

»»Datenvisualisierung in den Geisteswissenschaften««

Ein Laborbericht als
Tagungsprolog

Sven Peter¹, Thorsten Wübbena^{1,2}

2016

URN: urn:nbn:de:bsz:16-artdok-43488

URL: <http://archiv.ub.uni-heidelberg.de/artdok/volltexte/2016/4348>

¹ Deutsches Forum für Kunstgeschichte Paris / Centre allemand d'histoire de l'art Paris
² Kunstgeschichtliches Institut, Goethe-Universität (Frankfurt am Main)

Inhaltsverzeichnis

Einleitung	3
Die Datenbank des Kunstgeschichtlichen Instituts der Goethe-Universität Frankfurt	4
Visualisierungsarten und Datenvisualisierungstools	5
Klassische Diagramme: D3.js, vis.js	5
Geographische Visualisierung: Palladio	6
Netzwerkvisualisierung: Gephi	7
Gephi-Szenario: Geburts- und Sterbeorte von Künstlern	8
»Cultural Analytics«: ImagePlot (Lev Manovich)	12
Fazit	16
Literatur	18

Einleitung

Am 7. und 8. Dezember 2015 wurde vom Deutschen Forum für Kunstgeschichte Paris (DFK Paris) in Kooperation mit dem Kunstgeschichtlichen Institut der Goethe-Universität Frankfurt/Main (KGI Frankfurt) ein internationales Kolloquium mit dem Thema »Datenvisualisierung in den Geisteswissenschaften« veranstaltet, für welches Forscher unterschiedlicher Disziplinen (Informatiker, Historiker, Germanisten, Kunsthistoriker) aus den USA, Deutschland und Frankreich gewonnen werden konnten. In den dort vorgestellten Beiträgen und Diskussionen wurde intensiv der Einsatz der Datenvisualisierung als Forschungsmethode untersucht. Neben den Möglichkeiten jenseits reiner Illustration wurde auch den entstehenden Fragen z. B. hinsichtlich der Reflektion einer Nutzung »fachfremder« Werkzeuge und der zugrunde gelegten Datenbasis nachgegangen (siehe hierzu auch die Tagungswebsite unter <https://vis.dfgk.org>).

Im Vorfeld dazu und um sich dem Themenbereich »Visualisierung« im Zusammenhang mit spezifischen Anforderungen anstehender Projekte profunder stellen zu können, wurde im vergangenen Jahr von der Abteilung »Digital Humanities« am DFK Paris ein Kurzprojekt durchgeführt, in dessen Verlauf diverse frei verfügbare Visualisierungstools auf ihre Anwendbarkeit hin getestet wurden. Auf Grundlage von ca. 250.000 Entitäten, die am KGI Frankfurt im dortigen Bilddatenbanksystem (ConedaKOR) verwaltet werden, wurden dabei verschiedene Visualisierungsszenarien entwickelt. Unterschiedliche Visualisierungsausformungen wurden auf kunstgeschichtliche Daten angewendet und hierbei die fachspezifischen Fragestellungen sowie die gegenseitige Beeinflussung von Software und Gegenstand untersucht.

Das hier vorliegende Papier ist als »Laborbericht« zu diesem Kurzprojekt zu verstehen. Im Folgenden wird zunächst einmal das Datenbanksystem des KGI Frankfurt und sein zugrunde liegendes Datenmodell erläutert. Im Anschluss daran werden die zum Einsatz gekommenen Visualisierungsarten und Software-Tools ebenfalls kurz vorgestellt, bevor dann auf eine Auswahl der im Rahmen des Projektes entwickelte Szenarien näher eingegangen wird und die mittels der ausgewählten Werkzeuge erstellten Visualisierungen präsentiert werden. Die aus dem Einsatz und der Analyse der Visualisierungstools sowie der Daten aus der Bilddatenbank des KGI Frankfurt erworbenen Erkenntnisse werden in einem letzten Abschnitt zusammengefasst.¹

1 Ein besonderer Dank geht an Fabian Cremer und Henry Keazor für die Inspiration und tatkräftige Unterstützung.

Die Datenbank des Kunstgeschichtlichen Instituts der Goethe-Universität Frankfurt

Das KGI Frankfurt betreibt für Lehre und Forschung eine digitale Bilddatenbank, die aus der früheren, analogen Diathek hervorgegangen ist. Für diese Bilddatenbank² wird das quelloffene Datenbanksystem ConedaKOR³ verwendet, welches für die Archivierung, Verwaltung und Recherche von Bild- und Metadaten auf einer gemeinsamen webbasierten graphischen Oberfläche entwickelt wurde und konzeptuell als Graph-Datenbanksystem realisiert worden ist.

Der Datengraph stellt sich dabei als ein Netzwerk aus virtuellen Entitäten und deren Verknüpfungen zueinander dar und ein Graph besteht somit im Wesentlichen aus Knoten und Kanten, die jeweils zwei der Knoten miteinander verbinden. Im von ConedaKOR realisierten Graphmodell sind diese Knoten und Kanten typisiert, d. h. übergeordneten Typen zugeordnet, über die die möglichen Verknüpfungen im Graphen eingeschränkt werden können. Durch weitere freie Attribute können die Knoten zusätzlich näher bestimmt werden.

Exemplarisch ist ein winziger Teilgraph aus der Datenbank des KGI Frankfurt in Abbildung 1 dargestellt. Kanten zwischen Knoten sind immer beidseitig navigierbar. Für den hier zu sehenden Graphen bedeutet dies, dass der Bamberger Altar von Veit Stoß ein Teil des südlichen Querhauses des Bamberger Doms ist und entsprechend anders herum gelesen, das Querhaus des Doms als übergeordnetes Werk den Bamberger Altar beinhaltet.

