

# Mit neuen Methoden zu neuen Aussagen: Semantische Netzwerkanalyse am Beispiel der Europäischen Verfassung

Haiko Lietz<sup>1</sup>

*Universität Duisburg-Essen, Fachbereich Gesellschaftswissenschaften, Institut für Soziologie  
Hochschule Mittweida, Fachbereich für Mathematik/Physik/Informatik*

(Zuletzt bearbeitet am 05.09.07)

**Zusammenfassung:** Typischerweise werden bei der Inhaltsanalyse von Texten Häufigkeiten gesuchter Konzepte bestimmt. Dabei werden Konzepte jedoch von ihrem jeweiligen Kontext befreit. Die semantische Netzwerkanalyse betrachtet Texte als komplexe semantische Systeme, deren Aussagen sich mit graphentheoretischen Verfahren ermitteln und visualisieren lassen. Demnach bilden sich Makroaussagen aus Mikroaussagen wie Sätzen und Teilsätzen heraus. Statt der Isolierung von Konzepten beruht die Analyse auf semantischer Einbettung. Am Beispiel der Europäischen Verfassung wird gezeigt, dass sich die Hauptaussage eines Texts oder einer Textmenge sinnvoll in einer einzelnen Grafik abbilden lässt. Graphentheoretische Verfahren teilen Konzepte sinnvoll in Gruppen ein, in denen Konzepte unterschiedliche Rollen spielen. Die semantische Netzwerkanalyse ist somit eine viel versprechende neue Methode, um komplexe Informationen verschiedenster Arbeitsbereiche aufzubereiten und zu präsentieren.

**Stichworte:** Informationsmanagement, Informationsvermittlung, semantische Netzwerkanalyse, komplexe semantische Systeme, Europäische Verfassung

