

The background of the slide is a complex network visualization. It consists of a dense web of small, light blue dots (nodes) connected by thin, dark blue lines (edges). The overall shape is roughly circular, with the density of nodes and edges being higher in the center and tapering off towards the edges. The text is overlaid on this network.

EINFÜHRUNG IN

**GEPHI**

**VISUALISIERUNG GROSSER NETZWERKE**

Seminar: Einführung in die Korpuslinguistik, WS 16/17

Dozent: Fabian Klinker Referent: Roman Dorniok

# ALLGEMEINES

- in Frankreich an der Technische Universität Compiègne entwickelt
- quelloffene und kostenlose Software zur Visualisierung von Datensätzen
- basiert auf **NetBeans** und damit auf **Java**
- daher Verfügbarkeit auf allen Plattformen, die Java beherrschen (welche nicht?)

# 1. INSTALLATION UND PLATTFORMEN

- Windows:  
<https://github.com/gephi/gephi/releases/download/v0.9.1-windows.exe>
- Mac OS X:  
<https://github.com/gephi/gephi/releases/download/v0.9.1-macos.dmg>
- Linux:  
<https://github.com/gephi/gephi/releases/download/v0.9.1-linux.tar.gz>

*Eine laufende Java-Umgebung ist erforderlich!*

# 2. GRUNDLAGEN

## 2.1. WELCHE DATEN WERDEN GEBRAUCHT?

- Gephi verarbeitet *Nodes* und *Edges*
- jedes dieser Elemente erhält eine eindeutige ID + weitere mögliche Attribute
- 'network topology': alle Nodes und Edges
- 'network data': alle Attribute

## BEISPIEL NODES:

ID	Label	Modularity Class
0	Alpha	1
1	Beta	0
2	Gamma	0
3	Delta	0
4	Epsilon	1

## BEISPIEL EDGES:

Source	Target	Type	ID	Weight
0	1	undirected	0	0.05
1	2	undirected	1	0.01
2	3	undirected	2	0.01

...



## 2.2. DATEIFORMATE

besonders geeignet: **GEXF** (Graph Exchange XML  
Format)

- basiert auf XML
- sehr einfach
- Beispiel:

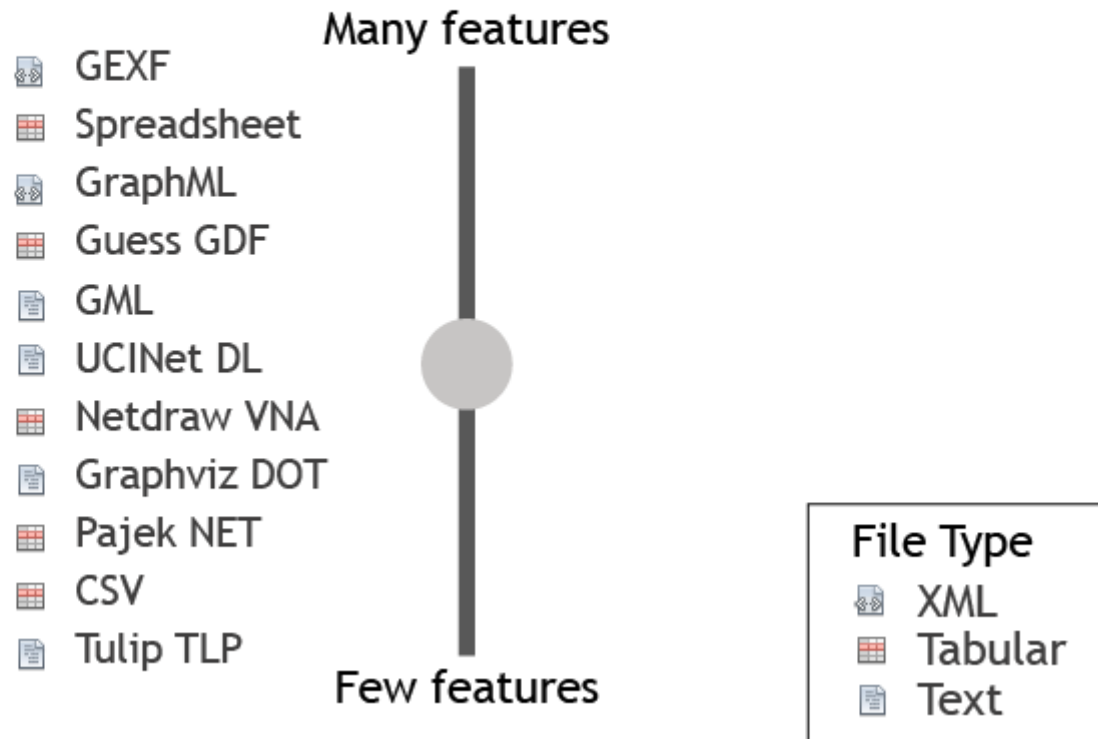
```
<?xml version="1.0" encoding="UTF-8"?>
<gexf>
  <meta>
    <creator>some corpus linguist</creator>
    <description>hello world</description>
  </meta>
  <graph mode="static" defaultedgetype="directed">
    <attributes class="node">
      <attribute id="0" title="POS" type="string"/>
    </attributes>
    <nodes>
      <node id="0" label="Hello">
        <attvalues>
```

```
        <attvalue for="0" value="..." />  
    </attvalues>  
</node>
```