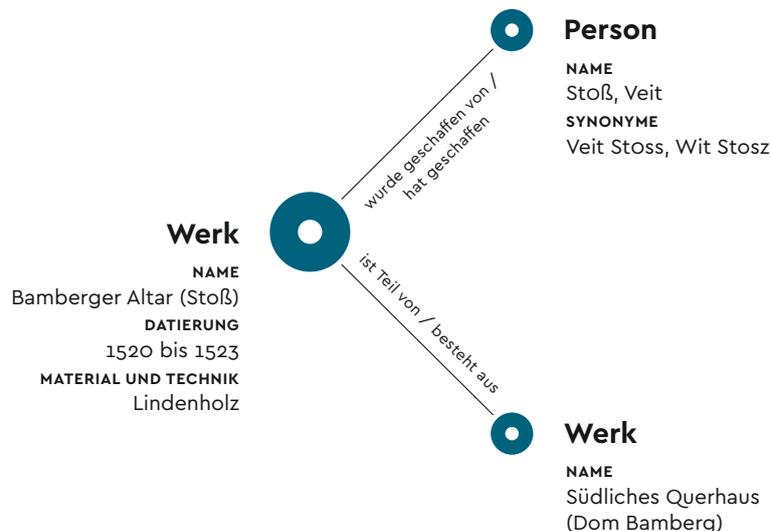


Abbildung 1

Modell eines Ausschnitts aus dem Datengraph der Datenbank des KGI Frankfurt

² Bilddatenbank des KGI Frankfurt: <https://kor.uni-frankfurt.de/> [zuletzt aufgerufen am 21.07.2016].

³ Coneda UG, mit Angaben zu ConedaKOR: <https://coneda.net/> [zuletzt aufgerufen am 21.07.2016].

Aktuell⁴ enthält die Datenbank des KGI Frankfurt 268.882 Entitäten (Knoten) und 551.750 Verknüpfungen (Kanten) im Datagraph. Die wichtigsten Entitätstypen repräsentieren Medien, Werke, Personen, Literatur, Institutionen und Orte, die mit einer Vielzahl von Relationen miteinander verbunden werden können.

Visualisierungsarten und Datenvisualisierungstools

Klassische Diagramme: D3.js, vis.js

Mit klassischen Diagrammen sind hier Darstellungen gemeint, die als Mittel zur graphischen Informationsvisualisierung bereits seit langem eine weite Verbreitung gefunden haben. Hierzu zählen beispielsweise Balken- oder Tortendiagramme, Scatter Plots, Zeitleisten usw.⁵

Zwei auf Javascript fußende, browserbasierte Visualisierungsprogrammbibliotheken, mit denen sich diese Art Diagramme realisieren lassen, sind beispielsweise vis.js und D3.js.

Vis.js⁶ stellt mehrere, leicht zu realisierende Module bereit, die es erlauben, Daten zu visualisieren, mit den Visualisierungen zu interagieren und diese zu manipulieren. Mit den zur Verfügung gestellten Visualisierungskomponenten lassen sich beispielsweise Netzwerke, Zeitleisten und Graphen erzeugen.

Die verbreitete Visualisierungsbibliothek D3.js⁷ stellt nicht wie vis.js fertige Visualisierungsmodule bereit, sondern enthält elementare Bausteine, mit denen sich eine praktisch unbegrenzte Menge an Visualisierungen realisieren lassen. Bei den mit 3D.js im Rahmen dieses Projektes realisierten Visualisierungen handelt es sich ausschließlich um klassische Diagramme im obigen Sinne, weswegen 3D.js seinen Platz in diesem Abschnitt erhalten hat.

In Abbildung 2 ist nun ein mittels 3D.js erstelltes Tortendiagramm zu sehen, welches die Verhältnisse von Material und Technik der repräsentierten Werke in der Datenbank des KGI Frankfurt widerspiegelt. Abbildung 3 stellt in einem Balkendiagramm die prozentuale Verteilung von Ölgemälden – in der Datenbank des KGI Frankfurt – auf zeitliche Abschnitte von je 50 Jahren zwischen 1300 und 2000 dar.

4 Mai 2016

5 Die Website <http://www.datavizcatalogue.com/> gibt einen Überblick über eine Vielzahl verschiedener Diagramme, die bei der Datenvisualisierung zum Einsatz kommen können [zuletzt aufgerufen am 21.07.2016].

6 <http://visjs.org/> [zuletzt aufgerufen am 21.07.2016].

7 Data-Driven Documents (3D.js): <https://d3js.org/> [zuletzt aufgerufen am 21.07.2016].

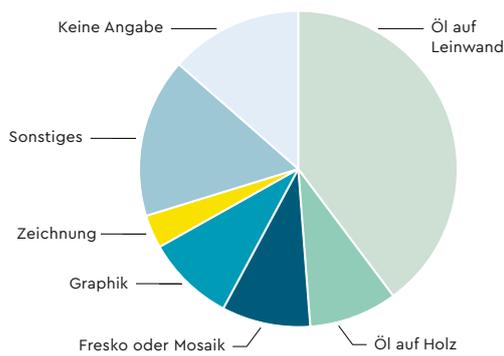


Abbildung 2

Tortendiagramm zur Visualisierung der Verhältnisse von Material und Technik der Werke in der Datenbank des KGI Frankfurt (realisiert mittels 3D.js)

