



Der vorliegende Beitrag erscheint auf Einladung der

*Professoren Barbara Job, Alexander Mehler und Tilmann Sutter im Tagungsband
Interdependenz und Dynamik sozialer und sprachlicher Netzwerke*

Zentrum für Interdisziplinäre Forschung (ZiF), Universität Bielefeld, 28.-30. April 2010.

Der vorliegende Preprint entspricht der beim Verlag eingereichten Version und ist nicht öffentlich.

Autorenkontakt: maximilian@schich.info

Netzwerke komplexer Netzwerke in der (Kunst)Wissenschaft

Maximilian Schich

Der vorliegende Beitrag beschreibt in einfachen Worten die konzeptuelle Relevanz komplexer Netzwerke für die Kunst- und Geisteswissenschaften im Zeitalter explosionsartig wachsender Datenmengen. Dabei wird eine quantitative Perspektive aufgezeigt, die vom Standpunkt der traditionell-qualitativ fokussierten Ausbildung in den Kunstwissenschaften bedrohlich erscheinen mag, jedoch in erweiterter Sicht weitreichende Chancen bietet, sowohl bezüglich des möglichen Fortschritts im Fach selbst, als auch bezüglich einer multidisziplinären Einbindung auf dem Weg zu einer allgemeinen, evidenzbasierten qualitativ wie quantitativ ausgerichteten Wissenschaft der Daten, mit fließenden Übergängen zwischen Kunst-, Geistes-, Sozial-, und Naturwissenschaften.

Wichtig erscheint dabei festzuhalten, dass komplexe Netzwerke und damit verbundene Konzepte hier im Sinne der diskreten Mathematik und der naturwissenschaftlichen Netzwerkforschung verstanden werden (vgl. Newman 2010 oder einführend Barabási 2002). Wichtig ist dies vor allem im Hinblick auf stärker theoretisch geprägte Ansätze wie etwa die so genannte Actor-Network-Theory (vgl. Latour 2005), deren Instrumentarium angesichts der zur Verfügung stehenden Daten ebenso unzureichend erscheint, wie viele andere philosophisch geprägte Theorien, die hier nicht diskutiert werden.

Komplexe Netzwerke

Betrachten wir zunächst eine Standardfrage der Kunstgeschichte. So gut wie jedes kunsthistorische Werk ordnet Objekte in irgendeiner Weise ein, etwa im Sinne einer Künstlerzuweisung, einer Datierung, einer geographischen oder auch nur thematischen Verortung. Prinzipiell lässt sich eine solche Einordnung als Klassifikation verstehen, in der die Objekte die Rolle des zu klassifizierenden Gegenstandes einnehmen sowie die jeweilige Einordnung diejenige des Klassifikationskriteriums. Dabei spielt es keine Rolle, ob das kunsthistorische Werk systematisch beispielsweise hunderte von Gemälden bestimmten fest umrissenen Themen in einem kontrollierten Index zuordnet, oder ob dabei ein einzelner unscharfer Werkzusammenhang exklusiv über hunderte von Seiten einer einzigen schwer umgrenzbaren Bedeutung zugewiesen wird. Zumindest theoretisch handelt es sich in beiden Fällen um den selben Vorgang.

Man könnte sich nun vorstellen eine Summe aller kunsthistorischen Einordnungen dieser Art zu erstellen, etwa indem man alle existierenden kunsthistorischen Werke auf einen großen Haufen wirft, selbstverständlich unter Erhalt der existierenden Multiplizität der Meinung. Jede kunsthistorische Einordnung in dem großen Haufen würde dabei einen Link herstellen zwischen einem Gegenstand und einem Klassifikationskriterium. Da jeder Gegenstand mit mehreren Klassifikationskriterien verbunden sein könnte, und jedes

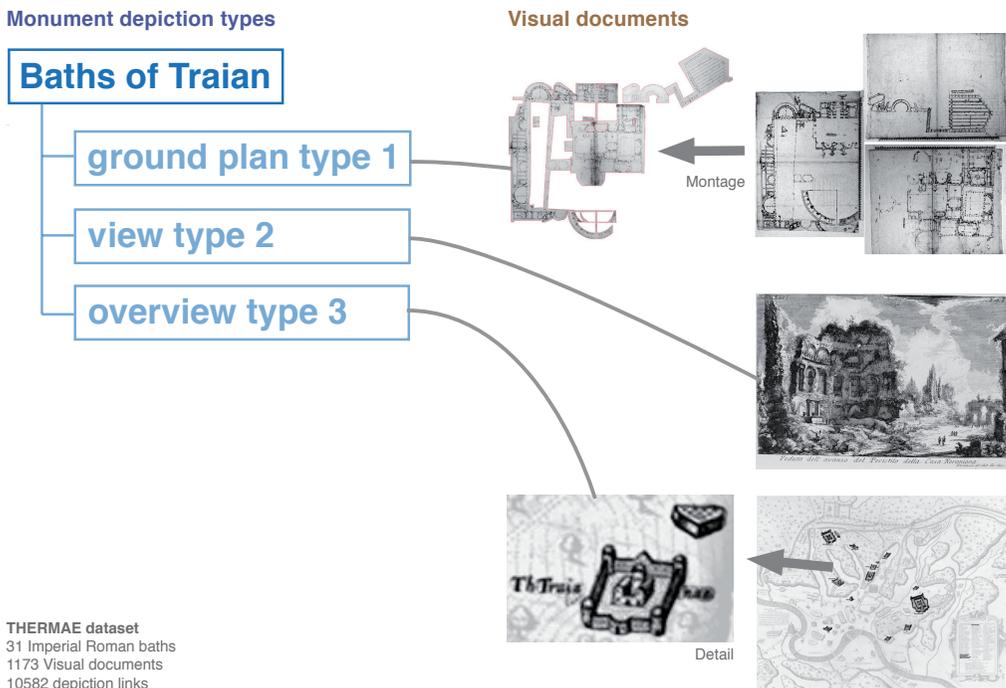


Fig. 1. Beispiel der Rezeption antiker Monumente in visuellen Dokumenten.

Klassifikationskriterium umgekehrt mit mehreren Gegenständen, entstünde dadurch ein Netzwerk, dessen globale Struktur, Dynamik und Evolution – da alles andere als Durchschnittlich und Zufällig – sich als lohnender Untersuchungsgegenstand anbietet.

Betrachten wir ein relativ einfaches, konkretes Beispiel der kunsthistorischen Einordnung – das Phänomen der Rezeption antiker Monumente in visuellen Dokumenten, so wie etwa im „Census of Antique Works of Art and Architecture Known in the Renaissance“ seit 1947 dokumentiert. Visuelle Dokumente entsprechen dabei dem Gegenstand der Klassifikation, antike Monumente den Klassifikationskriterien (oder umgekehrt). Unabhängig von der Perspektive ergibt sich ein Netzwerk mit zwei Knotenarten, den visuellen Dokumenten und den antiken Monumenten, sowie einer Linkart, dem Rezeptions- oder Klassifikationslink.

