richter und Schöffen		Aachen (Aache; Nordrhein-Westfalen)	lemma
13	A00000013	3	6	Schöffen		Aachen (Aache; Nordrhein-Westfalen)	lemma
14	A00000014	1	6	„Grafschaften“		Aachen (Aache; Nordrhein-Westfalen)	lemma
15	A00000015	1	6	„Grasgebot“		Aachen (Aache; Nordrhein-Westfalen)	lemma
17	A00000017	1	6	<u>Meierei</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma
19	A00000019	2	6	Brothaus		Aachen (Aache; Nordrhein-Westfalen)	lemma
21	A00000021	2	6	<u>Gewandhaus</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma
22	A00000022	2	6	Grashaus		Aachen (Aache; Nordrhein-Westfalen)	lemma
23	A00000023	2	6	<u>Plankenhaus</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma
24	A00000024	2	6	Tuchhaus		Aachen (Aache; Nordrhein-Westfalen)	lemma
25	A00000025	2	6	Haus zum <u>Haner</u>	(zum Hanen)	Aachen (Aache; Nordrhein-Westfalen)	lemma
26	A00000026	2	6	<u>zo der Geiss</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma
27	A00000027	1	6	<u>Rentmeister</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma
29	A00000029	1	6	„Stadtbücher“		Aachen (Aache; Nordrhein-Westfalen)	lemma
30	A00000030	1	6	<u>Vogtei</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma
31	A00000031	1	6	Kirchen und Klöster		Aachen (Aache; Nordrhein-Westfalen)	lemma
32	A00000032	2	6	St. <u>Adalbertstift</u>		Aachen (Aache; Nordrhein-Westfalen)	lemma

Abbildung 5: Liste aller Registereinträge im CSV-Format

Die nächste Abbildung zeigt als direkten Vergleich den Ausschnitt zum Registereintrag der Familie Volckamer aus der Nodes-Datei.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



nodeID	xmlID	empfaenger-ID	top-neo4j-node-ID	name1	name2	name3	type
45181	V00000763	2	45159	Felix		Voitsberg (Steiermark), Stadt	lemma
45182	V00000764	3	45159	seine Kinder		Voitsberg (Steiermark), Stadt	lemma
45184	V00000766	0		Voitscher, Jörg, Verweser der Hauptmannschaft in Steyr		Voitscher, Jörg, Verweser der	lemma
45189	V00000771	0		Vokhinger, Kaspar, Bürger in Wiener Neustadt		Vokhinger, Kaspar, Bürger in	lemma
45192	V00000774	0		Volckamer,	(Nürnberger) Familie	Volckamer, (Nürnberger) Fam	lemma
45193	V00000775	1	45192	Agnes, Frau des Hans		Volckamer, (Nürnberger) Fam	lemma
45194	V00000776	1	45192	Apollonia, Tochter des Niklas		Volckamer, (Nürnberger) Fam	lemma
45195	V00000777	1	45192	Berthold, Bürgermeister zu Nürnberg		Volckamer, (Nürnberger) Fam	lemma
45196	V00000778	1	45192	Franz, Bürger zu Nürnberg		Volckamer, (Nürnberger) Fam	lemma
45198	V00000780	1	45192	Hans		Volckamer, (Nürnberger) Fam	lemma
45199	V00000781	1	45192	Hans, Bruder Leonhards		Volckamer, (Nürnberger) Fam	lemma
45200	V00000782	1	45192	Hans, Sohn des Franz		Volckamer, (Nürnberger) Fam	lemma
45201	V00000783	1	45192	Leonhard		Volckamer, (Nürnberger) Fam	lemma
45202	V00000784	1	45192	Niklas		Volckamer, (Nürnberger) Fam	lemma
45203	V00000785	1	45192	Paul		Volckamer, (Nürnberger) Fam	lemma
45204	V00000786	1	45192	Peter, Sohn, Bertholds, Bürger zu Nürnberg		Volckamer, (Nürnberger) Fam	lemma
45205	V00000787	1	45192	Sebald, Sohn des Franz		Volckamer, (Nürnberger) Fam	lemma
45206	V00000788	1	45192	Stephan		Volckamer, (Nürnberger) Fam	lemma
45207	V00000789	1	45192	Veronika, Tochter des Niklas		Volckamer, (Nürnberger) Fam	lemma

Abbildung 6: Ausschnitt der Nodes-Datei zum Registereintrag der Familie Volckamer

Zu beachten ist, dass in die Nodes-Datei keinerlei Informationen zu den zum Registereintrag gehörigen Regestennummern übernommen werden, sondern nur die Registereinträge.

### 3.1.3 Die Regestenliste

regID	registennu	datum	type
114878	#13-187	1451-01-13	regist
114879	#13-188	1451-01-25	regist
114880	#13-189	1451-01-26	regist
114881	#13-190	1451-01-29	regist
114882	#13-191	1451-01-29	regist

Abbildung 7: Auszug aus der Regestenliste

In einer zweiten Datei werden nun alle im Register vorkommenden Regestennummern angelegt. Da es sich bei dem Register um das kumulierte Register aller Bände der Regesten Kaiser Friedrichs III. handelt, kann hierfür ein Auszug aus der Regestendatenbank mit allen Regesten zu Friedrich III. genutzt werden. Im Register setzen sich die Verweisnummern auf die Regesten aus zwei Bestandteilen zusammen: #13-189

– Das Doppelkreuz (#) markiert die Angabe zum jeweiligen Band und die mit dem Minus anschließende Nummer gibt die Regestennummer an. Analog werden die Regesten in der Regestenliste aufgelistet.

In der Abbildung sehen Sie einen Auszug aus der Regestenliste. In der ersten Spalte wird jedem Regest eine ID gegeben, die später für die Verknüpfung zu den Registereinträgen genutzt wird. Die zweite Spalte enthält die Regestennummer mit Heftangabe, die dritte Spalte das Datum des Regests und die vierte Spalte schließlich den Typ des Eintrages. Hierzu aber später mehr.



## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



### 3.1.4 Die Liste der Edges

nodeID	registernummer	emp	regID	datum
428	#24-9	0	119210	1442-07-18
428	#24-13	0	119214	1442-08-09
428	#24-47	0	119248	1445-03-05
429	#27-278	0	120842	1475-12-06
430	#13-189	0	114880	1451-01-26
431	#0-5004	0	105002	1467-05-22
431	#22-178	0	118282	1467-05-22
432	#19-111	1	117177	1451-02-19

Abbildung 8: Ausschnitt aus der Edgesdatei

Abschließend wird noch eine Liste mit den Verknüpfungen zwischen Registereinträgen (*Nodes*) und Regesten (auch *Nodes*) erstellt. Diese Verknüpfungen nennt man Kanten oder *Edges*. Quelle und Ziel der Verknüpfungen sind die in den vorherigen Listen erstellten IDs aus der ersten Spalte.

Die erste Spalte der Abbildung nennt die ID des Registereintrages. Die zweite Spalte enthält die Regestennummer mit Heftangabe. In der dritten Spalte wird angegeben, ob der Registereintrag Empfänger der im Regest beschriebenen Urkunde ist, hierzu aber mehr im nächsten Abschnitt. Spalte vier enthält die ID des Regests aus der Regestenliste und Spalte 5 das Datum des jeweiligen Regests.

### 3.1.5 ist\_empfaenger

nodeID	registernummer	empfaenger	regID	datum
1501	#13-177	1	114868	1450-09-25
1501	#13-183	1	114874	1450-11-16
1501	#13-189	1	114880	1451-01-26
1501	#13-217	1	114908	1451-09-07
1501	#26-187	1	119931	1446-05-13
1503	#26-187	1	119931	1446-05-13
1509	#26-187	1	119931	1446-05-13

Abbildung 9: Ausschnitt aus der Datei *ist\_empfaenger*

Dem Ausschnitt des Registers zur Familie Volckamer ist zu entnehmen, dass einige der Regestennummern fett gedruckt sind. Dies bedeutet, dass der Registereintrag der Empfänger jener im Regest zusammengefassten Urkunde ist. Auch diese Informationen sollen in die Graphdatenbank übernommen werden. In der Abbildung ist eine Zeile markiert, deren Werte hier erläutert werden. Die erste Spalte enthält die ID des Registereintrages. In der zweiten Spalte wird die Regestennummer des

Regests angegeben, dessen zugrunde liegende Urkunde die im Registereintrag genannte Person empfangen hat. Spalte 4 nennt die ID des Regests und in Spalte 5 wird das Datum des Regests bzw. der Urkunde angegeben.