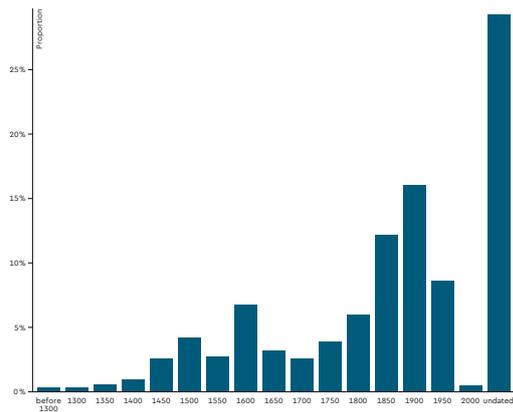


Abbildung 3

Balkendiagramm zur Visualisierung der Produktion von Ölgemälden in der Datenbank des KGI Frankfurt in Perioden von 50 Jahren (realisiert mittels 3D.js)

Geographische Visualisierung: Palladio

Bei der geographischen Visualisierung werden georeferenzierte Daten in einem geographischen Bezugssystem abgebildet.

Hierfür wurde im Projekt die bekannte Webanwendung Palladio⁸ genutzt, die vier verschiedene Visualisierungsarten von Daten erlaubt, so z. B. in einem Graphen, einer Tabelle oder einer Gallery mit Gitter-Layout. Als vierte Möglichkeit lassen sich georeferenzierte Daten in einer Weltkarte darstellen, ein Vorgehen, welches auf die hier vorliegenden Daten angewandt wurde. Palladio ist recht einfach in der Handhabung und so genügt es, die betreffenden Daten in Tabellenform, z. B. als Spreadsheet oder CSV-Dateien über die Website hochzuladen. Danach können die Daten in einer der oben genannten Visualisierungsarten dargestellt werden.

In Abbildung 4 sind die Orte aus der Datenbank des KGI Frankfurt dargestellt, in denen sich Institutionen befinden, die Kunstwerke des 18. Jahrhunderts in ihrer Sammlung aufweisen. Die Größe der dargestellten Kreise ist dabei proportional zur Anzahl der entsprechenden Kunstwerke.

⁸ <http://palladio.designhumanities.org/> [zuletzt aufgerufen am 21.07.2016].



Abbildung 4

Visualisierung von Sammlungsorten zur Kunst des 18. Jahrhunderts in der Datenbank des KGI Frankfurt (realisiert mittels Palladio)

Netzwerkvisualisierung: Gephi

Gephi⁹ ist ein quelloffenes Software-Tool zur Analyse und Visualisierung von Netzwerken bzw. Graphen. Die direkte Manipulation des Graphen über die graphische Benutzerschnittstelle und die Darstellung des Graphen in Echtzeit erlauben die interaktive Exploration und Interpretation der visualisierten Daten. Das Tool erlaubt dabei unter anderem, die Daten zu filtern, den Graphen statistisch zu analysieren, Knoten und Kanten des Graphen zu ordnen, zu gruppieren und das räumliche Layout des Graphen anzupassen.¹⁰

Leider war es mit Gephi im Rahmen dieses Projektes nicht möglich, den gesamten Datengraph der Datenbank des KGI Frankfurt (d. h. ca. 250.000 Knoten) zu verarbeiten, so dass nur kleinere Teilgraphen analysiert und visualisiert werden konnten. Als ein weiteres Problem hat sich herausgestellt, dass Gephi in der verwendeten Version 0.8.2 beta auch keine parallelen Kanten verarbeiten und darstellen konnte¹¹. Nichtsdestotrotz ist Gephi ein sehr ausgereiftes, mächtiges Werkzeug, das viele Möglichkeiten zur Analyse und Visualisierung von Datengraphen bietet, und darüber hinaus sehr einfach und intuitiv zu bedienen ist.

⁹ <https://gephi.org/> [zuletzt aufgerufen am 21.07.2016].

¹⁰ Bastian, M.; Heymann, S.; Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. ICWSM, 8, 361–362.

¹¹ Parallele Kanten werden in eine einzige Kante mit einem Kantengewicht, das der Summe der Kantengewichte der parallelen Kanten entspricht, überführt. Die neue Version 0.9 soll aber in Zukunft die Verarbeitung paralleler Kanten zulassen: <https://gephi.wordpress.com/2015/11/02/announcing-gephi-0-9-release-date/> [zuletzt aufgerufen am 12.02.2016].

Gephi-Szenario: Geburts- und Sterbeorte von Künstlern

Mittels Gephi wurde das hier zu sehende Netzwerk zwischen Orts- und Personenentitäten in der Graphdatenbank des KGI Frankfurt visualisiert (Abbildung 5).