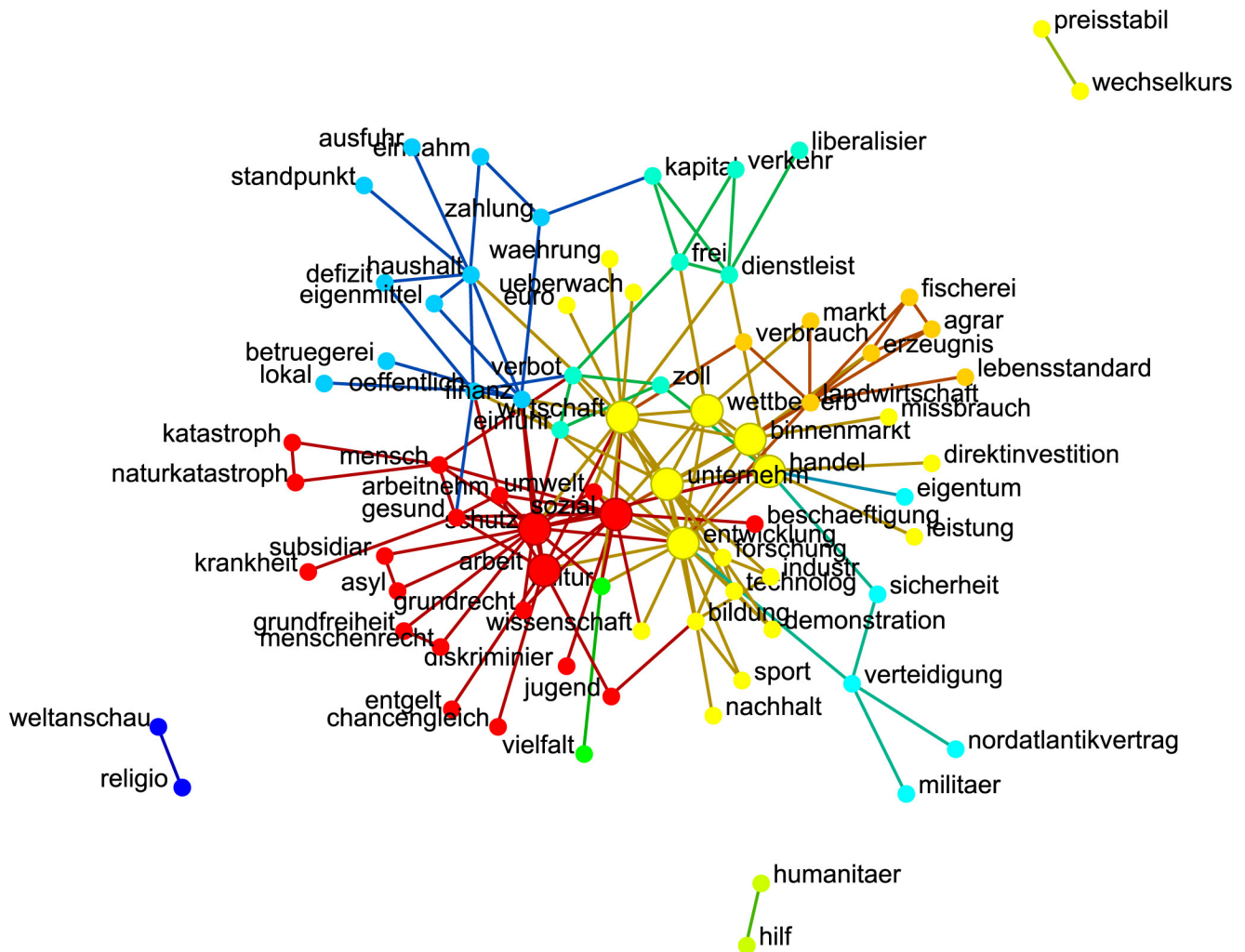
## Einführung

Die Siliziumrevolution hat enorme Datenmengen erfass- und verarbeitbar gemacht. Die Internetrevolution macht sie ortsungebunden zugänglich. Große Daten- und Informationsmengen stellen einen vor eine doppelte Herausforderung. Die erste Herausforderung ist eine des Informationsmanagements: wie finde ich das Signal im Rauschen? Die Nutzinformation muss erst gefunden und aus der Datensee gehoben werden. Dabei können unmöglich alle Informationen gesichtet und studiert werden. Die gesuchte Methode ist also einerseits ein *Data Mining*-Verfahren. Die zweite Herausforderung ist eine der Informationsvermittlung: wie vermittele ich die gefundene Information? Komplexe Informationen sollen effektiv und ansprechend konsumierbar sein. Große Datenmengen sind numerisch schlecht verdaubar und lassen viel Raum für Interpretationen. Die gesuchte Methode ist also auch ein Visualisierungsverfahren.

Übliche Methoden wie die quantitative oder qualitative Inhaltsanalyse ordnen Wörter in Texten Konzepten zu und zählen die Häufigkeit der Konzepte, die dann in Säulen- und Kuchendiagrammen kommuniziert werden. Dabei geht jedoch der Kontext verloren, Wörter stehen nicht mehr in Sinn gebenden Zusammenhängen. Konventionelle Inhaltsanalysen drehen Texte durch den Hexler und vermitteln atomisierte, von Kontext befreite Information.

---

<sup>1</sup> [hlietz@htwm.de](mailto:hlietz@htwm.de)



**Graph 1: Europäische Verfassung auf einen Blick (Präambel und Teile I-III)**

Nichttriviale semantische Aussagen entstehen aus dem Zusammenspiel von Wörtern. Sätze und Texte können also als Wortnetzwerke betrachtet werden. Die semantische Netzwerkanalyse ist eine Methode zur Extrahierung und Visualisierung von Informationen aus Texten unter Berücksichtigung der Wortrelationen. Wenn man Texte als *komplexe semantische Systeme* betrachtet, lassen sich Aussagen als emergente Texteneigenschaften verstehen. Komplexität drückt generell die Vielzahl der Systemzustände aus, die aus der Interaktion der Systemkomponenten resultieren können. Emergenz ist dann die Herausbildung einer Systemeigenschaft, die nicht in den Komponenten selbst begründet liegt, sondern erst aus der Interaktion entsteht. Die Gesamtaussage eines Textes oder einer Sammlung von Texten kann also als emergente Eigenschaft der semantischen Netzwerke modelliert werden.