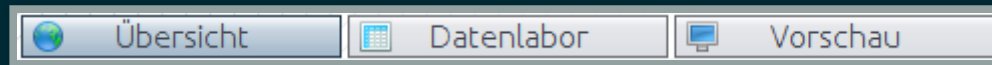
- weitere Formate:
- GML:  
[https://de.wikipedia.org/wiki/Geography\\_Markup\\_Lan](https://de.wikipedia.org/wiki/Geography_Markup_Lan)
- GDF: [https://de.wikipedia.org/wiki/Geographic\\_Data\\_I](https://de.wikipedia.org/wiki/Geographic_Data_I)
- GraphML: <https://de.wikipedia.org/wiki/GraphML>
- DOT: [https://de.wikipedia.org/wiki/DOT\\_\(GraphViz\)](https://de.wikipedia.org/wiki/DOT_(GraphViz))
- DL's (Description logics):  
<https://de.wikipedia.org/wiki/Beschreibungslogik>
- und weitere...
- ... und noch mehr.

# Do you need...



# 3. PROGRAMM-INTERFACE

- Hauptarbeitsfeld besteht aus 3 Reitern/Tabs, die frei konfigurierbar sind

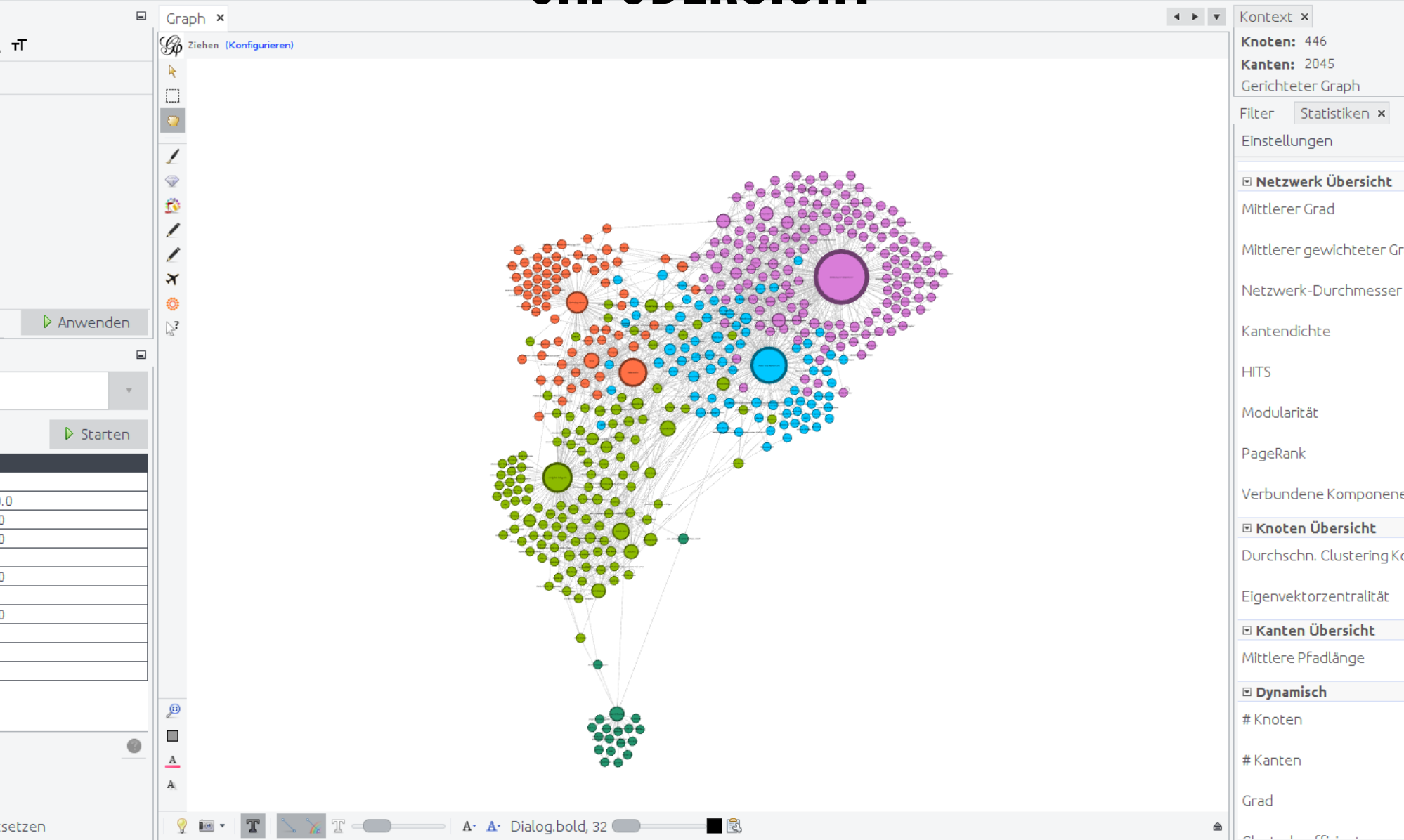


## 3.1. ÜBERSICHT

## 3.2. DATENLABOR

## 3.3. VORSCHAU

# 3.1. ÜBERSICHT



## 3.1. ÜBERSICHT > FUNKTIONEN

- Layout
- Ausgestaltung
- Statistik
- Filter

# 3.2. DATENLABOR

Operation Knoten hinzufügen Kante hinzufügen Suchen/Ersetzen Tabelle importieren Tabelle exportieren Weitere Aktionen Filter:  Id