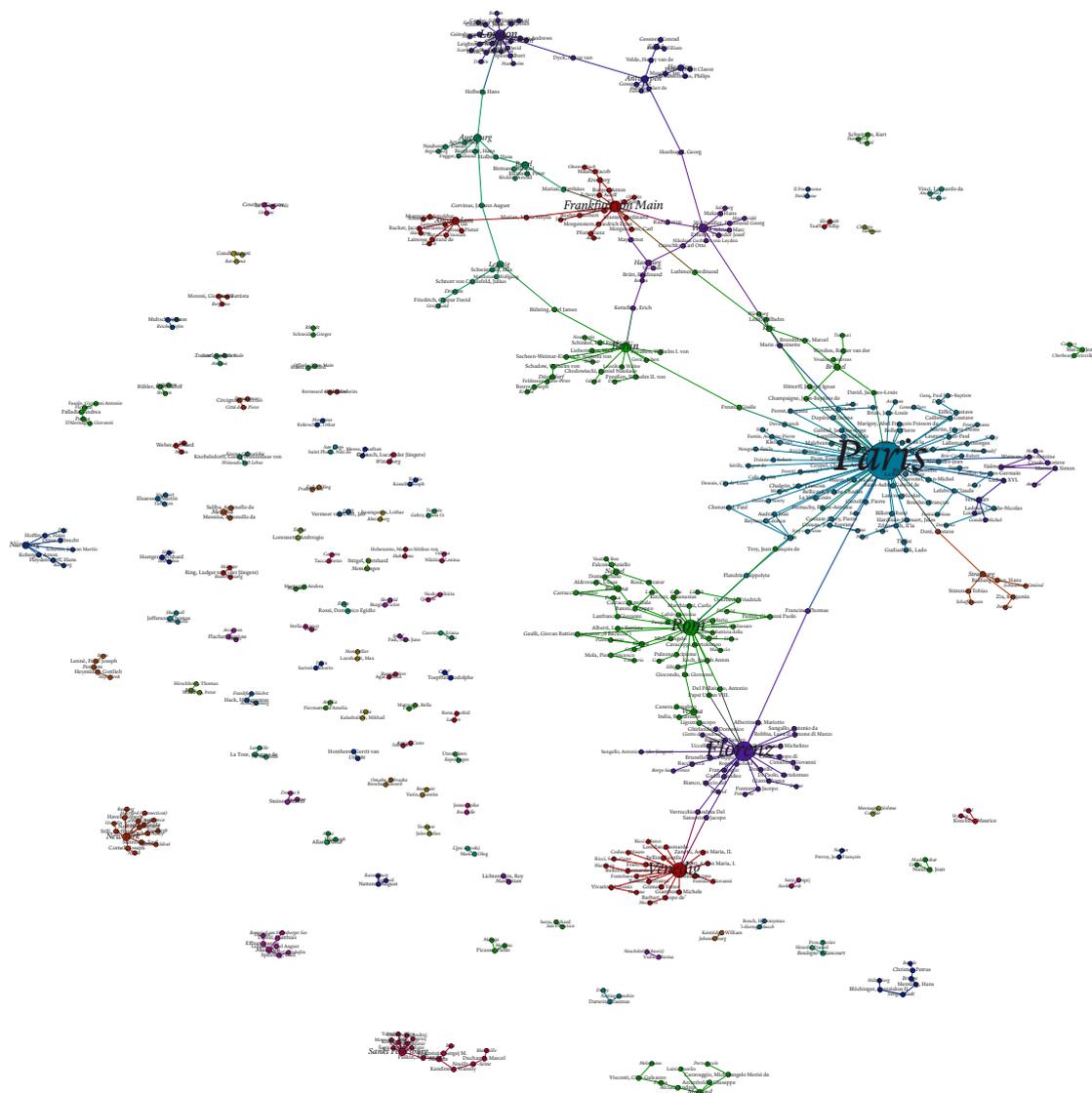


Abbildung 5

Netzwerkvisualisierung von Personen und deren Geburts- und Sterbeorten (realisiert mittels Gephi)

Zunächst wurden dazu alle relevanten Daten der Personen- und Ortsentitäten, sowie die zwischen den Entitäten dieser beiden Typen existierenden Geburts- und Sterbeort-Relationen in ein von Gephi verarbeitbares Format¹² exportiert.

Dieser Teilgraph wurde anschließend in Gephi importiert und die nicht verbundenen Entitäten herausgefiltert. Der resultierende Graph enthielt danach noch 2,91% der ursprünglichen Knoten. Parallele Kanten, die beispielsweise entstehen, wenn eine Person am selben Ort geboren und gestorben ist, konnten leider nicht in dieser Form von Gephi visualisiert werden (siehe oben). In diesen Fällen überführt Gephi die parallelen Kanten in eine Kante, deren Kantengewicht der Anzahl der ursprünglich parallelen Kanten entspricht.

Die importierten Graphdaten wurden anschließend analysiert. Dabei wurden die gewichteten Grade der Knoten berechnet, d.h. für jeden Knoten wurde ein Wert berechnet, der der Summe der Kantengewichte der am Knoten ein- und ausgehenden Kanten entspricht. Anschließend wurde darüber hinaus die Modularität des Graphen berechnet, wobei jeder Knoten in eine Modularitätsklasse eingeordnet wird, innerhalb derer die Knoten stärker untereinander verbunden sind als mit Knoten anderer Modularitätsklassen.

Auf Grundlage der berechneten Kennwerte wurde der Graph anschließend visualisiert: dabei wurden die Knoten des Graphen farblich nach Modularitätsklassen gruppiert und die Ordnung der gewichteten Knotengrade über die Knotengröße ausgedrückt. Anschließend wurde der Graph über einen sogenannten *force-based* Algorithmus in der zweidimensionalen Fläche angeordnet. Der *force-based* Algorithmus berechnet die Position jeden Knotens in Bezug auf die anderen Knoten des Graphen, indem eine Anziehung zwischen verbundenen Knoten mit einer Abstoßung zwischen unverbundenen Knoten verrechnet wird.

In der so entstandenen Visualisierung (Abbildung 5) kann man erkennen, dass wenige Orte besonders häufig Geburts- und Sterbeorte sind. Allen voran steht Paris, gefolgt von den italienischen Städten Florenz, Rom, Venedig. Aber auch Frankfurt, Berlin und London sind besonders häufig Geburts- und Sterbeorte. Zudem liegen Ortsknoten in einer Modularitätsklasse meist geographisch nahe beieinander und nur relativ wenige Personenknoten verbinden Modularitätsklassen untereinander.