Nehmen wir zum Beispiel den Text: Das Wetter ist schlecht. Die Textaussage ist nicht in den Systemkomponenten Wetter und schlecht enthalten, sondern entsteht erst aus der Interaktion der beiden Wörter, ist somit eine emergente Eigenschaft. Systemeigenschaften wie Makroaussagen lassen sich vermitteln, indem man das semantische Netzwerk visualisiert. Dieser Prozess der Informationserhebung und -vermittlung wird nun am Beispiel des Vertrags über eine Verfassung für Europa (2004) beispielhaft demonstriert. Ziel ist es, den Inhalt der Verfassung in einer einzelnen Grafik zu

kommunizieren. Diese Grafik ist Graph 1.<sup>2</sup> Der kurzen Beschreibung der Daten folgt die Beschreibung der Methode mit der *AutoMap*-Software. Wer möchte, kann diesen Teil überspringen und bei der Schilderung der Ergebnisse weiter lesen. Abschließend wird ein allgemeines Fazit gezogen.

## Daten

Der Vertrag über eine Verfassung für Europa ist am 29. Oktober 2004 von den Regierungschefs der Europäischen Mitgliedstaaten unterzeichnet worden und sollte der Europäischen Union eine einheitliche Struktur und Rechtspersönlichkeit geben. Da die Ratifikation an Volksabstimmungen in Frankreich und den Niederlanden scheiterte, erlangte der Vertrag bis heute keine Rechtsgültigkeit.

Der Vertrag besteht aus der Präambel (eine Seite ohne Namen), Teil I „Definition und Ziele der Union“ (30 Seiten), Teil II „Die Charta der Grundrechte der Union“ (14 Seiten), Teil III „Die Politikbereiche und die Arbeitsweise der Union“ (131 Seiten), Teil IV „Allgemeine und Schlussbestimmungen“ (sechs Seiten ohne Unterschriften) und 36 Protokollen und fünf Anhängen. Die Daten wurden von EUR-Lex als TXT-Dateien gespeichert und so bearbeitet, dass ein einzelner Artikel jeweils eine Zeile ausfüllt (kein Zeilenumbruch). Analysiert werden nur die Präambel (ohne Namen der Regierungschefs) und die ersten drei Teile. Teil IV und die Protokolle und Anhänge werden nicht analysiert, da sie den Inhalt der Verfassung nicht tangieren und zu allgemein oder länderspezifisch sind.<sup>3</sup>

## Methode

Die vier, den Inhalt der Verfassung betreffenden Texte werden mit der *AutoMap*-Software analysiert.<sup>4</sup> Das Verfahren wird beispielhaft am ersten Satz der Präambel erläutert. Bevor er eingelesen wird, werden Umlaute und ß ersetzt:

SCHOPFEND aus dem kulturellen, religiösen und humanistischen  
Erbe Europas, aus dem sich die unverletzlichen und  
unverausserlichen Rechte des Menschen sowie Freiheit,  
Demokratie, Gleichheit und Rechtsstaatlichkeit als universelle  
Werte entwickelt haben,

In *AutoMap* werden zuerst Zeichen und Zahlen entfernt:

SCHOPFEND aus dem kulturellen religiösen und humanistischen  
Erbe Europas aus dem sich die unverletzlichen und  
unverausserlichen Rechte des Menschen sowie Freiheit  
Demokratie Gleichheit und Rechtsstaatlichkeit als universelle  
Werte entwickelt haben

Dann werden Wörter auf Wortstämme zurückgeführt:

SCHOPFEND aus dem kulturell religiös und humanist Erbe Europas  
aus dem sich die unverletz und unverauss Recht des Mensch sowi  
Freiheit Demokrati Gleichheit und Rechtsstaat als universell  
Wert entwickelt hab

Dann werden Artikel, Pronomen, Konjunktionen und Präpositionen entfernt:

---

<sup>2</sup> Eine visuell aufbereitete Version der Grafik hängt diesem Dokument an.