### 3.1.6 is\_mother\_off

Die letzte CSV-Datei mit der Bezeichnung *is\_mother\_of.csv* enthält Informationen zur Hierarchie des Registers. Die zweispaltige Tabelle enthält in der ersten Spalte die ID eines Eintrages und in der zweiten Spalte die ID des jeweiligen übergeordneten Eintrages im Register. Mit diesen Informationen lässt sich die Hierarchie des Registers auch in der Graphdatenbank wieder aufbauen<sup>6</sup>.

	A	B
1	son	mother
2	2	1
3	7	6
4	8	6
5	9	6
6	10	6

Abbildung 10: Blick in die CSV-Datei *IS\_MOTHER\_OF*

<sup>6</sup> In einer neueren Version der Datenbank wurde die Angabe *is\_son\_of* durch die Bezeichnung *OBERLEMMA* ersetzt, da die Gefahr für Verwechslungen mit Verwandtschaftsbeziehungen verringert.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



### 3.1.7 Dateneingabe

Für den Datenimport in neo4j werden in diesem Beispiel csv-Dateien verwendet. Damit stehen uns nach den ganzen, eben beschriebenen Umwandlungen fünf csv-Dateien zur Verfügung: *nodes.csv*, *regestenliste.csv*, *edges.csv*, *ist\_empfaenger.csv* und *is\_son\_of*.

Die Installation von neo4j wird auf den Internetseiten des Anbieters beschrieben<sup>7</sup>. Es stehen Installationspakete für Windows, Linux und MacOSX zur Verfügung.

Nach der Installation rufen wir die *neo4j-shell* auf, die Kommandozeile von neo4j<sup>8</sup>.

Zum Import der ersten Datei werden folgende Befehle eingegeben (oder in die *neo4j-shell* kopiert):

```
// Lemmata
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///home/neo4j/RegisterFIII/nodes.csv"
AS line FIELDTERMINATOR '|'
// WITH line LIMIT 500
CREATE (:Lemma {id:line.nodeID, xmlid:line.xmlID, name1:line.name1, name2:line.name2, name3:line.name3,
type:line.type})
RETURN line.nodeID, line.name1;
```

Die erste Zeile erzwingt einen *commit* nach jeweils 1000 eingelesenen Zeilen. Die zweite Zeile enthält das Kommando für das Einlesen der CSV-Datei, wobei als Feldtrenner das Pipesymbol „|“ gewählt wird. Die Befehlssequenz *WITH HEADERS* verweist darauf, dass die erste Zeile der CSV-Datei die Spaltenüberschriften enthält. In der dritten Zeile des Befehlsbeispiels wird angegeben, dass für jede eingelesene Zeile der CSV-Datei ein *node*-Eintrag in der Datenbank erzeugt wird. Dabei werden die einzelnen Angaben für die Erstellung der Nodes mit Hilfe der Spaltenüberschrift in der CSV-Datei den einzelnen Werten zugeordnet. Im Folgenden werden die einzelnen Sequenzen des *CREATE*-Befehls nochmal erläutert:

Befehl	Erläuterung
<i>CREATE</i>	Mit dem <i>CREATE</i> -Befehl wird eine Node in neo4j angelegt.
<i>:Lemma</i>	Die angelegte Node repräsentiert ein Lemma des Registers
<i>id:line.nodeID</i>	Die Lemma-Node soll die ID <i>line.nodeID</i> haben.
<i>name1:line.name1</i>	Import des Feldes name1 (vgl. Abschnitt <i>die Lemmaliste</i> ).

<sup>7</sup> Vgl. <http://neo4j.com/download/> (abgerufen am 12.03.2015).

<sup>8</sup> Weitere Informationen zur neo4j-shell finden sich im neo4j-Nutzerhandbuch: <http://neo4j.com/docs/stable/shell.html> (abgerufen am 12.03.2015).

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



<i>name2:line.name2</i>	Import des Feldes name2 (vgl. Abschnitt <i>die Lemmaliste</i> ).
<i>name3:line.name3</i>	Import des Feldes name3 (vgl. Abschnitt <i>die Lemmaliste</i> ).
<i>type:line.type</i>	Ist eine Node vom Typ <i>Lemma</i> .

Dass es sich bei der angelegten Node um eine Node vom Typ *Lemma* handelt, wird zweimal berücksichtigt. Einmal als Node vom Typ *Lemma* und nochmals als Eigenschaft der Node. Diese Doppelung wird für den späteren Export nach Gephi und für die dortige Visualisierung benötigt.

Die Zeile *// WITH line LIMIT 5* ist mit zwei Schrägstrichen auskommentiert. Entfernt man den Kommentar, werden nur die ersten fünf Zeilen der CSV-Datei eingelesen. Damit lässt sich die Fehlerfreiheit des Skripts leichter und schneller testen, da ein schneller Rechner für die Abarbeitung des gestammten Materials mehrere Stunden benötigt.

Alle weiteren Befehle für die *neo4j-shell* sind ähnlich aufgebaut.

```
// Regesten
LOAD CSV WITH HEADERS FROM "file:///home/neo4j/RegisterFIII/regesten.csv"
AS line FIELDTERMINATOR '|'
// WITH line LIMIT 5
CREATE (:Regest {id:line.regID, nummer:line.regestennummer, date:line.datum, type:line.type})
RETURN line.regID;

// Edges
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///home/neo4j/RegisterFIII/edges.csv"
AS line FIELDTERMINATOR '|'
// WITH line LIMIT 5
MATCH (from {id:line.nodeID}), (to {id:line.regID}) create from-[r:GENANNT_IN{datum:line.datum}]->to
RETURN r;

// Empfänger-Edges
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///home/neo4j/RegisterFIII/empfaenger.csv"
AS line FIELDTERMINATOR '|'
// WITH line LIMIT 5
MATCH (from {id:line.nodeID}), (to {id:line.regID}) create from-[r:EMPFAENGER_IN{datum:line.datum}]->to
RETURN line.nodeID,r,line.regestennummer;

// is-son-of
```

### Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



```
USING PERIODIC COMMIT 1000
LOAD CSV WITH HEADERS FROM "file:///home/neo4j/RegisterFIII/is-son-of.csv"
AS line FIELDTERMINATOR '|'
// WITH line LIMIT 5
MATCH (from {id:line.son}), (to {id:line.mother}) create from-[:SON_OF]->to
RETURN *;
```

Es können auch alle Befehle auf einmal in die Shell kopiert werden. Dann dauert der Einspielprozess einige Stunden. Anschließend können erste Versuche mit dem Material unternommen werden.

## 3.2 Datenausgabe

Nach dem Ende des Einspielprozesses sind alle Informationen in der Graphdatenbank vorhanden. Im nächsten Schritt werfen wir einen Blick auf das Webfrontend von neo4j.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



### 3.2.1 Neo4j-Webfrontend

Für die Anzeige des Webfrontends von Neo4j benötigt man einen aktuellen Browser (Firefox, Safari oder Chrome). Dort gibt man in die Adresszeile ein:

`http://localhost:7474`

Anschließend erscheint das Webfrontend von Neo4j.