Natürlich können diese Ergebnisse nur als Indizien für eventuelle realweltliche Sachverhalte dienen. Die Datenbank des KGI Frankfurt ist qua ihrer Natur in keinem Fall eine systematisch angelegte Datenbank für Künstlerbiographien, sondern wurde zur Unterstützung der universitären Lehre konzipiert, wobei das Bereitstellen von digitalisiertem Bildmaterial im Vordergrund steht. Für eine systematisch angelegte biographische Datenbank wäre die hier vorgestellte Methode aber sicherlich ein adäquates Mittel, um große biographische Datenmengen nach aussagekräftigen Mustern zu durchforschen. In dem Zusammenhang sei auf ein

¹² Es wurde das Graph Exchange Format (GEXF) verwendet. Homepage: <https://gephi.org/gexf/format/> [zuletzt aufgerufen am 12.02.2016].

Projekt von Maximilian Schich verwiesen, welches er – in der Grundidee dem o.g. ganz ähnlich – 2014 auf Basis der weitaus größeren Freebase-Datenbank von Metaweb bzw. Google durchgeführt hat¹³. Ganz unabhängig von der Qualität der Datenbasis ist natürlich auch die Methode an sich kritisch zu beleuchten. Die Gefahr einer vereinfachten Ableitung bei der Reduktion eines Menschenlebens auf Geburt und Tod berücksichtigt – auch wenn es nur um den migratorischen Aspekt geht – nicht einmal in Ansätzen die Komplexität realweltlicher Phänomene. So ist die menschliche Migration natürlich nicht wie die Visualisierung in einem YouTube-Video¹⁴ zum Charting Culture-Projekt nahelegt, ein linearer Migrationsprozess vom Geburts- zum Sterbeort, sondern setzt sich in Wirklichkeit ja aus einer Vielzahl komplexer Wanderbewegungen im Laufe eines Menschenlebens zusammen.

Die in Abbildung 5 dargestellte Visualisierung, bei der sich die Bedeutung weniger Orte als Geburts- und Sterbeorte von Personen der Frankfurter Datenbank erkennen lässt, kann allenfalls als Ausgangspunkt für weitere Fragen dienen. Der große Vorteil eines interaktiven Tools wie Gephi ist in diesem Zusammenhang, dass die Visualisierungen je nach geändertem Interesse und modifizierter Fragestellung direkt manipuliert werden können.

Im Falle der Visualisierung von Sterbe- und Geburtsorten im Datengraph der Datenbank des KGI Frankfurt wurde so, von der bereits erstellten Visualisierung ausgehend, in einem zweiten Schritt getrennt die Bedeutung der Städte als Geburts- und Sterbeorte betrachtet. Hierzu wurden alle Kanten, die Personen mit ihrem Sterbe- bzw. Geburtsort verbinden, herausgefiltert und die Knotengrade im jeweils resultierenden Graphen neu berechnet, d. h. es wurde für jeden Ortsknoten die Anzahl der Personen ermittelt, für die der Knoten Sterbe- bzw. Geburtsort ist. Dann wurden auch die Personenknoten und die jeweils verbleibenden Kanten aus dem Graphen herausgefiltert und die Größe der verbliebenen Ortsknoten dem berechneten Knotengrad entsprechend angepasst. Anschließend wurden die Knoten in einem kreisförmigen Layout der Größe nach angeordnet.

¹³ Schich, M.; Song, C.; Ahn, Y. Y.; Mirsky, A.; Martino, M.; Barabási, A. L.; Helbing, D. (2014). A network framework of cultural history. *Science*, 345(6196), 558–562

¹⁴ <https://www.youtube.com/watch?v=4glhRkCcD4U> [zuletzt aufgerufen am 21.07.2016].

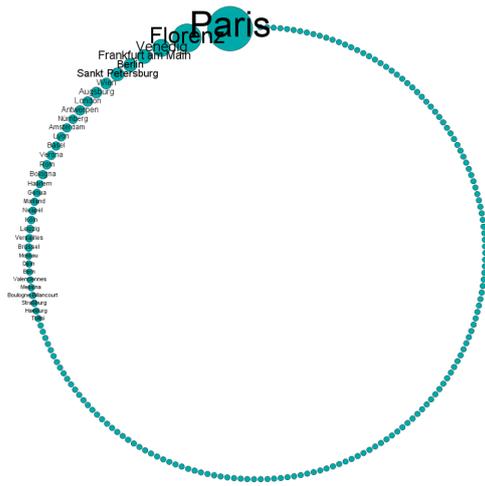


Abbildung 6

Nach ihrer Bedeutung als Geburtsorte der Größe nach angeordnete Ortsknoten (realisiert mittels Gephi)

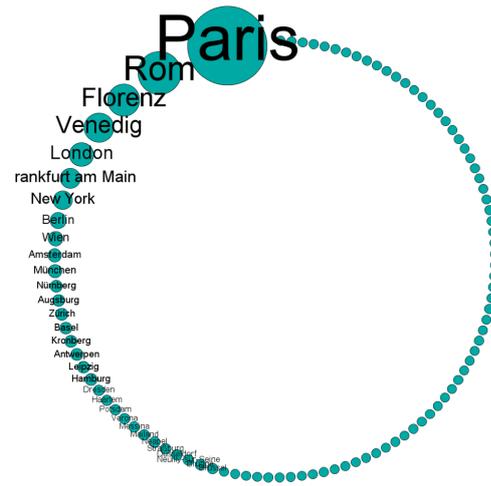


Abbildung 7

Nach ihrer Bedeutung als Sterbeorte der Größe nach angeordnete Ortsknoten (realisiert mittels Gephi)