Verbinden  
Ethernet-Netzwerk

Value	users_can_post	link	category	username	talking_about_count	post_activity	fan_count	Modularity Class	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality
no		https://w...	News/Me...	heiseonline	5705	0.39	101635	4	2.0	0.509737	0.519101
yes		https://w...	Musician/...	jannikbrun...	96	0.02	14246	3	6.0	0.222611	0.256592
yes		https://w...	Media/Ne...	heise.busin...	93	0.07	610	4	0.0	0.0	0.0
yes		https://w...	Podcast	ctuplink	21	0.01	1507	4	3.0	0.338146	0.341948
yes		https://w...	Magazine	technology...	27	0.17	1848	4	3.0	0.503394	0.566667
yes		https://w...	Community	MakerFaire...	17	0.04	3324	0	4.0	0.448137	0.549064
no		https://w...	Education	IPv6Kongress	1	0.0	237	4	0.0	0.0	0.0
no		https://w...	Magazine	ctmagazin	1489	0.1	59287	4	3.0	0.425837	0.452434
no		https://w...	Website	heiseautos	559	0.03	3670	4	3.0	0.424618	0.449064
yes		https://w...	Magazine	ix.magazin	51	0.22	2604	4	3.0	0.42543	0.451311
no		https://w...	News/Me...	TechStage...	24	0.04	20167	4	3.0	0.424618	0.449064
yes		https://w...	Magazine	makemaga...	90	0.04	4742	1	3.0	0.502257	0.621348
no		https://w...	Media/Ne...	telepolis.de	1212	0.3	18814	4	0.0	0.0	0.0
yes		https://w...	Website	heisedown...	7	0.04	1829	4	0.0	0.0	0.0
yes		https://w...	Magazine	ctdigifoto	605	0.12	34633	2	3.0	0.476956	0.564794
yes		https://w...	Media/Ne...	heisegruppe	16	0.05	2014	4	3.0	0.501691	0.517603
yes		https://w...	Magazine	ct.Mac.and.i	201	0.23	7493	4	3.0	0.42543	0.450562
no		https://w...	Website	heise.de	846	0.02	13321	4	3.0	0.424618	0.449064
yes		https://w...	Musician/...	KimLeiting...	15	0.01	3467	3	0.0	0.0	0.0
yes		https://w...	Musician/...		3	0.01	1363	3	7.0	0.182228	0.197202
yes		https://w...	Sports Tea...	MagicSistas...	3	0.01	266	3	7.0	0.182153	0.196078
yes		https://w...	Photograp...	MirkosFotos	13	0.01	5817	2	5.0	0.223058	0.236816
yes		https://w...	Media/Ne...	MEDIABERL...	37	0.01	1181	3	7.0	0.182228	0.197202
yes		https://w...	Musician/...	FeliciaLuKu...	104	0.01	2671	3	0.0	0.0	0.0
yes		https://w...	Musician/...	Bwoodwar...	21	0.01	953	3	0.0	0.0	0.0
yes		https://w...	Musician/...	officialches...	271	0.0	287847	3	1.0	1.0	1.0
yes		https://w...	Musician/...	MaxSchnei...	10585	0.04	939324	3	0.0	0.0	0.0
yes		https://w...	Musician/...	KurtHugoS...	12220	0.06	1375928	3	1.0	1.0	1.0
yes		https://w...	Musician/...	SamTsuiMu...	12305	0.02	1615893	3	0.0	0.0	0.0
yes		https://w...	Musician/...	TheFray	9019	0.02	4315851	3	0.0	0.0	0.0
yes		https://w...	Musician/...	TylerWard...	3867	0.01	819993	3	0.0	0.0	0.0
yes		https://w...	Musician/...	gootmusic	1784	0.0	1219224	3	0.0	0.0	0.0
yes		https://w...	Company	EFDDeutschl...	1470	0.06	66571	3	0.0	0.0	0.0
yes		https://w...	Cause	newdoo clips	1	0.0	998	3	0.0	0.0	0.0
yes		https://w...	Company	you.messe	7	0.02	11613	3	6.0	0.254722	0.272622
yes		https://w...	Local Busi...		3	0.0	791	3	1.0	1.0	1.0

Spalten  
schmelzen

Spalte  
löschen

Spalte  
leeren

Daten in andere  
Spalte kopieren

Spalte mit einem  
Wert füllen

Spalte  
duplizieren

Erzeuge Spalte vom Typ  
Boolean aus Regex-Treffern

Erzeuge Spalte mit Liste der den regulären  
Ausdrücken entsprechenden Teilausdrücke