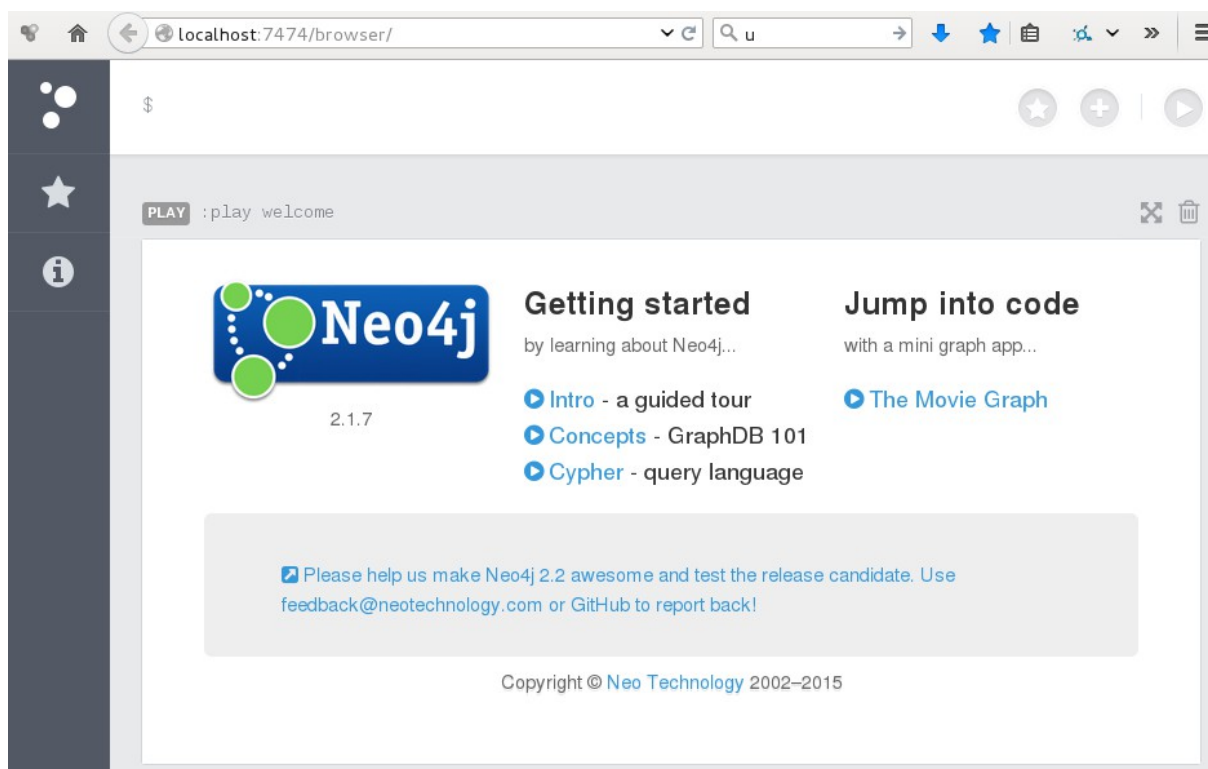


Abbildung 11: Webfrontend von Neo4j

Zunächst wenden wir uns den Symbolen am linken Rand des Browserfensters zu, zum einen den drei verschieden großen Kreisen und zum anderen dem Stern-Symbol.

Nach Auswahl des Symbols mit den drei verschieden großen Kreisen blendet Neo4j eine weitere Anzeigespalte ein, welche die in der Datenbank verfügbaren Nodes und Edges auflistet.

Wählt man unter *Node labels* die Schaltfläche *Lemma* aus, werden einige Lemmata aus dem Register angezeigt.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).

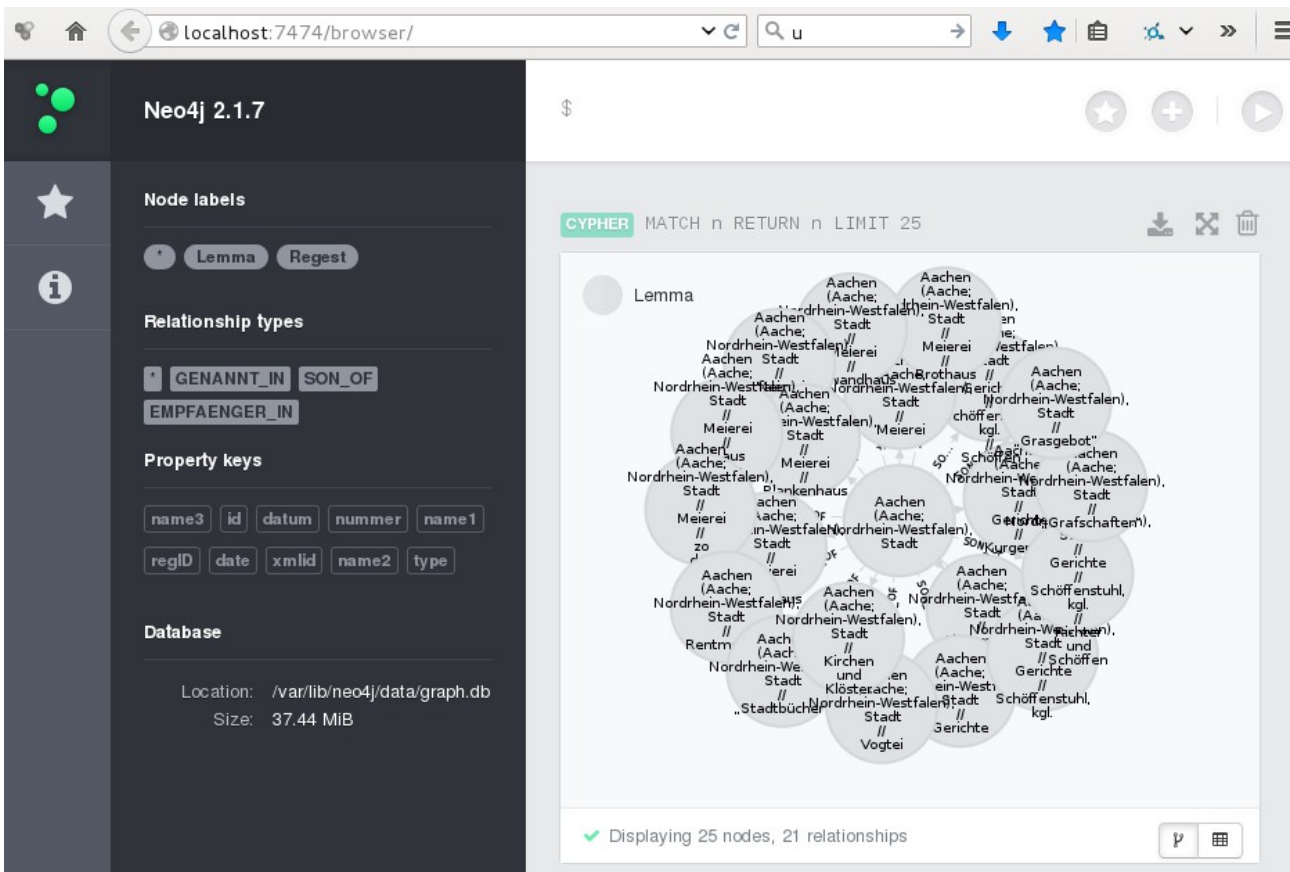


Abbildung 12: Einige Lemmata aus dem Register in Neo4j

Nach Belieben können nun andere Ansichten des Registers über die Auswahl der weiteren Schaltflächen angezeigt werden. Unter *Relationship types* sind die verschiedenen Typen von Verknüpfungen, also Edges, aufgelistet. Schließlich werden unter *Property keys* noch alle vorhandenen Eigenschaften von Lemmata, Regesten und Verknüpfungen aufgelistet.