Die Betrachtung beider Visualisierungen in Abbildung 6 und Abbildung 7 bestätigt noch einmal die bereits bei der ersten Gephi-Visualisierung gemachten Beobachtung, dass nur wenige Orte herausragende Geburts- und Sterbeorte sind, und die Bedeutung – an der Größe der Knoten gemessen – in der Rangfolge schnell abfällt. Paris ist sowohl bedeutendster Geburts- als auch Sterbeort. Unter den nachfolgenden Städten sind vor allem italienische Städte und größere regionale Zentren nördlich der Alpen prominent. Etwas erstaunlich ist die gute Positionierung der Stadt Frankfurt am Main als Geburts- und Sterbeort (4. Geburtsort, 6. Sterbeort), was wohl damit erklärt werden kann, dass in der Lehre am KGI Frankfurt lokale Künstler überdurchschnittlich repräsentiert sind. Insgesamt gibt es weniger Städte, die Sterbeorte, als Städte, die Geburtsorte für Personen in der Datenbank sind. Die Bedeutung der Städte als Sterbeorte ist insgesamt auch höher als ihre Bedeutung als Geburtsorte, dafür spricht der Vergleich der Größen derselben Knoten in beiden Visualisierungen. Recht erstaunlich in diesem Zusammenhang ist auch der große Unterschied der Stadt Rom in ihrer Bedeutung als Geburts- und Sterbeort: während Rom an zweiter Stelle gleich nach Paris als Sterbeort von 26 Personen steht, ist sie als Geburtsort nur für 4 Personen vermerkt¹⁵.

Man stellt also fest – immer auf das vorliegende Datenkonvolut bezogen –, dass es im Laufe des Lebens der Künstler (oder zumindest zu deren Lebensende) zu einer Konzentration in größeren Städten kommt, bei denen sich wenige ganz besonders hervortun. Während beispielsweise in Paris, Florenz und Venedig auch viele Künstler geboren werden, ist die geringe Bedeutung

¹⁵ Auch im Fall von New York überwiegt die Bedeutung als Sterbeort der Bedeutung als Geburtsort ganz erheblich.

Roms als Geburtsort im Vergleich auffällig. Es wäre interessant, hier einmal genauer nach den Gründen dieser Diskrepanz im Datenpool zu suchen¹⁶.

»Cultural Analytics«: ImagePlot (Lev Manovich)

Mit ImagePlot¹⁷, einem freien Visualisierungstool bzw. einem Makro für das auf Java basierende Bildverarbeitungsprogramm ImageJ¹⁸ ist es möglich, Muster in großen Bildsammlungen zu entdecken. ImagePlot wurde von der Software Studies Initiative¹⁹ entwickelt und ist als Werkzeug im Zusammenhang mit Lev Manovichs kulturwissenschaftlichem Paradigma »Cultural Analytics«²⁰ zu sehen. Demnach gilt es, sich der explodierenden Menge kultureller Daten, die im Zeitalter des New Media entstehen, mit neuen Methoden und Werkzeugen zu nähern, indem computerbasiert große kulturelle Datenmengen analysiert und visualisiert werden. Denn den riesigen, allein im World Wide Web von Nutzern generierten Inhalten ist laut Manovich mit den althergebrachten Methoden und Werkzeugen nicht mehr Herr zu werden: So meldete Google bereits 2008, dass ihr Webseitenindex bereits 3 Billionen Webseiten erfasst hätte.²¹ Bei YouTube werden aktuell (2015) 400 Stunden Videomaterial pro Minute hochgeladen²², über 10 Milliarden Bilder sind bei Flickr verfügbar²³.

Durch ImagePlot werden Datenpunkte in einem zweidimensionalen kartesischen Datensystem angeordnet. Das besondere an einem ImagePlot-Diagramm im Vergleich zu einem traditionellen Streudiagramm ist allerdings, dass mittels ImagePlot nicht nur abstrakte Datenpunkte im Diagramm dargestellt werden können, sondern Bilder selbst. ImagePlot bietet sich daher besonders dazu an, digitalen Bildern eingeschriebene Eigenschaften wie z. B. Helligkeit, Farbton und Farbsättigung zu visualisieren. Da die Verortung der Bilder im Diagramm dann unmittelbar mit der menschlichen Wahrnehmung korreliert, kann die Visualisierung so direkt validiert und in Bezug auf die bildimmanenten Informationen interpretiert werden. Im Gegensatz zu anderen Werkzeugen zur Visualisierung, mit denen sich hauptsächlich die Metadaten aus der jeweiligen Bilddatenbank visualisieren lassen, bietet die Kombination ImageJ/ImagePlot also die Besonderheit, in erster Linie für die Visualisierung der Bilddaten »an sich«²⁴ konzipiert zu sein. Darüber hinaus erlaubt es ImagePlot auch eine sehr hohe Anzahl von

16 Eine Hypothese wäre, dass Rom zwar viele bereits ausgebildete Künstler von außerhalb angezogen hat, in Rom selbst aber keine erfolgreiche Künstlerausbildung entwickelt wurde. An dieser Stelle ist aber auch der zeitliche Faktor von Bedeutung, da Rom ab dem 15. bis hin zum 17. Jahrhundert ein künstlerisches Zentrum darstellte, welches internationale Anziehungskraft besaß – im späten 17., frühen 18. Jahrhundert wurde Rom dann aber zunehmend als weniger relevant erachtet, da sich der Fokus nach Paris verlagerte. Wenn Künstler zu der Zeit noch nach Rom kamen, dann eher um die Vergangenheit zu studieren und weniger um sich dort konkret schulen zu lassen. Um diese Beobachtung zu einer untersuchungswürdigen Hypothese auszubauen sind weitere, mitunter auch nicht-empirische Beobachtungen notwendig.