<sup>3</sup> <http://eur-lex.europa.eu/de/treaties/dat/12004V/hm/12004V.html>

<sup>4</sup> Diese Software wird von Jana Diesner am CASOS-Labor der *Carnegie Mellon University* entwickelt und ist frei verfügbar unter: <http://www.casos.cs.cmu.edu/projects/automap/index.html>

## Mit neuen Methoden zu neuen Aussagen

SCHOPFEND kulturell religios humanist Erbe Europas unverletzt  
unverauss Recht Mensch Freiheit Demokrati Gleichheit  
Rechtsstaat universell Wert entwickelt hab

Als nächstes wird ein Generalisierungsthesaurus geschrieben, der diejenigen Wörter enthält, die aus dem Text extrahiert werden sollen. Wie bei der konventionellen Inhaltsanalyse können und werden verschiedene Wörter zu einheitlichen Konzepten zusammengefasst. Die Wörter arbeitsmarkt, arbeitsplatz, arbeitsumwelt, etc. werden zum Beispiel dem Konzept arbeit zugeordnet. In den vier Texten werden vier verschiedene Konzeptklassen identifiziert, deren semantische Kombination die Analyse verschiedener Netzwerkklassen ermöglicht. In der folgenden Metamatrix sind die vier Konzeptklassen und die sich daraus ergebenden Netzwerkklassen dargelegt:

|           | Inhalt          | Akteur            | Beziehung             | Prozess             |
|-----------|-----------------|-------------------|-----------------------|---------------------|
| Inhalt    | Inhaltsnetzwerk | Policy-Netzwerk   | Kontext-netzwerk      | Umsetzungs-netzwerk |
| Akteur    |                 | Soziales Netzwerk | Interaktions-netzwerk | Handlungs-netzwerk  |
| Beziehung |                 |                   | Beziehungs-netzwerk   | Verfahrens-netzwerk |
| Prozess   |                 |                   |                       | Prozess-netzwerk    |

### Metamatrix 1

Nach Identifikation der Konzepte durch den Generalisierungsthesaurus sieht der Beispielsatz so aus:

kultur religio erbe europa recht mensch freiheit demokrat  
gleich rechtsstaat wert

Mit Hilfe eines Metamatrixthesaurus, der Konzepte Klassen zuordnet, werden Konzeptklassen im Text identifiziert:

<Inhalt> <Inhalt> <Inhalt> <Prozess> <Inhalt> <Inhalt>  
<Inhalt> <Beziehung> <Inhalt> <Inhalt>

Dem Ziel gemäß, den Inhalt der Verfassung in einer einzelnen Grafik zu kommunizieren, werden hier nur Inhaltsnetzwerke diskutiert (siehe Metamatrix). Der Beispielsatz würde somit auf folgende Konzepte reduziert:

kultur religio erbe mensch freiheit demokrat rechtsstaat wert

Zur Erstellung des Netzwerks wird eine Fenstergröße von 9 angesetzt. Das heißt, zwei Konzepte werden miteinander in Verbindung gesetzt, wenn nicht mehr als acht Wörter zwischen ihnen stehen. Fenster reichen nicht über Absätze hinaus. Deswegen sind alle Verfassungsartikel auf einzelne Absätze formatiert worden. Bei der verwendeten Fenstergröße von 9 wird der Beispielsatz also als vollständig verbundenes Netzwerk dargestellt, in dem jedes der acht Konzepte mit jedem anderen verbunden ist.

## Ergebnisse

Der Generalisierungsthesaurus identifiziert 665 Wörter, die mindestens zweimal in allen Texten vorkommen, und ordnet sie 461 Konzepten zu. Diese teilen sich auf vier Klassen auf. 178 Konzepte gehören zur analysierten Klasse „Inhalt“. Zwischen diesen 178 Konzepten gibt

es 3670 Beziehungen der maximalen semantischen Distanz 9. Das Netzwerk in Graph 1 ist reduziert auf Beziehungen der Mindeststärke 7. Das heißt, alle Konzeptpaare, die bis zu sechsmal in allen Texten erscheinen, werden nicht dargestellt. Es bleiben 79 Konzepte und 347 Beziehungen. Wie bei einer Bergkette, die nach und nach überflutet wird bis nur noch ihre Spitzen aus dem Wasser ragen, sieht man nur die häufigsten Konzepte und Konzeptbeziehungen. Das Netzwerk repräsentiert die Hauptaussagen der Texte.