## Zitation:

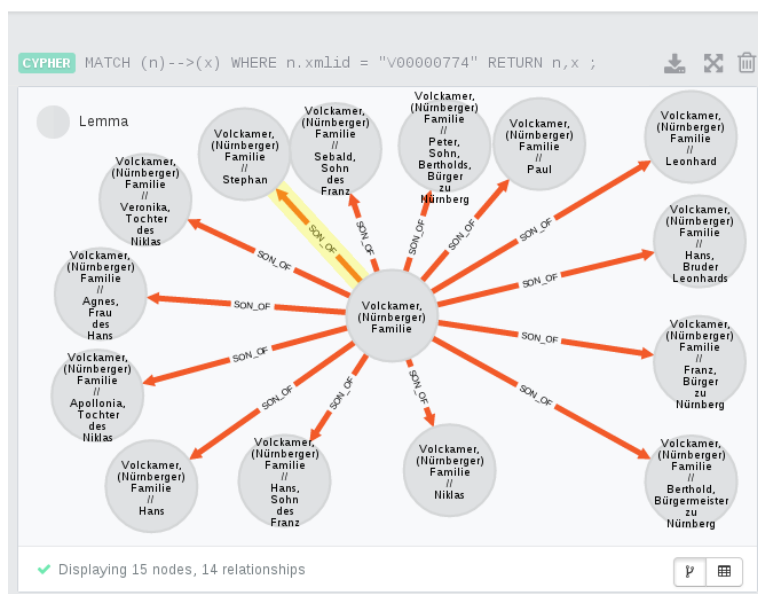
Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



```
$ MATCH (n)-->(x) WHERE n.xmlid = "V00000774"  
RETURN n,x ;
```



### 3.2.2 Die Abfragesprache cypher



Im vorherigen Abschnitt haben wir zur Ansicht des Materials nur die bereits angebotenen Abfragen genutzt. In diesem Abschnitt sollen erste eigene Schritte unternommen werden. Neo4j bringt eine eigene Abfragesprache mit. Sie nennt sich *cypher*. Da hier nur erste Schritte erläutert werden können, sei an dieser Stelle auf die Dokumentation von *cypher* im Neo4j-Handbuch verwiesen<sup>9</sup>.

Als ersten Beispiel rufen wir die oben bereits angesprochene Familie Volckamer auf und nutzen hierfür die XML-ID des Eintrages im Register.

Abbildung 13: Der Registereintrag der Familie Volckamer in Neo4j

Der entsprechende *cypher*-Befehl lautet:

```
MATCH (n)-->(x) WHERE n.xmlid = "V00000774" RETURN n,x;
```

und ergibt die in der Abbildung zu sehende Ausgabe. Einen solchen Teil der gesamten Graphdatenbank bezeichnet man als *Subgraph*. Der Befehl *MATCH* liefert als Treffer alle *Nodes x* zurück, die in irgendeiner Beziehung zur *Node n* mit der *xmlid V00000774* stehen. Im XML-Register hat das Hauptlemma der Familie Volckamer genau diese XML-ID, und rundherum werden alle Untereinträge angezeigt, die mit dem Hauptlemma über Edges mit dem Typ *SON\_OF* verknüpft sind.

<sup>9</sup> <http://neo4j.com/docs/stable/cypher-query-lang.html> abgerufen am 13.03.2015.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



### 3.2.3 Visualisierungsoptionen im neo4j-Webfrontend

Klickt man auf eine Node, werden in einem separaten Fenster die Eigenschaften angezeigt.

```
$ MATCH (n)-->(x) WHERE n.xmlid = "V00000774" RETURN n,x;
```



The screenshot shows the Neo4j web interface. At the top, a CYPHER query is entered: `MATCH (n)-->(x) WHERE n.xmlid = "V00000774" RETURN n,x;`. Below the query, a graph visualization is displayed with a central node labeled "Volckamer, (Nürberger) Familie" and several surrounding nodes connected by orange arrows labeled "SON\_OF". A properties window for the selected node is open on the right, showing the following details:

Lemma [27028]	
Properties	
id	45195
name3	Volckamer, (Nürberger) Familie // Berthold, Bürgermeister zu Nürnberg
name2	
name1	Berthold, Bürgermeister zu Nürnberg
xmlid	V00000777
type	lemma

At the bottom of the interface, it says "Displaying 15 nodes, 14 relationships".

Abbildung 14: Anzeige der Eigenschaften einer Node

Das Fenster ist in zwei Tabs aufgeteilt. Im ersten Tab *Properties* werden die Werte der Node oder der Edge aufgelistet. Im zweiten Tab mit dem symbolisierten Auge können Eigenschaften der Anzeige wie Größe und Farbe der *Nodes* oder der *Edges* geändert werden sowie die jeweils angezeigte *Property* eingestellt werden. Damit kann die Anzeige etwas übersichtlicher gestaltet werden.

In einem zweiten Beispiel soll der Subgraph zum Nürnberger Bürger Niklas Muffel und den anderen mit ihm über Regesten verknüpften Registereinträgen dargestellt werden. Die Befehlszeile sieht folgendermaßen aus:

```
MATCH (n)-[r1]->(m)-[r2]->(x) WHERE n.xmlid="M00002942" RETURN n,m,x,r1,r2 LIMIT 500;
```

Hier wird zunächst die Node (n) gesucht, welche, gesteuert über die Angabe

```
WHERE n.xmlid="M00002942"
```

auf den Registereintrag von Niklas Muffel weist. Ausgehend von diesem Startpunkt werden alle mit diesem Lemma verknüpften Einträge aufgelistet. Die Verknüpfung wird repräsentiert durch

```
<-[r1]->
```



Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



und die verküpften Einträge durch die Variable ( $m$ ).

Dann werden ausgehend von jedem einzelnen Treffer ( $m$ ) alle mit diesem verküpften Einträge ( $x$ ) aufgelistet, jeweils verbunden über die Verknüpfung

<-[r2]->

Das abschließende Statement

```
RETURN n,m,x,r1,r2 LIMIT 500
```

gibt neo4j die Anweisung, die Werte  $n,m,x,r1,r2$  bis zu einer Gesamttreffermenge von 500 zurück zu liefern und anzuzeigen.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).

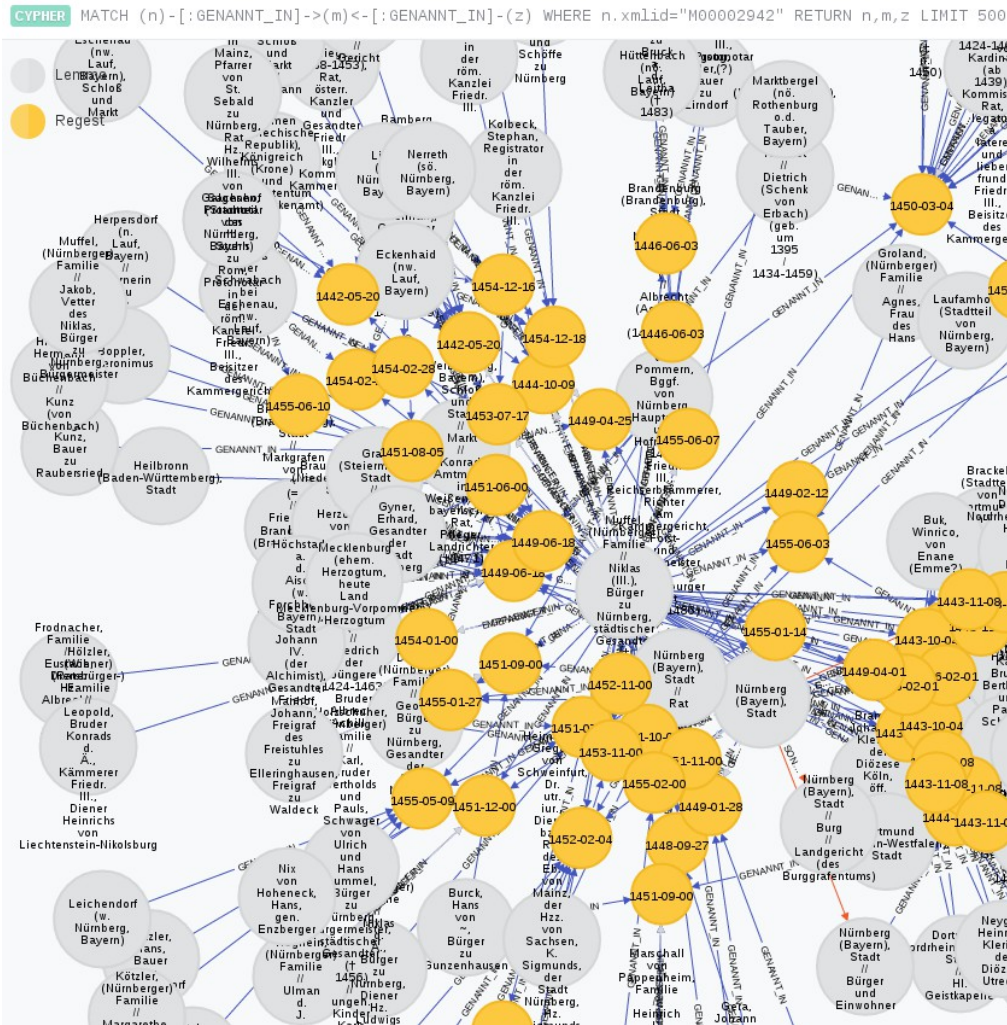


Abbildung 15: Das Hauptlemma der Nürnberger Familie Muffel mit allen Unterlemmata und den genannten Regesten