17 ImagePlot: <http://lab.softwarestudies.com/p/imageplot.html> [zuletzt aufgerufen am 21.07.2016].

18 <http://imagej.nih.gov/ij/> [zuletzt aufgerufen am 28.03.2016].

19 Mit je einer Dependence in New York (The Graduate Center, CUNY) und La Jolla, California (UCSD).

20 Ein von Lev Manovich – nach eigener Aussage im Jahr 2007 – eingeführter Begriff, der laut ihm wie folgt zu definieren ist: „Cultural analytics is the use of computational and visualization methods for the analysis of massive cultural data sets and flows.“, <http://lab.softwarestudies.com/p/cultural-analytics.html> [zuletzt aufgerufen am 28.03.2016].

21 Manovich, L. (2009). Cultural analytics: visualizing cultural patterns in the era of "more media", http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf, [zuletzt aufgerufen am 28.03.2016].

22 <http://www.googlewatchblog.de/2015/07/aktuelle-statistiken-youtube-pro/>

23 <http://expandedramblings.com/index.php/flickr-stats/>

24 An dieser Stelle sei selbstverständlich auch auf Pias, C. (2003). Das digitale Bild gibt es nicht – Über das (Nicht-)Wissen der Bilder und die informatische Illusion, in: *zeitenblicke* 2 (2003), Nr. 1 [08.05.2003], <http://www.zeitenblicke.historicum.net/2003/01/pias/index.html> hingewiesen.

Bildern zu verarbeiten²⁵ und so hat Lev Manovich in seiner Studie zur Stilvarianz in Mangas²⁶ in einer einzigen Visualisierung über eine Million Bilder visualisiert.

ImagePlot-Szenario: Historische Helligkeitsentwicklung der Malerei

Da in der Datenbank des KGI Frankfurt nicht explizit zwischen Werkgattungen wie Gemälde, Skulptur oder Architektur unterschieden wird, wurde aus den in der Datenbank existierenden Angaben zu Material und Technik der Kunstwerke eine Liste erarbeitet, welche die am häufigsten vorkommenden Material- und Technikangaben enthält, die darauf schließen lassen, dass der jeweilige Kunstwerkdatensatz ein Gemälde repräsentiert. Auf Grundlage dieser Zusammenstellung wurden dann alle Entitäten vom Typ Werk in der Datenbank, die durch eine der in der zuvor erstellten Liste vorkommenden Angaben näher bestimmt sind, ausgewählt. Daran anschließend wurden die Verknüpfungen zu den diese Gemälde repräsentierenden Entitäten vom Typ Medium extrahiert und damit die digitalen Bilder (jeweils die erste Gesamtdarstellung des Werks).

Die Bilddateien der so ermittelten Gemäldedarstellungen wurden mittels ImageJ u. a. in Bezug auf ihren Helligkeitswertmedian analysiert. Der Helligkeitswertmedian ist der mittlere Wert aller im digitalen Pixelbild für jedes Pixel vorkommenden Helligkeitswerte.

Mit dem berechneten Helligkeitswertmedian für jedes Bild, wurden die einzelnen Bilder mit ImagePlot in einem Streudiagramm dargestellt, wobei für jeden Datenpunkt im zweidimensionalen Koordinatensystem eine miniaturisierte Version des Bildes dargestellt wird. An der x-Achse wurde dabei der Datierungswert des jeweiligen Kunstwerks im Intervall zwischen 1300 und 2050, und an der y-Achse der Helligkeitswertmedian des digitalen das Gemälde repräsentierenden Bildes abgetragen (Abbildung 8).

²⁵ Laut Projektwebsite gibt es „[...] no theoretical limit to the number of images that can be included in a single visualization“, <http://lab.softwarestudies.com/p/imageplot.html> [zuletzt aufgerufen am 28.03.2016].

²⁶ Manovich, L. (2012). How to compare one million images?. In: Understanding Digital Humanities. Palgrave Macmillan UK, 249–278.