Figur 1 zeigt einen kleinen Ausschnitt aus einem solchen Netzwerk der Rezeption. Visuellen Dokumenten zu römischen Kaiserthermen zur Rechten stehen dabei Abbildungstypen antiker Monumente zur Linken gegenüber (vgl. Schich 2009). Bei den Dokumentknoten handelt es sich dabei um einfache Veduten, um Montagen aus Zeichnungsfragmenten oder um Ausschnitte aus einem Stadtplan. Die Abbildungstypen werden im Beispiel durch eine einfache Hierarchie von Verbalkonzepten repräsentiert. Rezeptionslinks erscheinen als graue Linien zwischen den visuellen Dokumenten und den verbalen Konzepten. Wollte man das Beispiel weiter vereinfachen, so könnte man sich auch Photos als Dokumente sowie einfache Tags oder Schlagworte als Klassifikationskriterien vorstellen. Die Links wären auch in diesem Fall als graue Linien zwischen Gegenstand und Klassifikationskriterien denkbar.

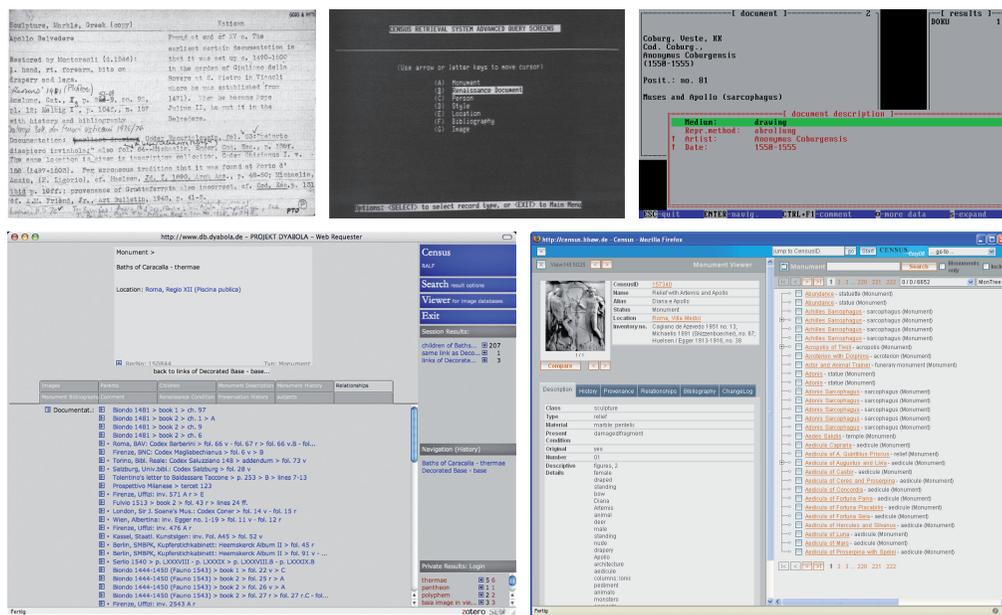


Fig. 2. Evolution der Benutzeroberfläche einer kunstwissenschaftlichen Datenbank

Interessant ist hier zunächst, wie ein solches Klassifikationsnetzwerk traditionell in der Kunstwissenschaft abgebildet und erschlossen wird. Prinzipiell ist man zumeist gezwungen über Suchformulare oder Listen in die Klassifikation einzusteigen, worauf man diese von Knoten zu Knoten navigiert. Nach wie vor zeigen kunstwissenschaftliche Datenbanken beispielsweise Listen von Gegenständen beim Klassifikationskriterium sowie Listen von Klassifikationskriterien beim Gegenstand – analog zum herkömmlichen Museumskatalog mit Index auf Papier, oder zum doppelten Zettelkasten. Figur 2 zeigt die Evolution entsprechender Benutzeroberflächen für den Fall des oben bereits erwähnten Census von 1947 bis 2006. Figur 3 zeigt den Standard als Karikatur: Man befindet sich entweder beim Dokument (links) und sieht eine Liste von darin vorkommenden Monumenten, oder beim Monument (rechts) mit einer Liste von entsprechenden Dokumenten. Was man beispielsweise nicht sieht ist eine Übersicht von Dokumenten die sich aufgrund einer Anzahl darin erscheinender Monumente ähneln. Will man als Benutzer die Frage beantworten, in welchen Dokumenten die aufgelisteten Monumente beim Dokument X sonst noch vorkommen, so ist man gezwungen langwierig in der Datenbank (entlang einzelner Links wie dem einzelnen grauen in Figur 3) zu navigieren. Im Standardfall gibt es derzeit keine globale Kartierung des Klassifikationsnetzwerks anhand dessen sich unsere Frage schnell und einfach beantworten ließe.

Glücklicherweise bedeutet die Definition unserer Klassifikation als Netzwerk in dieser Sache einen Lichtblick. Aus der Graphentheorie, einem Zweig der Mathematik, lernen

Document – London, RIBA: vol. VII, fol. 7 r
 Author: Giovannantonio Dosio – Date: 1562
 RecNo 4322

Monuments

- Aqua Claudia - aqueduct
- Arch of Constantine - triumphal arch
- Arch of Gallienus - honorific arch
- Arch of Septimius Severus - triumphal arch
- Arch of Titus - triumphal arch
- Aurelian Walls - wall
- Baths of Caracalla - thermae
- Baths of Constantine - thermae
- Baths of Diocletian - thermae
- Baths of Titus - thermae
- Baths of Traian - thermae**
- Colosseum - amphitheatre
- Column of Marcus Aurelius - honorific column
- Column of Trajan - honorific column
- Domus Augustana - palace
- Ianus Quadrifrons - quadrifrons
- Mausoleum of Augustus - mausoleum
- Mausoleum of Hadrian - sepulchral monument
- Pantheon - temple
- Porta Maggiore - gateway
- Pyramid of C. Cestius - pyramid
- S. Croce in Gerusalemme - church
- Septonium - facade
- S. Giovanni in Laterano - church
- Temple of Castor and Pollux - temple
- Temple of Hercules Victor - temple
- Temple of Sol - pavilion

Monument – Baths of Traian
 Location: Roma, Regio III, Thermae Traianae
 RecNo 152816