In diesem Beispiel wurden alle Regestennodes in gelber Farbe, die Lemmata in grauer Farbe, die *SON\_OF*-Edges in roter Farbe und die *GENANNT\_IN*-Edges in blauer Farbe dargestellt. Dies verbessert zwar die Erkennbarkeit der Strukturen, jedoch ist das Anzeigefenster zu klein, um alle Treffer in lesbarer Art und Weise sichtbar zu machen.

Um diesen Subgraphen weiter auswerten zu können, kann man einerseits versuchen, die Ergebnismenge einzuschränken. Andererseits könnten aber auch Fragestellungen auftreten, bei denen weit größere Ergebnismengen vorliegen als im oben dargestellten Beispiel. Für diese Fälle gibt es die Möglichkeit, den Subgraphen zu exportieren und in einem anderen Programm weiter zu analysieren und zu bearbeiten.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



### 3.2.4 Export von Subgraphen zur Weiterbearbeitung

Für den Export von Daten aus Neo4j ist die Installation der *neo4j-shell-tools* notwendig. Das Paket wird unter <https://github.com/jexp/neo4j-shell-tools> (abgerufen am 13.03.2015) im Internet angeboten. Dort wird auch die Installation des Pakets beschrieben. Mit der Installation stehen nun in der *neo4j-shell* zusätzliche Befehle zur Verfügung, darunter auch der Befehl *export-graphml*.

Nach Aufruf der Shell können Sie nun folgenden Befehl eingeben:

```
export-graphml -t -o /home/neo4j/RegisterFIII/subgraph.graphml MATCH
(n)<-[r1:GENANNT_IN]->(m)<-[r2:GENANNT_IN]->(z) WHERE n.xmlid="M00002942"
RETURN n,m,z,r1,r2 LIMIT 500
```

Der Befehl liefert bei Erfolg folgendes Ergebnis zurück:

```
Wrote to GraphML-file /home/neo4j/RegisterFIII/subgraph.graphml 0. 100%: nodes = 192 rels = 559 properties = 1593
time 1849 ms total 1849 ms
```

Der Parameter *-t* führt zu einer speziell für die Weiterbearbeitung in Gephi notwendigen Formatierung der graphml-Datei. Der Parameter *-o* gibt den Pfad zur Datei an. Hier ist besonders auf die Schreibberechtigung von neo4j zu achten, da es ansonsten zu Fehlern kommt. Zum Schluss wird der *cypher*-Befehl zur Erstellung des Subgraphen angegeben. Zu beachten ist, dass die Reihenfolge der Kommandozeilenparameter nicht beliebig ist, sondern genau in der oben angegebenen Weise erfolgen muss, sonst funktioniert der Export nicht. Zum Schluss haben wir eine Datei *subgraph.graphml*, die wir mit Gephi öffnen können.

## 4 Visualisierung von Netzwerken in Gephi<sup>10</sup>

### 4.1 Das Programm Gephi

Gephi ist Open-Source-Software und kann auf der Seite [www.gephi.org](http://www.gephi.org) heruntergeladen werden<sup>11</sup>. Nach der Installation können die exportierten Daten importiert werden.

### 4.2 Import und Aufbereitung der Daten

Nach dem Öffnen von Gephi erstellt man ein neues Projekt und kann anschließend die GraphML-Datei öffnen. Die vorgeschlagenen Importparameter können so übernommen werden. Im oberen Bereich des Programmfensters finden sich die drei Reiter *Overview*, *Data Laboratory* und *Preview*. Falls im Bereich *Overview* kein Graph angezeigt wird, kann man die Anzeige über das Menü

<sup>10</sup> Das Kapitel ist teildentisch mit Andreas Kuczera: Digitale Farbenspiele oder nützliches Werkzeug – Visualisierung von Netzwerken aus den Registern von Editions- und Regestenwerken, in: Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte, 8. Januar 2015, <http://mittelalter.hypotheses.org/5089> (ISSN 2197-6120).

<sup>11</sup> Auf der Seite [www.gephi.org](http://www.gephi.org) findet man unter dem Reiter *Download* die Softwarepakete für verschiedene Betriebssysteme. Installationsanleitungen für die einzelnen Betriebssystemumgebungen finden sich unter <https://gephi.github.io/users/install/> (abgerufen am 30.3.2015).

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



*Fenster – Graph* erzwingen. Das gleiche gilt für den Fall, dass im Bereich *Data Laboratory* nichts angezeigt wird. Auch hier kann über das Menü *Fenster – Data Table* die Anzeige erzwungen werden.

Zunächst sehen wir uns im Bereich *Overview* die verschiedenen Visualisierungsmöglichkeiten näher an.

### 4.3 Visualisierungsmöglichkeiten

Wählt man im oberen Bereich des Programmfensters den Reiter *Data Laboratory* aus, erscheinen die der Netzwerkvisualisierung zu Grunde liegenden Daten. Unter *Nodes* erkennen wir unsere Personen aus dem Register wieder. Der Eintrag mit der ID n15957 ist z.B. unser Niklas Muffel III.

Nodes	name3	name1	xmlid
n15861	n1n1München (Bayern), Stadt // Tag zwischen Fürsten und Reichsstädten	Tag zwischen Fürsten und Reichsstädten	M00002805
n15956	n1n1Muffel, (Nürnberger) Familie // Jakob, Vetter des Niklas, Bürger zu N	Jakob, Vetter des Niklas, Bürger zu Nürn	M00002941
n15957	n1n1Muffel, (Nürnberger) Familie // Niklas (III.), Bürger zu Nürnberg, städ	Niklas	M00002942
n16004	n1n1Murach (Obermurach bei Oberviechtach, nö. Schwandorf, Bayern), He	Friedrich zu Flügelsberg, Pfleger zu Holns	M00003049
n16228	n1n1Neitperg (Neuberg, w. Hartberg, Steiermark), Herren von ~ (Neitperç	Johann	N00000285
n16254	n1n1Nerreth (sö. Nürnberg, Bayern)	Nerreth	N00000322
n16509	n1n1Neyger, Heinrich, Kleriker der Diözese Utrecht	Neyger, Heinrich, Kleriker der Diözese Ut	N00000792
n16630	n1n1Nikolaus V., (Tommaso Parentucelli) Papst (1447–1455)	Nikolaus V.,	N00001044
n16645	n1n1Nithart, Meister Peter, Gesandter des Schwäbischen Städtebundes ar	Nithart, Meister Peter, Gesandter des Sch	N00001072
n16651	n1n1Nix von Hoheneck, Hans, gen. Enzberger	Nix von Hoheneck, Hans, gen. Enzberger	N00001081
n16667	n1n1Nördlingen (Bayern), Stadt	Nördlingen	N00001098
n16693	n1n1Nördlinger, Stefan, aus dem Bistum Bamberg, öff. Notar	Nördlinger, Stefan, aus dem Bistum Bamt	N00001130
n16779	n1n1Nürnberg (Bayern), Stadt	Nürnberg	N00001303
n16780	n1n1Nürnberg (Bayern), Stadt // Bürger und Einwohner	Bürger und Einwohner	N00001304
n16783	n1n1Nürnberg (Bayern), Stadt // Botschaften	Botschaften	N00001311
n16787	n1n1Nürnberg (Bayern), Stadt // Rat	Rat	N00001321
n16796	n1n1Nürnberg (Bayern), Stadt // Burg // Landgericht (des Burggrafentums	Landgericht	N00001339
n30265	n3n3	#0- 1442	
n44942	n4n4	#14 1442	
n45034	n4n4	#14 1443	
n45035	n4n4	#14 1443	
n45036	n4n4	#14 1443	
n45037	n4n4	#14 1443	

Abbildung 16: Liste der importierten Nodes im Data Laboratory von Gephi

Unter dem Reiter *Edges* werden die Verknüpfungdaten (*Edges*) angezeigt. Die markierte Edge verweist auf eine gemeinsame Nennung von Bamberg in Bayern (mit der ID n1096) im Regest 463

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



aus Heft 14 mit dem Datum 1449-04-29. Im Datenlabor hat man also direkten Zugriff auf alle der Visualisierung zu Grunde liegenden Daten.