Wahrheitswerte  
negieren

## 3.2. DATENLABOR > FUNKTIONEN

- Separate Ansicht von Nodes und Edges
- Zahlreiche Möglichkeiten zur Manipulation des Datensatzes



# 3.3. VORSCHAU

walten

	1.0
	custom [0,0,0]
	100.0
	<input type="checkbox"/>
	Arial 12 Normal
	<input checked="" type="checkbox"/>
	custom [0,0,0]
	30
	0.0
	custom [255,255,255]
	80.0
	<input type="checkbox"/>
	parent
	100.0
	<input checked="" type="checkbox"/>
	1.0
	<input type="checkbox"/>
	mixed
	100.0
	<input checked="" type="checkbox"/>
	0.0
	3.0
	<input type="checkbox"/>
	Arial 10 Normal
	original
	<input type="checkbox"/>

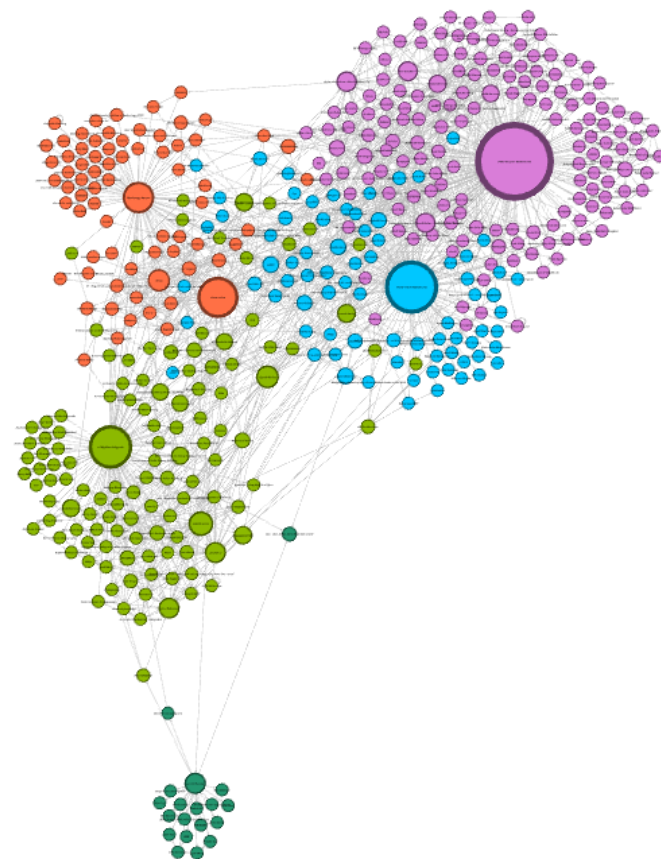
Aktualisieren

Graph x

Ziehen (Konfigurieren)



Dialog.bold, 32



### 3.3. VORSCHAU > FUNKTIONEN

- Dient allgemein zum optischen Feinschliff, v.a. mit Blick auf den Export als Bild- oder Videodatei
- Anwendung von Voreinstellungen möglich

## 4. TUTORIALS UND WIKIS

- Offizielles Gephi-Wiki:  
<https://github.com/gephi/gephi/wiki>
- Offizieller Quick-Guide (pdf):  
[https://gephi.org/tutorials/gephi-tutorial-quick\\_start.pdf](https://gephi.org/tutorials/gephi-tutorial-quick_start.pdf)
- Visualisieren eines Facebook-Netzwerks:  
<https://www.youtube.com/watch?v=N3yv5E-hjbc>
- Visualisierungen in der Linguistik:  
<http://linguistik.zih.tu-dresden.de/lehre/blogs/visualisierungen/>

# 5. LITERATUR

- Bastian, Mathieu; Heymann, Sebastien; Jacomy, Mathieu: Gephi. An open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. 2009.

## USED SOFTWARE

- Präsentation: [reveal.js](#)
- Gephi-Container: [DIT4C](#)