Documents

- Albertini 1510 (Opusculum), book 1, ch. 07, fol. F v, lines 15-18
- Albertini 1510 (Opusculum), book 1, ch. 07, fol. Fii r, lines 08-15
- Berlin, SMBPK, Kunstbibliothek: inv. HdZ 4151, fol. 021 r, C
- Berlin, SMBPK, Kupferstichkabinett: Heemskerck Album I, fol. 69 v, A
- Berlin, SMBPK, Kupferstichkabinett: Heemskerck Album II, fol. 53 v, B
- Berlin, SMBPK, Kupferstichkabinett: Heemskerck Album II, fol. 91 v - fol. 92 r
- Biando 1444-1450 (Fauno 1543), book 2, fol. 25 r, A
- Biando 1444-1450 (Fauno 1543), book 2, fol. 27 v, fol. 27 v.D - fol. 28 r.A
- Biando 1444-1450 (Fauno 1543), book 2, fol. 30 v, B
- Biando 1481, book 2, ch. 001, A
- Biando 1481, book 2, ch. 012-013
- Chicago, Univ.Lib.: Speculum Romanae Magnificentiae, no. B 242
- Delle miracole di Roma prima della meta e dello chastello, fol. 42v, l. 11-19
- Delle miracole di Roma prima della meta e dello chastello, fol. 49r, l. 01-06
- Delle miracole di Roma prima della meta e dello chastello, fol. 52r, l. 19-26
- Du Perac 1575, fol. 17
- Du Perac 1575, fol. 18
- Edificazione di molti palazzi e tempi di Roma, fol. 10 v, l. 17-18
- Ferrara, Bibl. Com. Aristotea: MS. I 217, fol. 66 r
- Firenze, Bibl. Medicea-Laurenziana: Queste sono le miracole de Roma, fol. 44 r, l. 17-33
- Firenze, Bibl. Medicea-Laurenziana: Queste sono le miracole de Roma, fol. 47 v, l. 13-31
- Firenze, Bibl. Medicea-Laurenziana: Queste sono le miracole de Roma, fol. 49 v, l. 05-20
- Firenze, Bibl. Medicea-Laurenziana: Redi 77, Map of Rome by Alessandro Strozzi
- Firenze, BNC: Le Antiquita della alma Roma, fol. 10v, l. 17-18
- Firenze, BNC: Le Antiquita della alma Roma, fol. 11v, l. 01-06
- Firenze, BNC: Le Antiquita della alma Roma, fol. 18v, l. 10-14
- Firenze, Uffizi: inv. 1537 A v
- Firenze, Uffizi: inv. 2523 A r
- Fulvio 1527, book 3, fol. XXXVI v, B
- London, RIBA: vol. IV, fol. 1 r
- London, RIBA: vol. IV, fol. 2 r
- London, RIBA: vol. IV, fol. 3 r, D
- London, RIBA: vol. VII, fol. 7 r
- New York, Morgan Lib., Feltrinelli Coll.: Du Perac Sketchbook, fol. 11 r, C
- Palermo, Biblioteca Comunale: Libro d'Arabeschi, fol. 132
- Palermo, Biblioteca Comunale: Libro d'Arabeschi, fol. 150
- Paris, Fondation Custodia / Frits Lugt Collection: inv. 6006
- Paris, Musée du Louvre: Hieronymus Cock, Thermae Diocletiani imp., 1558, part I, p. 1, C
- Real Monasterio El Escorial: Codex Escorialensis, fol. 40 v
- Roma, Città del Vaticano, BAV: Descriptio Urbis Romae, ch. 05, fol. 2 v, l. 21.D - l. 22.A

Fig. 3. Karikatur einer herkömmlichen Benutzeroberfläche

tenbank und als Karikatur in Figur 3 finden, ist nichts anderes als ein äußerst begrenzter Ausschnitt aus der Adjazenzliste des gesamten Netzwerks aller möglichen Einordnungen.

Figur 5 zeigt eine Alternative: Das Netzwerk der Rezeption ist hier als eine mit Bildern angereicherte Matrix dargestellt. Jedem Monument ist dabei eine Spalte zugeordnet, jedem Dokument eine Zeile. Anstelle des jeweiligen Links erscheint eine Detailabbildung des Dokuments – im Beispiel je ein Ausschnitt aus einem Stadtplan mit einer Abbildung des jeweiligen Monuments. Die beiden Standardlisten aus Figur 3 entsprechen jeweils einer Zeile bzw. einer Spalte in der Matrix, so wie in Figur 5 markiert. Es ist einfach zu erkennen was dem Benutzer einer herkömmlichen kunstwissenschaftlichen Datenbank, oder auch eines traditionellen Ausstellungskataloges, entgeht: Die Matrix liefert den bisher unsichtbaren Kontext. Ohne weiteres wird erkennbar in welchen weiteren Dokumenten die beim markierten Dokument aufgelisteten Monumente sonst noch vorkommen. Die Dokumentenfrage und die Monumentenfrage – aus denen sich im Übrigen die künstliche Spaltung von Kunstgeschichte und Archäologie ergibt – verbinden sich hier auf natürliche Weise in einer einzigen zweidimensionalen Abbildung.

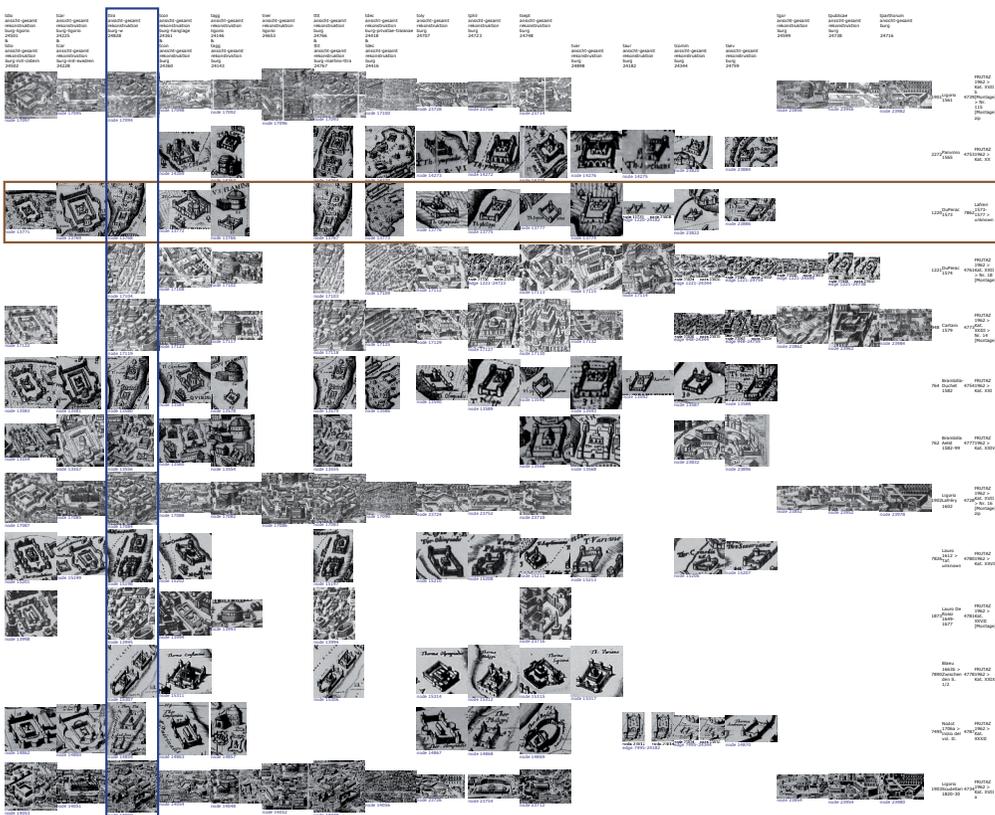


Fig. 5. Bildmatrix mit einem Ausschnitt aus dem Netzwerk der Rezeption. [Markierung entspricht den Spalten in Figur 2].