Source	Target	Type	Id	datum
n848 - n848	n46963 - n46963	Directed	e2590	1 1451-05-14
n889 - n889	n48166 - n48166	Directed	e3049	1 1450-03-04
n1021 - n1021	n48166 - n48166	Directed	e3791	1 1450-03-04
<b>n1096 - n1096</b>	<b>n45301 - n45301</b>	<b>Directed</b>	<b>e4352</b>	<b>1 1449-04-25</b>
n1096 - n1096	n45315 - n45315	Directed	e4354	1 1449-06-18
n1096 - n1096	n45316 - n45316	Directed	e4355	1 1449-06-18
n1096 - n1096	n45317 - n45317	Directed	e4356	1 1449-06-18
n1096 - n1096	n45318 - n45318	Directed	e4357	1 1449-06-18

Abbildung 17: Liste der Edges im Data Laboratory von Gephi

Wechselt man in den Bereich *Overview*, zeigt Gephi zunächst die entsprechenden Nodes und Edges in einem Würfel an.

## 4.4 Auswahl des Layout: *ForceAtlas2*

Auf der linken Seite des Programmfensters können im Fenster *Layout* verschiedene Layouts für die weitere Bearbeitung der Daten ausgewählt werden. In unserem Beispiel wenden wir nun das Layout *ForceAtlas2* an. Nach kurzer Zeit hat sich unser Würfel auseinandergezogen und es werden verschiedene Zentren bzw. Gruppen von Nodes sichtbar.



Abbildung 18: Gephi zeigt die erste Visualisierung der Daten als Würfel an.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).

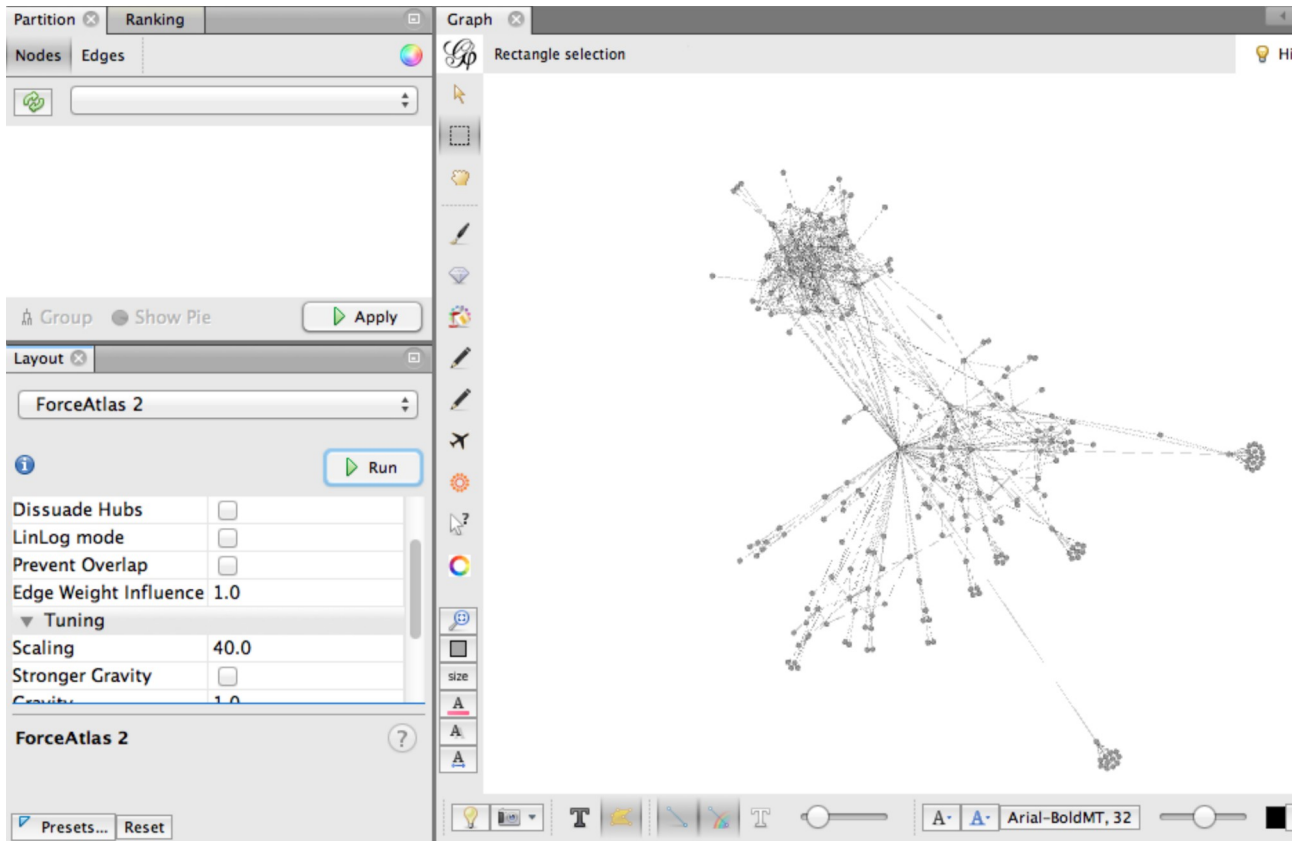


Abbildung 19: Nach Anwendung des Layouts ForceAtlas2

In den folgenden Abschnitten werden noch weitere Funktionen erläutert, mit denen charakteristische Merkmale der Daten sichtbar gemacht werden können.

### 4.5 Lemmata und Regesten farbige darstellen

Für den besseren Überblick lassen sich Regesten und Lemmata verschiedene Farben zuordnen. Hierfür benötigt Gephi aber Informationen zum Typ der jeweiligen Nodes. Daher wurde in vorher in Neo4j den Nodes jeweils die Angabe hinzugefügt, ob es sich um Regesten oder Lemmata handelt.

Diese Angaben können jetzt für die unterschiedliche Farbzweisung verwendet werden.

Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).