Die Farben der Knoten (Konzepte) kennzeichnen Gruppen von Konzepten, die häufig miteinander Erwähnung finden.<sup>5</sup> Betrachtet man diese Gruppen gesondert und charakterisiert sie anhand des zentralsten Konzepts – das innerhalb die meisten Beziehungen hat –, erhält man die Gruppen der Hauptkonzepte in Übersicht 1.

|                           |                      |                    |                       |
|---------------------------|----------------------|--------------------|-----------------------|
| <b><u>entwicklung</u></b> | <b><u>schutz</u></b> | <b>haushalt</b>    | <b>landwirtschaft</b> |
| wirtschaft                | sozial               | ausfuhr            | lebensstandard        |
| unternehmen               | arbeit               | betruegerei        | markt                 |
| <u>innenmarkt</u>         | grundfreiheit        | standpunkt         | verbrauch             |
| wettbewerb                | jugend               | einnahm            | agrar                 |
| <u>handel</u>             | arbeitnehm           | oeffentlich        | fischerei             |
| eigentum                  | beschaeftigung       | zahlung            | erzeugnis             |
| demonstration             | chancengleich        | eigenmittel        |                       |
| nachhalt                  | entgelt              | defizit            | <b>verteidigung</b>   |
| wahrung                   | gesund               | finanz             | nordatlantikvertrag   |
| leistung                  | asyl                 | lokal              | sicherheit            |
| ueberwach                 | krankheit            |                    | militaer              |
| missbrauch                | menschenrecht        | <b>dienstleist</b> |                       |
| bildung                   | diskriminier         | verkehr            |                       |
| direktinvestition         | grundrecht           | verbot             |                       |
| wissenschaft              | subsidiar            | einfuhr            |                       |
| technolog                 | mensch               | kapital            |                       |
| forschung                 | katastroph           | frei               |                       |
| sport                     | umwelt               | zoll               |                       |
| euro                      | naturkatastroph      | liberalisier       |                       |
| industr                   |                      |                    |                       |

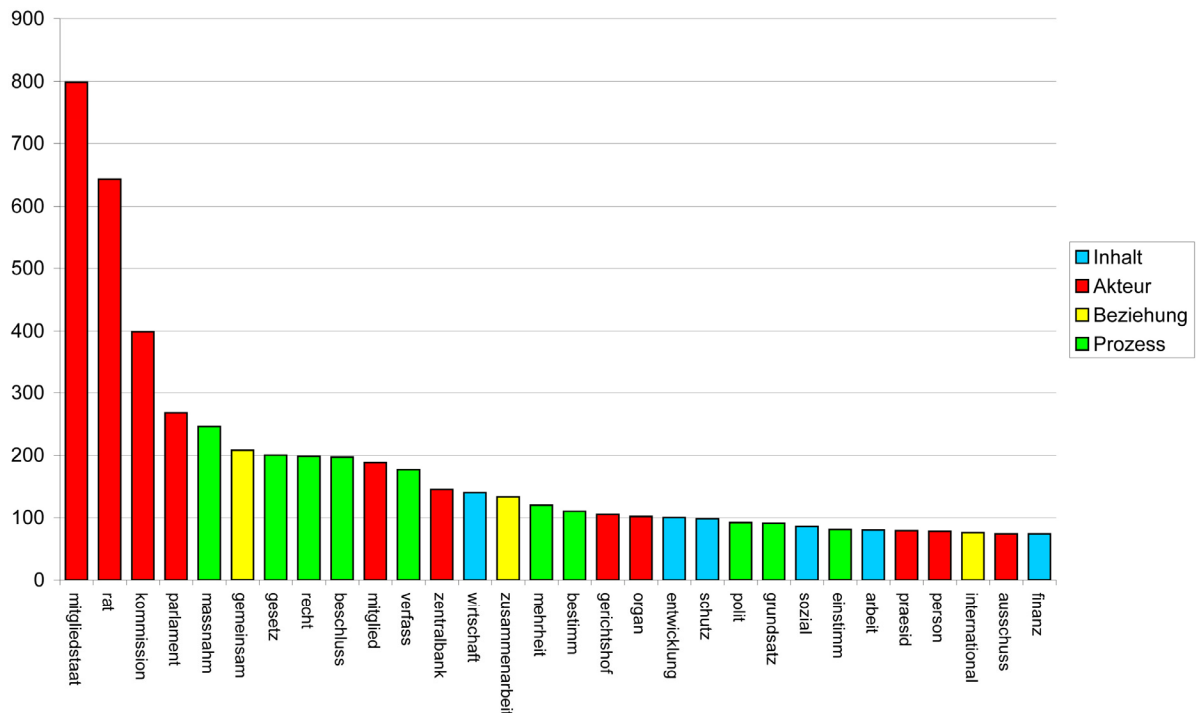
**Übersicht 1: Gruppen der Hauptkonzepte (Mittlerkonzepte sind unterstrichen)**