Es stellt sich die Frage warum solche, ohne Zweifel wünschenswerten, Übersichten nicht längst allgegenwärtig sind? Einen Teil der Antwort liefert ein kurzer Blick auf die Adjazenzmatrix der entsprechenden Gesamtklassifikation in Figur 6, aus der die Übersicht in Figur 5 mit Hilfe eines speziellen Algorithmus (Ahn Bagrow Lehmann 2010) extrahiert wurde. Die Konstruktion der Übersicht ist alles andere als trivial, da die Matrix des entsprechenden Gesamtnetzwerkes schwach besetzt (sparse) ist, so wie im Fall der meisten realen komplexen Netzwerke. Dichte und vor allem sinnvolle Bereiche müssen in der Matrix eigens zusammensortiert (permutiert) werden anhand von Algorithmen, deren Entwicklung inzwischen einen wahrnehmbaren Teil der naturwissenschaftlichen Netzwerkforschung ausmacht (vgl. Porter Onnela Mucha 2009 bzw. Fortunato 2010). Allein die Anzahl der Sortiermöglichkeiten von n Zeilen mal m Spalten der Matrix sprengt angesichts der Möglichkeiten, die bereits im Beispiel von Figur 6 in die Hunderttausende gehen, das kognitive Limit eines jeden einzelnen Forschers. Wählt man ein weniger plakatives Beispiel als in Figur 5, mit Dokumenten aus zahlreichen verschiedenen Quellen, statt einer einzelnen Gattung zu der es einen guten Katalog gibt (Frutaz 1962), so wird der Einsatz des Rechners und vor allem die fachliche Entgrenzung zwischen Kunst- und Naturwissenschaft unumgänglich. Rezeption im speziellen und kunsthistorische Einordnung im Allgemeinen erweisen sich als komplexes Netzwerk im Sinn der diskreten Mathematik und wollen als solche behandelt werden.

Selbstverständlich bedeutet die vorgestellte Sichtweise nicht das Ende der klassischen Kunstwissenschaft. Dennoch sollten wir uns jedes mal, wenn wir in einer Datenbank navigieren, einen Katalog durchblättern, oder auch nur eine einzelne kunsthistorische Einordnung in einem Aufsatz lesen, bewusst machen, dass wir prinzipiell einen winzigen Ausschnitt des Netzwerks aller existierenden kunsthistorischen Einordnungen vor Augen haben. Die Gesamtstruktur, Dynamik und Evolution dieses komplexen Netzwerks kennen wir nur zu sehr geringem Teil. Da in komplexen Systemen mehr immer auch anders ist als erwartet (Anderson 1972), ist die Erforschung des Großen und Ganzen nicht nur ein Desiderat, sondern eine missionskritische Aufgabe für die Kunstwissenschaft.

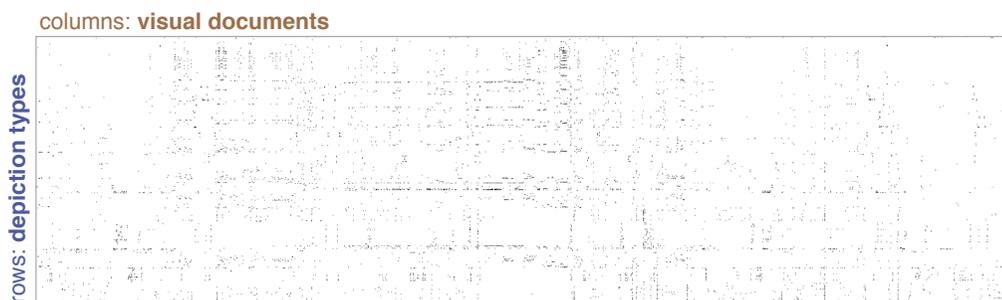


Fig. 6. Matrix des zugehörigen Gesamtnetzwerkes.

Netzwerke komplexer Netzwerke

Sowohl die Summe aller kunsthistorischen Einordnungen im Allgemeinen wie die Rezeption antiker Monumente in visuellen Dokumenten im Speziellen sind als komplexes Netzwerk vorstellbar. Geht diese Vorstellung bereits über den Rahmen der traditionellen Kunstwissenschaft hinaus, so ist sie doch nur ein Teil eines noch größeren Zusammenhanges. Figur 7a zeigt ein Knoten-Link-Diagramm des Netzwerks der Rezeption in einer weiteren kunsthistorischen Datenbank, dem „Corpus der Antiken Denkmäler die J.J. Winkelmann und seiner Zeit bekannt waren“. Für sich gesehen erscheint die Rezeption auch hier eindeutig als komplexes Netzwerk. Im Datenmodell der Datenbank erscheint die Rezeption jedoch nur als eine Linkart zwischen zwei Knotenarten, eingebunden in ein übergeordnetes Netzwerk mehrerer Arten von Links und Knoten. Figur 7b zeigt das gesamte Datenmodell des Corpus im Vergleich. Zusätzlich zur Erkenntnis, dass kunstwissenschaftliche Einordnungen als komplexe Netzwerke vorstellbar sind, kommt hier also die Einsicht hinzu, dass Datenmodelle kunstwissenschaftlicher Datenbanken ebenfalls als Netzwerke zu verstehen sind.

Spinnen wir den oben eingeführten Faden konsequent weiter, so liegt es folglich nahe die Konzeption des Datenmodells als Netzwerk auszunutzen um der tatsächlichen Gesamtstruktur einer Datenbank näher zu kommen. Wie an anderer Stelle für den Census gezeigt (Schich 2010), können wir das Datenmodell einer kunstwissenschaftlichen Datenbank nicht nur als Knoten-Link-Diagramm darstellen, sondern auch als Matrix. Jeder Knotenart wird dabei sowohl eine Zeile als auch eine Spalte zugewiesen. Die jeweiligen Linktypen erscheinen in den entsprechenden Zellen der Matrix, wo die Zeile der Ausgangsknotenart auf die Spalte der Zielknotenart trifft. Figur 8 zeigt das entsprechend dem Census verfeinerte Ergebnis für den Winkelmann Corpus. Wir erkennen, dass nicht nur die Kartierung der Rezeption ein komplexes Netzwerk widerspiegelt, sondern jede einzelne Linkart der Datenbank als komplexes Netzwerk verstanden werden kann. In anderen Worten:

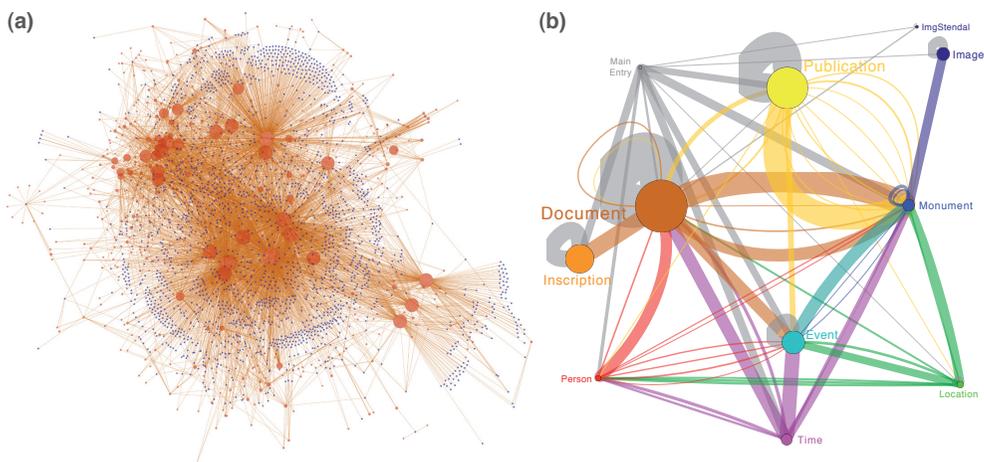


Fig. 7. Netzwerk der Rezeption (a) sowie Datenmodell (b) für den Winkelmann Corpus.

Kunstwissenschaftliche Datenbanken sind Netzwerke von komplexen Netzwerken mit einer Vielzahl nicht-trivialer Verbindungen zwischen Objekten, Personen, Örtlichkeiten, Perioden, Ereignissen und anderen Konzepten – ein weiterer lohnenswerter Untersuchungsgegenstand der Netzwerkforschung.