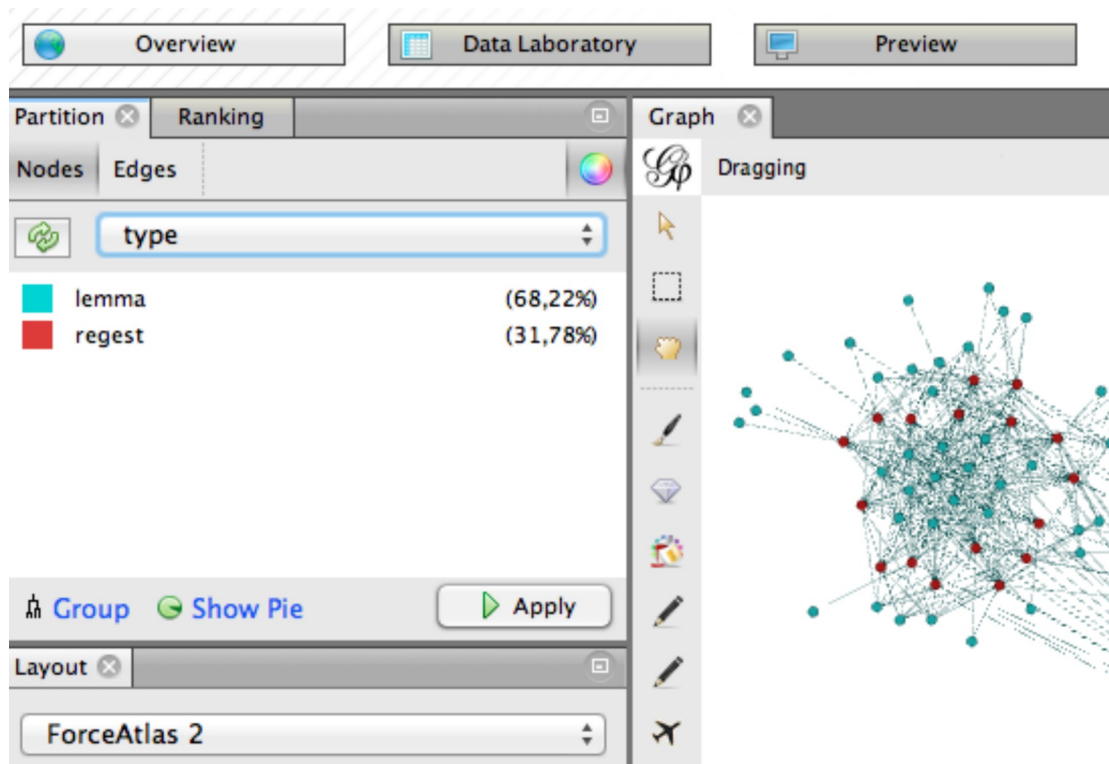


Abbildung 20: Zuweisung unterschiedlicher Farben für Regesten und Lemmata

Hierfür wird oben links im Bereich Partition „Nodes“ ausgewählt. Anschließend werden über den grünen Aktualisierungsknopf die verfügbaren Eigenschaften aktualisiert. Dann werden bei der Auswahl unter *type* für *lemma* und *regest* verschiedene Farben angeboten. Mit *Apply* werden die angebotenen Farben den Nodes zugeordnet.

#### Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



## 4.6 Degree Range

Unter dem Reiter *Filters* am rechten Bildrand gibt es noch weitere Analysemöglichkeiten. Unter dem Punkt *Topology* befindet sich der Unterpunkt *Degree Range*. Mit diesem kann man auswählen, wieviele Verknüpfungen eine Node haben muss, damit sie noch angezeigt wird. Für unser Beispiel haben wir die alle Nodes ausgeblendet, die 42 oder weniger Verknüpfungen haben.

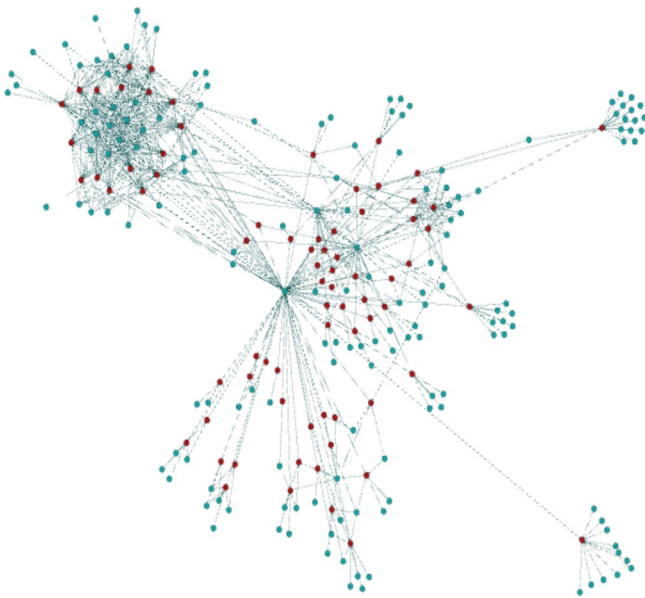


Abbildung 21: Gesamtbild des Graphen

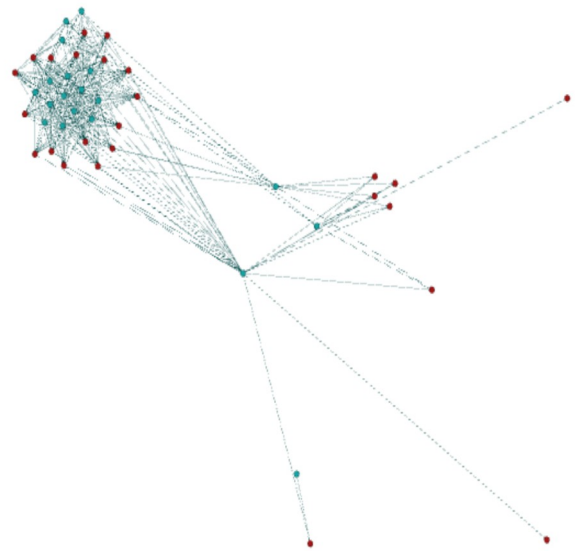


Abbildung 22: Nach Anwendung des Degree-Filters

In der linken Abbildung wird das gesamte Netzwerk abgebildet, im rechten Bild nur noch jene Nodes, die 10 oder mehr Verknüpfungen haben. Zu diesen „wichtigen“ Nodes lassen sich auch die Namen einblenden.



Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



Abbildung 23: Nodes mit eingblendeter Beschriftung

Wechselt man von dieser Ansicht ins Datenlabor, werden dort nur noch jene Nodes angezeigt, die in der Auswahl vorher sichtbar waren.

## Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



No...	Id	L...	name3	n...	date
<input type="checkbox"/>	r n144	n144	Beheim, (Nürnberger) Familie // Michael d. Ä., Bürger zu Nürnberg		
<input type="checkbox"/>	r n416	n416	Ebner, (Nürnberger) Familie // Matthias (Mathes), Bürger zu Nürnberg		
<input type="checkbox"/>	r n422	n422	Eckenhaid (nw. Lauf, Bayern)		
<input type="checkbox"/>	r n487	n487	Erengroß, Leonhard, von Gastein, Prokurator des ksl. Kammergerichts, S		
<input type="checkbox"/>	r n502	n502	Eschenloer, Konrad (Kunz), Bürger zu Nürnberg		
<input type="checkbox"/>	r n689	n689	Geuder, (Nürnberger) Familie // Georg, Bürger zu Nürnberg		
<input type="checkbox"/>	r n741	n741	Graser, (Nürnberger) Familie // Hans, Bürger zu Nürnberg		
<input type="checkbox"/>	r n763	n763	Groland, (Nürnberger) Familie // Hans, Bürger zu Nürnberg		
<input type="checkbox"/>	r n778	n778	Grundherr, (Nürnberger) Familie // Paul, Bürger zu Nürnberg, Bürgerme		
<input type="checkbox"/>	r n832	n832	Haller, (Nürnberger) Familie // Georg, Sohn Peters d. Ä., Vetter Ludwigs		
<input type="checkbox"/>	r n835	n835	Haller, (Nürnberger) Familie // Ulrich (d. Ä. oder d. J.), Bürger zu Nürnbr		
<input type="checkbox"/>	r n882	n882	Hegnein, (Nürnberger) Familie // Ulman (Ulrich), Bürger zu Nürnberg		
<input type="checkbox"/>	r n999	n999	Holzschuher, (Nürnberger) Familie // Karl, Bruder Bertholds und Pauls, S		
<input type="checkbox"/>	r n122	n122	Krebs, Wilhelm vom ~, Bürger zu Köln		
<input type="checkbox"/>	r n146	n146	Marquard, Goswin, Lic. utr. iur., Kanoniker zu Utrecht, Bruder Johanns /,		
<input type="checkbox"/>	r n151	n151	Mendel, (Nürnberger) Familie // Peter d. Ä., Bürgermeister zu Nürnberg		
<input type="checkbox"/>	r n159	n159	Muffel, (Nürnberger) Familie // Niklas (III.), Bürger zu Nürnberg, städtisc		
<input type="checkbox"/>	r n167	n167	Nürnberg (Bayern), Stadt		
<input type="checkbox"/>	r n167	n167	Nürnberg (Bayern), Stadt // Rat		

Abbildung 24: Blick ins Datenlabor. Hier werden nur noch jene Nodes angezeigt, die vorher in der Auswahl sichtbar waren.