Große Knoten im Graphen stellen Konzepte dar, die zentrale Bedeutung in den Texten haben. Diese Konzepte, unterstrichen in Übersicht 1, verbinden verschiedene Konzeptgruppen. Es ist auffällig, dass Mittlerkonzepte nur den Gruppen „entwicklung“ und „schutz“ angehören und in ersterer sämtlich wirtschaftlichen Inhalts sind.

Der Mehrwert, den die relationale Analyse gegenüber der einfachen Häufigkeitsauszählung schafft, liegt also darin, dass Konzepte in ihrem Kontext gesehen werden. Eine nichtrelationale Inhaltsanalyse würde nur die in Diagramm 1 gezeigten Häufigkeiten schildern, nicht aber deren Gruppenzugehörigkeiten und Rollen im Gesamttext.

Die semantische Netzwerkanalyse beruht auf der Annahme, dass sich ein Text oder eine Menge von Texten eine nicht bewusst formulierte – in der Sprache der Komplexitätstheorie: selbstorganisierte, emergente – Gesamtaussage besitzen. Graph 1 verbildlicht diese. Ein Blick auf die Graphen der einzelnen Teile des Verfassungstexts offenbart, welcher Teil welchen Beitrag zur Gesamtaussage spielt (siehe Anhang). So sind „verteidigung“, „sozial“ und „wirtschaft“ zentrale Konzepte in den Definition und Zielen der Union (Teil I). Bemerkenswert ist, wie sinnvoll der Algorithmus Konzepte in Gruppen aufteilt. Dem Inhalt von Teil II gemäß sind „verbot“, „schutz“, „arbeit“ und „religio“ zentrale Konzepte der Grundrechte der Union, die die Gruppen, in denen sie eingebettet sind, sinnvoll charakterisieren. Der Graph für Teil III sagt aus, dass Politikbereiche oft wirtschaftlich assoziiert sind.

<sup>5</sup> Verwendet wurde der Clustering-Algorithmus von Newman. Zur Visualisierung wurde der *Network Visualizer* verwendet, der *AutoMap* beiliegt.



**Diagramm 1: Häufigkeiten der 30 häufigsten Konzepte**

### Zusammenfassung und Ausblick

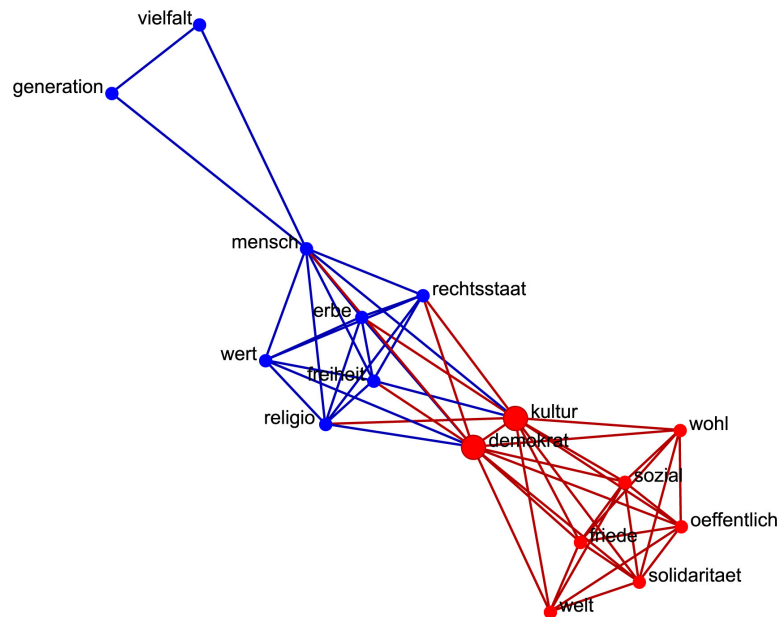
Texte besitzen Makroaussagen, die sich aus dem Zusammenspiel und dem gegenseitigen Verstärken von Sätzen und Teilsätzen (Mikroaussagen) herausbilden. Am Beispiel des Vertrags über eine Verfassung für Europa wurde gezeigt, dass sich die Hauptaussage eines Texts sinnvoll in einer einzelnen Grafik abbilden lässt. Mit graphentheoretischen Verfahren lassen sich Konzepte Gruppen zuordnen, die sich als unterschiedliche Kontexte interpretieren lassen, in denen verschiedene Konzepte unterschiedliche Rollen spielen. Für das Beispiel konnte gezeigt werden, dass die Europäische Verfassung stark den Charakter eines Wirtschaftsvertrags hat.