Bereits ohne weiteren Vergleich mit anderen Datenbanken ergibt sich durch die Kartierung der gesamten Datenbank eine erhebliche Erweiterung des Handlungsspielraums für die kunstwissenschaftliche Forschung. Ohne Weiteres wird in Figur 8 ersichtlich, dass in der Datenbank, neben dem zentralen Anliegen der Rezeption (in der Zelle Document–Monument), zahlreiche weitere interessante Phänomene festgehalten sind. Aus der Sicht der Sozialwissenschaft erscheint die Zelle Person–Location interessant, weil darin offenbar komplexe Bezüglichkeiten zwischen Akteuren, Geburts-, Arbeits- und Todesorten festgehalten sind. Für die Provenienzforschung würde es sich lohnen die Zellen Monument–Event, Event–Location, und Event–Time in einer übergeordneten Studie zu verbinden. Im Hinblick auf die weitere Geschichte der Rezeption selbst ist es sicherlich ertragreich einen Blick auf die moderne Bibliographie zu antiken Monumenten (in der Zelle Publication–Monument) zu werfen, deren Struktur der Rezeption zur Zeit Winkelmanns (in der Zelle Document–Monument) und der Renaissance im Censur in punkto Komplexität in nichts nachsteht (vgl. Schich Barabási 2009). Evaluatoren der Datenbank werden

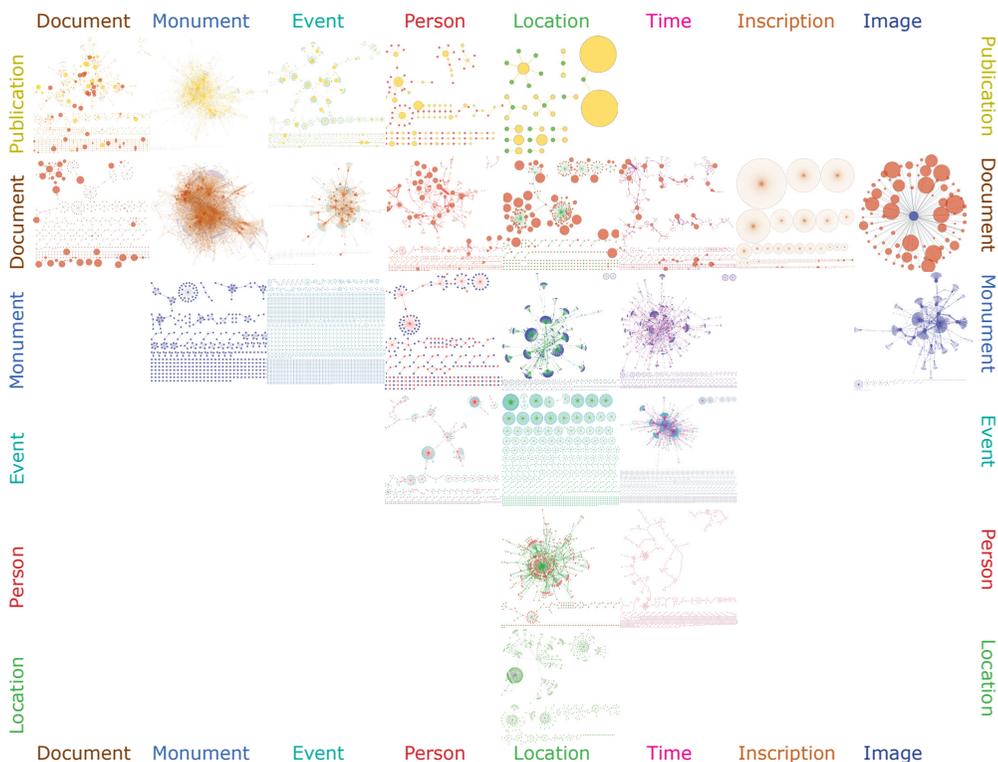


Fig. 8. Mit Knoten-Link-Diagrammen angereicherte Datenmodell-Matrix für den Winkelmann Corpus.

Netzwerke von Netzwerken komplexer Netzwerke

Aus der Sicht eines traditionellen Kunstwissenschaftlers erscheint es sicher utopisch vom großen Haufen aller kunsthistorischen Einordnungen zu träumen. Betrachtet man die Situation jedoch aus einem etwas anderen Blickwinkel, nähern wir uns diesem Ziel gerade mit großen Schritten von zwei Seiten an. Einerseits werden immer mehr strukturierte Datenmengen in maschinenlesbarer Form produziert und veröffentlicht; andererseits arbeiten sowohl Wissenschaft als auch Industrie eifrig daran vorhandene relativ unstrukturierte (Text-)Information in strukturierte Daten zu verwandeln.

Was die Produktion von entsprechend strukturierten Datenmengen angeht, so reicht vermutlich ein Verweis auf Facebook – nicht nur das größte soziale Netzwerk mit über einer Milliarde Benutzern, sondern derzeit auch die größte Bilddatenbank der Welt (Johnson 2010). Die Plattform speichert nicht nur alle Daten in Form einer verteilten Graphdatenbank, d.h. als Netzwerk; sie stellt die Daten mit Hilfe des so genannten Graph API auch als solches zur Verfügung (<https://graph.facebook.com/>). In dem wir Facebook oder ähnliche Plattformen benutzen, erstellen wir also ohne notwendige Intention große Mengen sauber strukturierter Daten in Form eines Netzwerks komplexer Netzwerke.

Ein gutes Beispiel für die zunehmende Veröffentlichung strukturierter Datenmengen, und vermutlich gegenwärtig auch das herausragendste, ist die so genannte Linked Open Data Cloud (vgl. Heath 2010), in der Datenmengen nach den Richtlinien des World Wide Web Consortium (W3C) als Graphen, d.h. Netzwerke hinterlegt werden. Figur 9 zeigt die gegenwärtig darin enthaltenen Datenmengen inklusive ihrer Verlinkung untereinander. Neben DBpedia, einem strukturierten Auszug aus Wikipedia, finden sich darin immer mehr bekannte Standarddatenmengen, aus der Biologie, dem Bibliothekswesen, der Medienindustrie, sozialer Netzwerke, zur Geographie, etc.

Die Links in Figur 9 zeigen Querverbindungen zwischen den enthaltenen Datenmengen, die sich aus dem gleichzeitigen Vorhandensein der selben Knoten ergeben – wir sehen also beispielsweise einen Link wenn eine Reihe von Örtlichkeiten sowohl in Geonames als auch in DBpedia vorkommt. Was auf den ersten Blick auffällt, ist dass wir auch hier wieder ein komplexes Netzwerk vor Augen haben. Die Linked Data Cloud ist also ein Netzwerk von Datenmengen, d.h. von Netzwerken komplexer Netzwerke – in anderen Worten: ein Eldorado für die multidisziplinäre Netzwerkwissenschaft.