## 4.7 Timeline

Abschließend soll noch die Verwendung der *Timeline* in Gephi erläutert werden. Die Datengrundlage unserer Visualisierung bilden zum einen die aus dem Register extrahierten Registereinträge und Regesten (*Nodes*) und die aus den Regestenangaben im Register gewonnenen Verküpfungen (*Edges*). Jede Verküpfung zwischen den Registereinträgen geht auf die gemeinsame Zuordnung zu einem Regest zurück. Das Regest selbst hat ein Datum und dieses kann man den Verküpfungen zuordnen, so dass sie zeitlich eingeordnet werden können. Hier soll nur auf die zeitliche Einordnung eingegangen werden.

Momentan erstellt Gephi aus importierten Datumsangaben nicht automatisch eine Timeline, auch wenn die Daten in der normierten Form YYYY-MM-DD vorliegen. Es gibt aber einen Workaround, mit dem man Gephi überzeugen kann, doch die importierten Datumsangaben für die Visualisierung mit einer Timeline zu verwenden. Hierfür wechselt man ins Datenlabor in den Bereich Edges. Im unteren Bereich des Programmfensters wählt man nun *merge columns* aus. Im sich nun öffnenden Dialogfenster wählt man unter *available columns* die Spalte *datum* und bringt sie mit dem Pfeilsymbol in den Bereich *columns to merge*. Im sich nun öffnenden neuen Fenster wählt man *time interval creation options*. Hier muss man nun *parse dates* auswählen und das Format zu *yyyy-mm-dd* anpassen. Nach dem OK-Click erscheint die Timeline.

### Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



Aktiviert man am unteren Ende des Programmfensters in Gephi die *Timeline*, kann ein bestimmter Zeitintervall für die Anzeige ausgewählt werden.

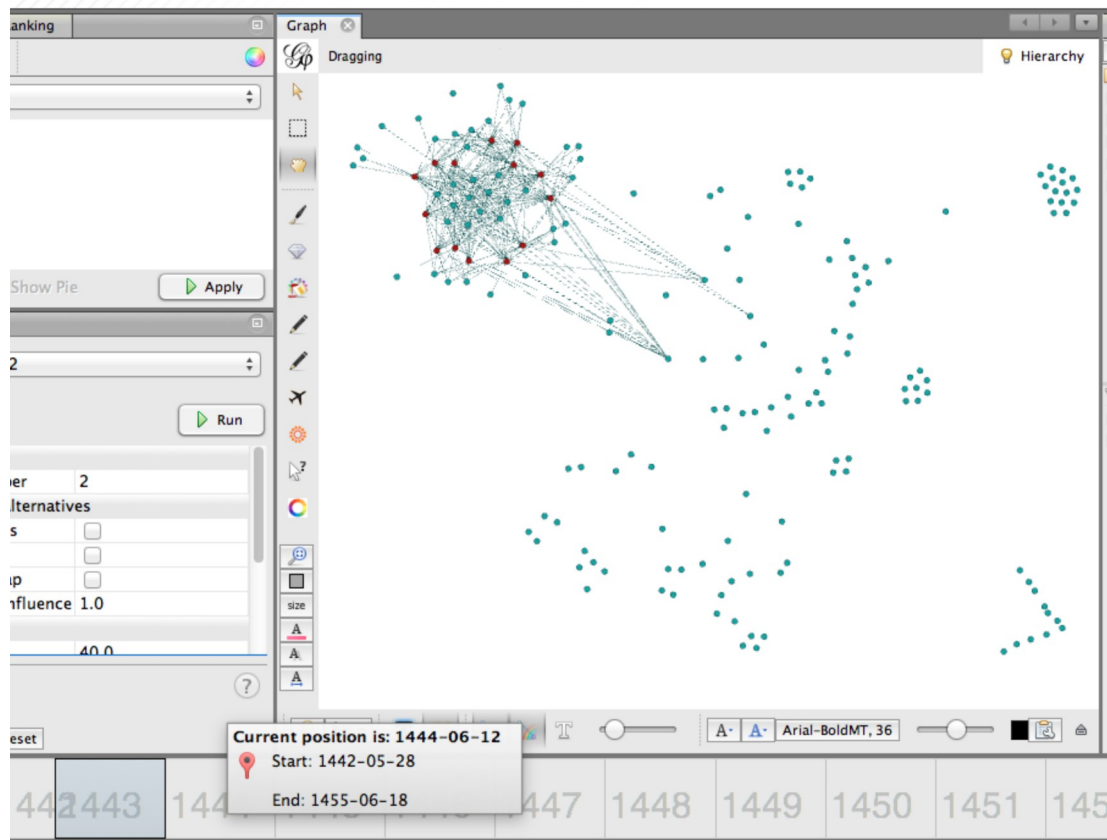


Abbildung 25: Gephi mit aktivierter Timeline. Hier werden nur die Verknüpfungen des Jahres 1443 angezeigt.

Unser Beispiel enthält Daten über einen Zeitraum von 1442 bis 1454. Zunächst sind alle Jahre ausgewählt, so dass alles angezeigt wird. Schränkt man die Auswahl mit Hilfe des Zeitbalkens am unteren Fensterrand jedoch ein, werden nur noch jene *Edges* angezeigt, deren Datumsangaben innerhalb des ausgewählten Intervalls liegen.

Je nach Fragestellungen lassen sich hier also Netzwerkstrukturen für bestimmte Zeiträume visualisieren. Es ist auch möglich, ein ausgewähltes Zeitintervall über den gesamten Zeitraum quasi filmartig ablaufen zu lassen, jedoch lässt sich dieses Feature hier nicht darstellen.

Abschließend noch der Verweis auf die von mir verwendete Neo4-Datenbank:

Graphdatenbank des kumulierten Registers der Regesten Kaiser Friedrichs III. (Stand Anfang 2014)<sup>12</sup>: <https://www.dropbox.com/s/ciq6wc4qzf3wer1/data-FIII.tar.gz?dl=0>

<sup>12</sup> Es handelt sich um eine komprimierte-Datei mit der gesamten Neo4j-Datenbank. Nach der erstmaligen Installation von Neo4j müssen Sie im neo4j-Verzeichnis den Ordner *data* löschen und durch den entsprechenden Ordner in der

#### Zitation:

Andreas Kuczera: Graphdatenbanken für Historiker. Netzwerke in den Registern der Regesten Kaiser Friedrichs III. mit neo4j und Gephi, in: *Mittelalter. Interdisziplinäre Forschung und Rezeptionsgeschichte*, 5. Mai 2015, <http://mittelalter.hypotheses.org/5995> (ISSN 2197-6120).



## 5 Zusammenfassung

Die zunehmende Menge an Volltexten in den Geschichtswissenschaften bietet neue Chancen für die Forschung, erfordert aber auch neue Methoden und Sichtweisen. In diesem Artikel wurde der Import des kumulierten Registers der Regesten Kaiser Friedrichs III. in die Graphdatenbank Neo4j vorgestellt. Anschließend konnte ein Teil des Registers, ein sog. Subgraph, exportiert und in Gephi weiterbearbeitet werden. Graphdatenbanken können sich für Historiker zu einer Möglichkeit für die transparente Speicherung von Beziehungs- und Interpretationsstrukturen entwickeln.

Aktuelle Visualisierungswerkzeuge wie Gephi ([www.gephi.org](http://www.gephi.org)) können auch für die Visualisierung von Netzwerkstrukturen in historischen Quellenbeständen, Regesten- und Editionswerken genutzt werden.

Zu beachten ist die Qualität der Verknüpfungsdaten. Bei Registern lassen gemeinsame Nennungen von Personen in Urkunden oder Regesten nur sehr begrenzte Schlüsse auf ihre tatsächlichen Beziehungen zu.

---

komprimierten Datei ersetzen.