Indem Kontext gewahrt wird, geht die semantische Netzwerkanalyse über bloße Analysen von Konzepthäufigkeiten hinaus. Sie ist eine neue Methode, um Kernaussagen und Strukturen aus beliebig großen Textmengen – die Grenze wird von der verfügbaren Rechenleistung bestimmt und schnell erreicht – zu extrahieren und darzustellen. Sie ist daher eine viel versprechende Methode, um komplexe Informationen verschiedenster Arbeitsbereiche aufzubereiten und zu präsentieren. Visuell aufbereitete Informationen versprechen eine Zeiteinsparung. Planungsentscheidungen können vorbereitet werden.

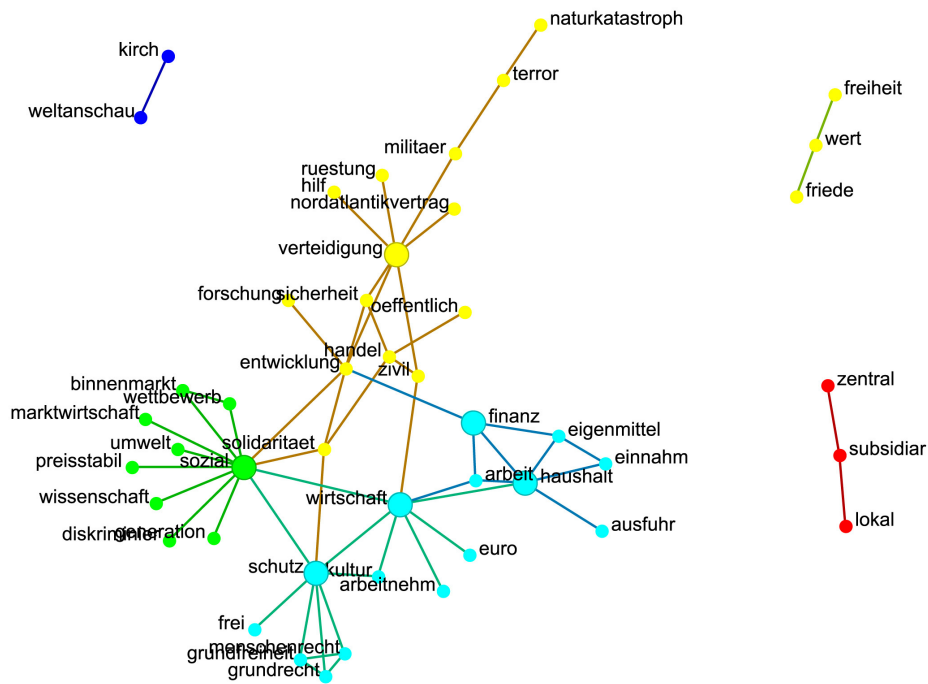
Letztendlich handelt es sich um eine neue Methode der Informationsverarbeitung. Viele Datenbestände sind nie derart analysiert worden. Auch in alten Daten können demnach noch neue Erkenntnisse schlummern.