Wie komplex die zu erwartende Großstruktur der Linked Data Cloud ist, zeigt ein kurzer Blick auf die in den Datenmengen enthaltenen Arten von Knoten und Links. So gut wie jede Diskussion über Datenbanken in den Geisteswissenschaften oder Daten für das Semantic Web spricht auch über Standards, die eingehalten werden müssen, sollen, oder werden. Figur 10 stellt solchen zumeist a priori Diskussionen eine komplexe Realität entgegen – die so genannte UMBEL-Ontologie (UMBEL.org), die, hier in der Version 0.80, als übergeordnete Verbindungsschicht anhand knapp 22.000 Referenzkonzepten und über 63.000 Links die einzelnen Ontologien oder Modelle der Datenmengen in der Linked Data Cloud und anderswo zu verbinden sucht. Naiv gesehen ist die UMBEL-Ontologie

damit vergleichbar mit dem Datenmodell in Figur 7b: jede Knotenart darin repräsentiert Tausende oder gar Millionen von Instanzen, jede Linkart entspricht einem komplexen Netzwerk wie in den Zellen von Figur 8. Angesichts der Tatsache dass sich die UMBEL-Ontologie allein – d.h. also nur das idealisierte Referenzdatenmodell – in der selben Größenordnung bewegt wie der gesamte Census, den man noch vor Kurzem als riesig empfand, könnte einem der Atem stocken. Fakt ist jedoch, dass die Analyse von Netzwerken mit hunderten von Millionen von Knoten und Links bereits heute technisch möglich ist und praktiziert wird (vgl. SNAP). Statt in diskreten Fachbereichen, bewegen wir uns in einem Kosmos von sich verschneidenden Datenmengen, Modellen und Spezialfragen, der die definierten Fachgrenzen selbst zwischen Natur- und Geisteswissenschaft nicht nur in Frage stellt sondern praktisch auflöst.

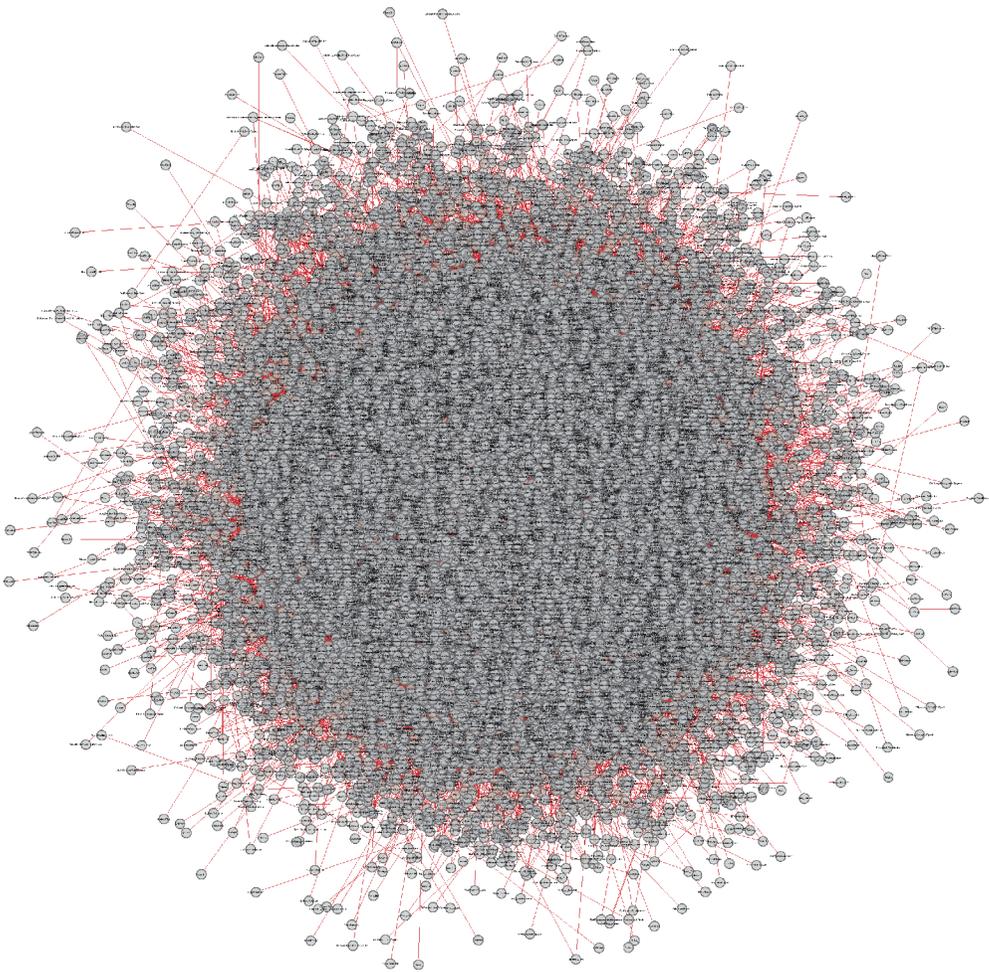


Fig. 10. Ein Ridiculogram der UMBEL-Ontologie nach Structured Dynamics LLC/Ontotext AD. <http://www.umbel.org/resources/graphs>

Vom Text zum Netzwerk

Skeptiker werden an dieser Stelle anmerken, dass die meisten kunstwissenschaftlichen Einordnungen nach wie vor in Büchern und Aufsätzen publiziert werden, das heißt aus der Sicht der Datenanalyse in Form von halb- oder gänzlich unstrukturiertem Text. Doch auch hierzu gibt es beispielsweise mit dem Google-Books-Projekt Entwicklungen, deren Tragweite nicht zu unterschätzen ist. Auch hier wird das kognitive Limit eines einzelnen oder gar einer Gruppe von Forschern bei weitem gesprengt. Im Literaturverzeichnis einer umfangreichen Dissertation in den Kunstwissenschaften finden sich zum Beispiel circa 500 bis 1000 Publikationen; In einem großen Forschungsprojekt wie dem Census sammeln 35 Mitarbeiter über 60 Jahre um die 35.000 Rezeptionslinks; Google Books hingegen hat seit 2004 nach internen Angaben um die 20.000.000 Bücher gescannt – und dabei geht es nicht nur um die Digitalisierung, Auffindung, und zur Verfügungstellung von Büchern. Ein mindestens genauso wertvolles Ziel für Google ist es die in den Büchern implizite strukturierte Information zu extrahieren, um diese zum Beispiel zu nutzen, die allgemeine Suche im Web zu verbessern. Auch hier jagt man nach einem Netzwerk komplexer Netzwerke.

Figur 11 zeigt eine Annäherung an die Struktur von Google Books als Diagramm. Im Zentrum stehen dabei die gescannten Bücher. Links davon erkennt man die Benutzer, die mit den Digitalisaten interagieren. Dabei ist die Beziehung zwischen Benutzern und Büchern asymmetrisch. Benutzer erzeugen durch ihre Aktivität mit der Maus ein Profil ihrer

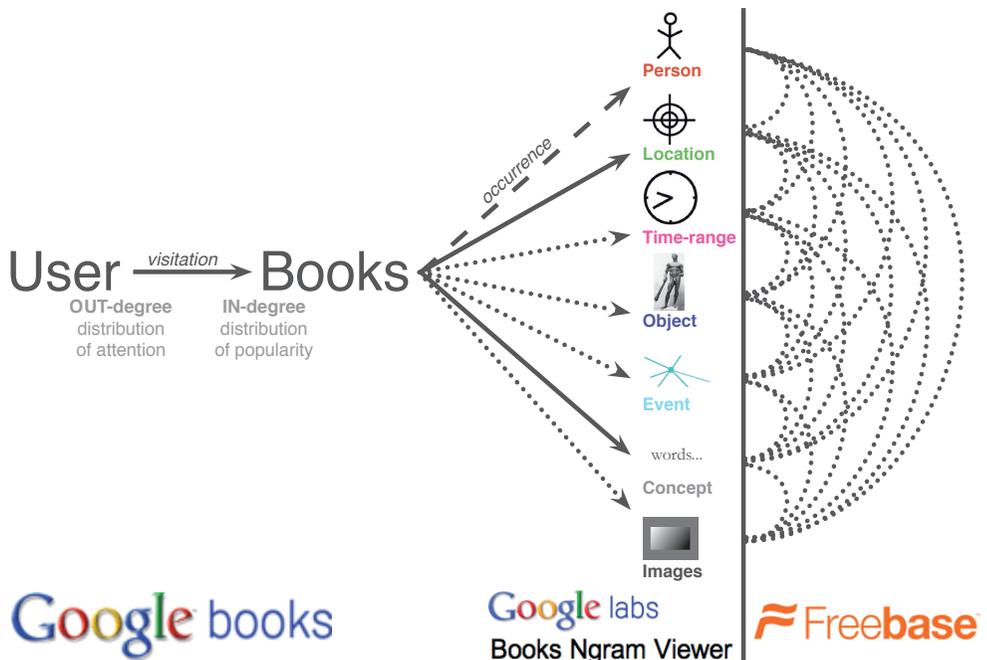


Fig. 11. Annäherung an die Struktur von Google Books. Fig. 1. Beispiel der Rezeption antiker Monumente in visuellen Dokumenten.

selbst, gesteuert durch ihre jeweilige Aufmerksamkeit und begrenzt durch ihr kognitives Limit – jeder Mensch kann nur eine gewisse Anzahl von Büchern lesen, was sich in der Häufigkeitsverteilung des Besucherlinks niederschlägt. Der begrenzten Aufmerksamkeitsverteilung auf der Seite der Benutzer steht auf der Seite der Bücher eine nahezu unbegrenzte Popularitätsverteilung gegenüber – prinzipiell gibt es keine Begrenzung wie viele Benutzer wie oft ein Buch aufrufen. Bereits die Benutzung der digitalen Bücher durch die Benutzer erzeugt somit Daten in Form eines komplexen Netzwerks, mit Häufigkeitsverteilungen wie man sie in vielen anderen realen Beispielen findet. Die Struktur, Dynamik und Evolution dieses Netzwerks zu verstehen bildet eine der wichtigsten Geschäftsgrundlagen des Suchmaschinenanbieters.

Rechts der Bücher erkennt man in Figur 11 einen Fächer von Links zu Gegenständen, die es aus den Büchern zu extrahieren gilt. Besucht man die Webseite von Google Books (<http://books.google.com/>), so findet man derzeit bei vielen Büchern bereits zwei Produkte einer entsprechenden Explikation: Eine so genannte Word Cloud mit Worten, die für das Buch typisch sind, sowie eine Landkarte mit im jeweiligen Buch zu findenden Örtlichkeiten. Google ist also drauf und dran nicht nur Worte aus Büchern zu extrahieren, sondern auch deren Bedeutung im Zusammenhang zu erschließen.

Prinzipiell entsteht dabei zunächst ein Netzwerk zwischen Büchern und Worten bzw. Wortgruppen, so genannten n-Grams (im Sinne von 1-Wort-, 2-Wort-, bis zu n-Wort-Gruppen). Knotenarten des Netzwerks sind dabei Bücher auf der einen Seite und Worte bzw. Wortgruppen auf der anderen. Der Link bezeichnet das Auftreten (occurrence) eines Wortes bzw. der Wortgruppe im Buch. Interessanterweise können wir auch hier ein asymmetrisches Verhältnis erwarten. Bücher werden gekennzeichnet sein durch eine begrenzte Aufmerksamkeitsverteilung, während die Popularität von Worten oder Konzepten prinzipiell unbegrenzt, d.h. skalenfrei ist – einfacher gesagt kein Buch kann alle Kunstwerke der Welt aufzählen, ein Kunstwerk kann jedoch von beliebig vielen Büchern genannt werden.

Bis hierher sprechen wir nur über Worte oder Zeichenketten in Büchern. Eigentliches Ziel ist es jedoch sicherlich die darin enthaltenen Konzepte, Objekte, Personen, Örtlichkeiten, Perioden, Ereignisse, etc. zu identifizieren. Es ist schließlich von Interesse ob die Zeichenkette „Faust“ im jeweiligen Buch eine geballte Hand oder einen Verweis auf Goethe bezeichnet. Jenseits des bloßen Auftretens liegt des Pudels Kern außerdem nicht in der Beziehung zwischen Büchern und Konzepten, sondern in den mannigfaltigen Beziehungen zwischen den Konzepten untereinander. Wir wollen wissen was Faust und der Teufel mit einander zu tun haben; ob beide auf der selben Buchseite erscheinen ist sekundär. Ziel ist die Explikation des Netzwerks komplexer Netzwerke aller in Büchern getroffenen Aussagen, von denen alle kunsthistorischen Einordnungen im Übrigen nur ein kleiner Teil sind.

Es steht außer Frage, dass Google von diesem Ziel noch ein gutes Stück entfernt ist. Nichts desto trotz befinden wir uns auf einem Weg der in diesem Jahrhundert, angesichts der endlichen Zahl aller bisher publizierten Bücher, zu einem guten Teil beschreibbar erscheint. Ohne Zweifel die Aufgabe ist groß, doch sie ist bereits in Angriff genommen. Es ist wohl kein Zufall dass Google gleichzeitig mit dem Wiedereinstieg der Zeitschrift

Science in die Kulturwissenschaften (Michel et al. 2011) die Häufigkeit aller 1- bis 5-Wort-Gruppen in 5% aller bisher publizierten Bücher der Wissenschaft zur Verfügung gestellt hat. Ebenso wenig zufällig erscheint Googles Akquisition der Netzwerkdatenbank Freebase (Freebase.com), deren Aufgabe es ist das Netzwerk komplexer Netzwerke explizierter strukturierter Daten sinnvoll zu kuratieren sowie zugänglich und analysierbar zu machen.

Wie auch immer diese Entwicklung weitergeht, die Wissenschaft komplexer Netzwerke wird dabei weiterhin eine wichtige Rolle spielen. Die Kunstwissenschaft kann dies ebenso wenig ignorieren wie jede andere Disziplin im Ökosystem der Netzwerke. Wie in der Biologie und anderen Fächern zuvor müssen wir lernen immer größere Datenmengen zu erschließen, die richtigen Fragen zu stellen, die richtigen Werkzeuge und Arbeitsabläufe zu entwickeln, und unsere Fachgrenzen zu hinterfragen bzw. zu vergessen.

MS – Boston, 12. März 2011

Danksagung

Der Autor dankt Yong-Yeol Ahn für die großartige Zusammenarbeit bei der Erstellung von Figur 5. Der vorliegende Beitrag entstand im Rahmen eines Forschungsstipendiums der Deutschen Forschungsgemeinschaft (DFG) am Center for Complex Network Research der Northeastern University in Boston. Besonderer Dank gilt dem Gastgeber des Autors Albert-László Barabási.