## Anhang: Graphen einzelner Verfassungsteile

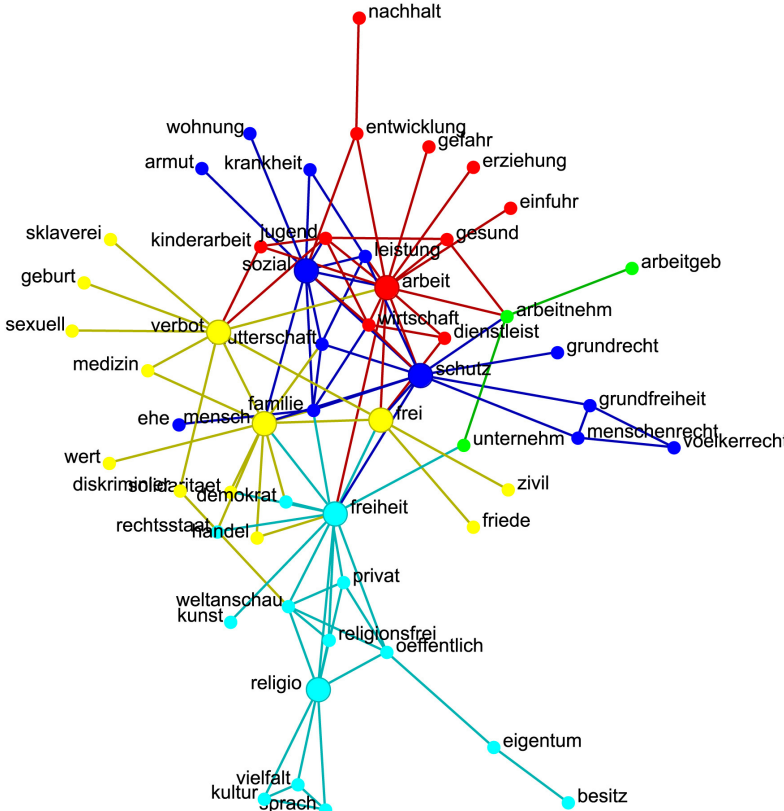


Graph 2: Präambel, alle Nennungen

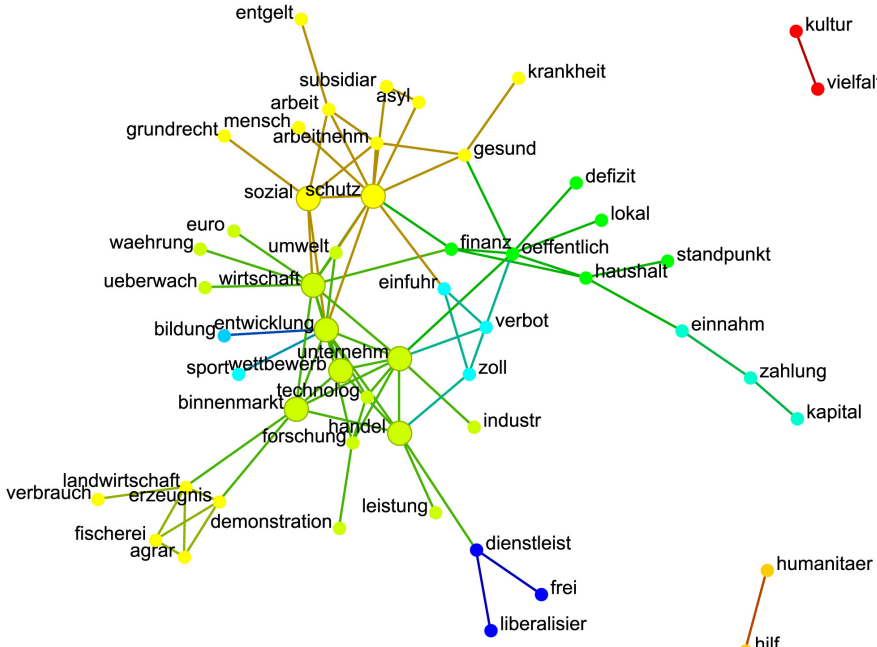


Graph 3: Definition und Ziele der Union (Teil I), mindestens drei gemeinsame Nennungen

## Mit neuen Methoden zu neuen Aussagen



**Graph 4: Grundrechte der Union (Teil II) , mindestens zwei gemeinsame Nennungen**



**Graph 5: Politikbereiche und die Arbeitsweise der Union (Teil III), mindestens acht gemeinsame Nennungen**



# Vertrag über eine Verfassung für Europa