Literatur:

Ahn, Yong-Yeol Ahn, James P. Bagrow, und Sune Lehmann. 2010. *Link communities reveal multiscale complexity in networks*. *Nature* 466: 761-764.

Anderson, Philip Warren 1972. *More is different*. *Science* 177, no. 4047: 393–396.

Barabási, Albert-László. 2002. *Linked. The New Science of Networks. How everything is Connected to Everything Else and What it Means for Science, Business and Everyday Life*. Cambridge/MA: Perseus Publishing.

Bertin, Jacques. 2010. *Semiology of Graphics: Diagrams, Networks, Maps*. Redlands/CA: ESRI Press.

Census of Antique Works of Art and Architecture Known in the Renaissance. 1947-2011. Hrsg. Berlin-Brandenburgische Akademie der Wissenschaften and Humboldt-Universität zu Berlin. Online-Datenbank. <http://www.census.de> (Stand: 12. März 2011)

Johnson, Bobby. 2010. *Lessons of Scale at Facebook*. USENIX WebApps'10 Conference Keynote Video. <http://www.usenix.org/events/atc10/stream/johnson/index.html> (Stand: 12. März 2011)

Fortunato, Santo. 2010. *Community detection in graphs*. *Physics Reports* 486, 75-174.

Freebase.com. An entity graph of people, places and things, built by a community that loves open data. 2007-2011. San Francisco/CA: Google/Metaweb. Online-Datenbank. <http://www.freebase.com> (Stand: 12. März 2011)

Frutaz, Amadeo P. 1962. *Le piante di Roma*. Rom: Stabilimento Arti Grafiche Luigi Salomone.

Latour, Bruno. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford: Oxford University Press.

Heath, Tom (on behalf of the Linked Data community). 2010. *Linked Data – Connect Distributed Data across the Web*. Informationswebseite. <http://linkeddata.org/> (Stand: 12. März 2011)

Michel, Jean-Baptiste*, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, und Erez Lieberman Aiden*. 2011. *Quantitative Analysis of Culture Using Millions of Digitized Books*. *Science* 331, no. 6014: 176-182.

Newman, Mark E. J. 2010. *Networks. An Introduction*. Oxford: Oxford University Press.

Porter, Mason A., Jukka-Pekka Onnela, und Peter J. Mucha. 2009. *Communities in networks*. *Notices of the American Mathematical Society* 56, 1082.

Schich, Maximilian. 2009. *Rezeption und Tradierung als komplexes Netzwerk. Der CENSUS und visuelle Dokumente zu den Thermen in Rom*. München: Verlag Biering & Brinkmann.

Schich, Maximilian, und Albert-László Barabási. 2009. *Human activity – from the Renaissance to the 21st century*. In: *Cultures of Change. Social Atoms and Electronic Lives*, Ausstellungskatalog: Arts Santa Mònica, Barcelona, 11. Dezember 2009 – 28. Februar 2010, Hrsg. Gennaro Ascione, Cinta Massip, und Josep Perelló, 86-89. Barcelona: Arts Santa Monica.

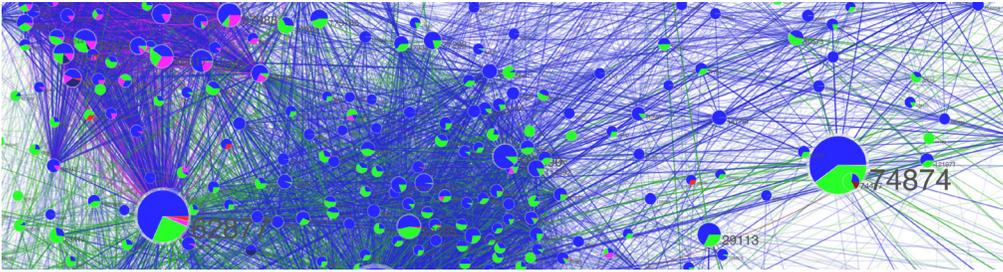
Schich, Maximilian Schich. 2010. *Revealing Matrices*. In: *Beautiful Visualization. Looking at Data through the Eyes of Experts*, Hrsg. Julie Steele, und Noah Iliinsky, 227-254. Sebastopol/CA: O'Reilly 2010.

SNAP. Stanford Large Network Dataset Collection. 2009-2011. Hrsg. Jure Leskovec. Palo Alto: Stanford University. Datenrepositorium. <http://snap.stanford.edu/data/> (Stand: 12. März 2011)

Umbel. Reference Concept Ontology and Vocabulary. 2008-2011. Hrsg. Michael Bergman and Frédéric Giasson. Coralville/IA: Structured Dynamics LLC und Fairfield/CT: Ontotext AD. Projektwebseite. <http://www.umbel.org/> (Stand: 12. März 2011)

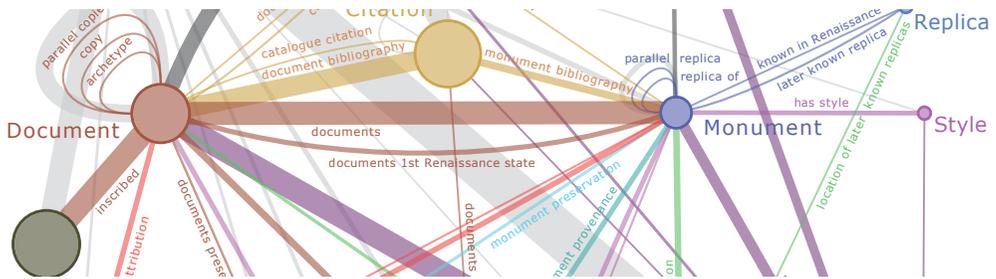
Corpus der antiken Denkmäler, die J.J. Winckelmann und seine Zeit kannten. 2000. Hrsg. Winckelmann-Gesellschaft Stendal. DVD und Online-Datenbank. München: Verlag Biering & Brinkmann / Stiftung Archäologie. <http://www.dyabola.de> (Stand: 12. März 2011) [seit 2009 wird der Corpus fortgeführt unter <http://www.census.de>]

Weitere Beiträge:



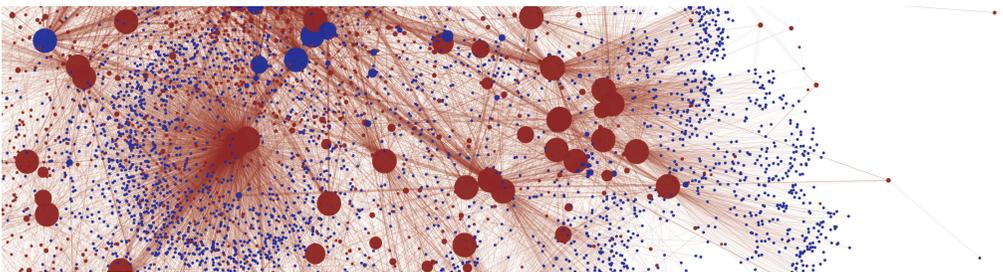
Schich Coscia 2011 (Proceedings)

Zur Entwirrung von Themen in der Klassischen Archäologie!



Schich 2010 (Revealing Matrices)

Inhalt jenseits der Benutzeroberflächen!



Schich 2009 (Rezeption und Tradierung)

Antikenrezeption als komplexes Netzwerk!



Schich BCom 2010 (L'immagine delle Terme)

Die Realität spricht gegen Palladio!