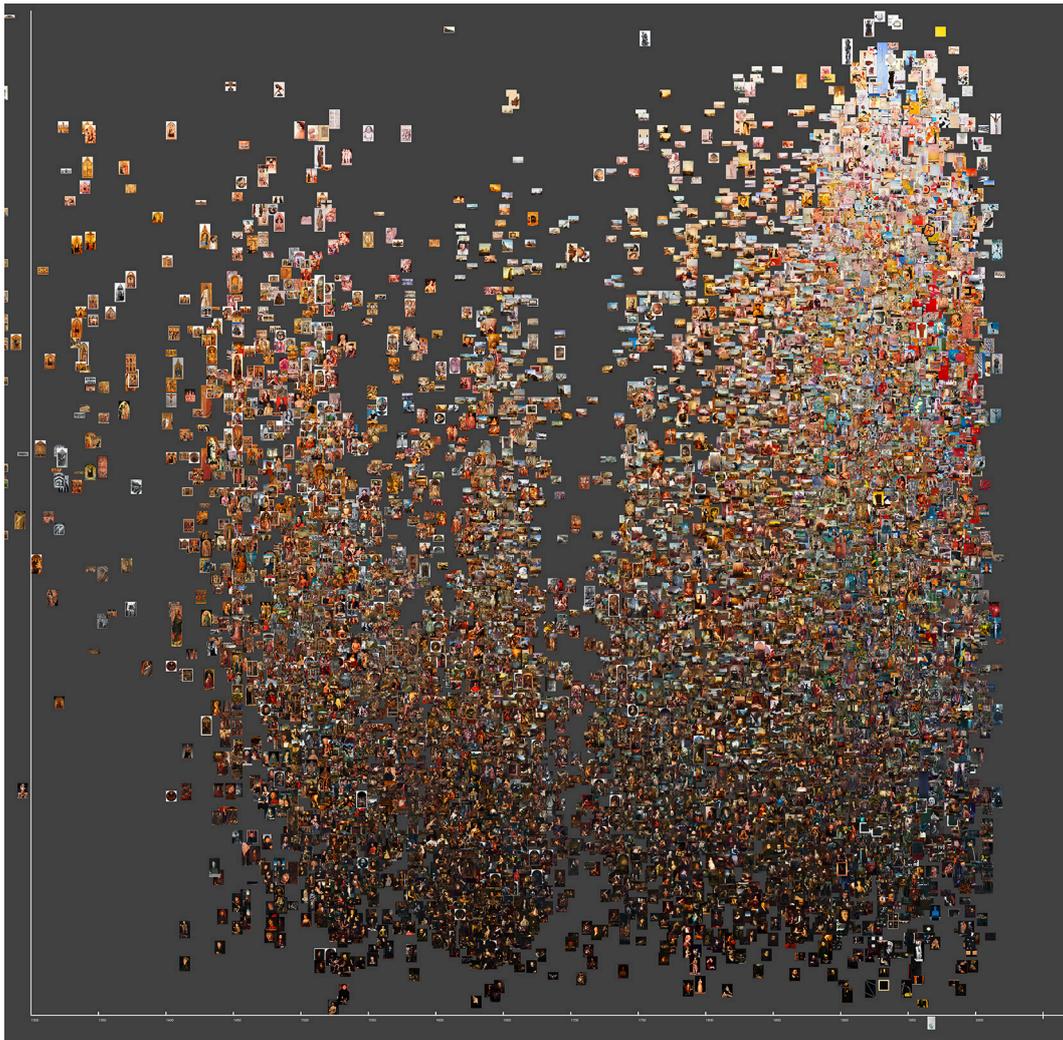


Abbildung 8

Nach Entstehungsdatum und Helligkeit angeordnete Gemälde in der Datenbank des KGI Frankfurt (x-Achse: Entstehungsdatum, y-Achse: Helligkeitswertmedian) (realisiert mittels ImagePlot)

Durch die direkte Korrelation zwischen der Verortung der miniaturisierten Bilder nach Helligkeitswertmedian im Koordinatensystem der Visualisierung und der Wahrnehmung der Helligkeit der miniaturisierten Bilder durch den menschlichen Wahrnehmungsapparat, kann das Ergebnis direkt validiert und interpretiert werden. So erkennt man beispielsweise, dass in der Frankfurter Datenbank sehr helle Gemälde vermehrt ab dem letzten Viertel des 19. Jahrhunderts auftreten, während vor 1700 und ganz besonders um 1600 dunklere Gemälde zu dominieren scheinen. Auffällig ist ebenfalls die sehr geringe Dichte von Bildmaterial aus der Zeit um 1700.

Diese Phänomene könnten nun mit dem aufkommenden Impressionismus und der zunehmenden Autonomie der bildnerischen Mittel ab dem letzten Viertel des 19. Jahrhunderts, und dem Einfluss der Chiaroscuro-Malerei um 1600 in Zusammenhang erklärt werden.

Allerdings muss dem Betrachter des Streudiagramms nach wie vor bewusst sein, dass die im Diagramm dargestellte Situation lediglich die in der Bilddatenbank des KGI Frankfurt vorgehaltene Datenlage widerspiegelt, und selbst diese nicht vollständig, da wie oben beschrieben lediglich eine heuristische Datenauswahl verwendet werden konnte. Die Analyse der Bilddatenbank des KGI Frankfurt liefert deshalb auch hier keine verlässlichen Aussagen über allgemeine realweltliche Sachverhalte kunsthistorischen Interesses.

Wendet man sich allerdings von dem Anspruch ab durch die Interpretation der Visualisierungsergebnisse Aussagen über allgemeine kunsthistorische Sachverhalte treffen zu wollen und begnügt sich stattdessen damit, die Lehrsituation am KGI Frankfurt zu untersuchen, so verfügt man eben doch über eine repräsentative Datenbasis, auf die man seine Interpretationen stützen kann. Unter diesem Aspekt ließe sich annehmen, dass in der Frankfurter Lehre nur wenige Gemälde, die um 1700 entstanden sind, gezeigt werden und diese Epoche daher insgesamt von der Frankfurter Lehre wenig behandelt wird. Natürlich würde auch diese Interpretation der quantitativen Visualisierung einer qualitativen Validierung bedürfen.

Eine vollständige Datenbank aller Gemälde wäre natürlich utopisch. Eine mit Anspruch auf Repräsentativität methodisch angelegte Datenbank – die sich beispielsweise wie die Datenbank *Census of Antique Works of Art and Architecture Known in the Renaissance*²⁷ auf einen gewissen kunstgeschichtlichen Corpus beschränkt – könnte über das hier dargestellte Verfahren aber durchaus auch über die traditionellen Objekte kunstgeschichtlicher Forschung interessante und valide Ergebnisse liefern²⁸.

Ein großes methodisches Problem, das sich bei der Arbeit mit den digitalen Bildern immanenten Daten in ImageJ offenbart hat, ist der Umgang mit der äußerst unterschiedlichen Reproduktionsqualität der Gemälde in der Datenbank. Manche Gemälde sind nur in Grautönen, manche mit, manche ohne Rahmen, oder manche mit breitem, nicht zum eigentlichen Bild gehörenden Rand reproduziert. Schon allein die Existenz von Datensätzen mit mehreren digitalen Bildern, die dasselbe Gemälde reproduzieren, in Bezug auf ihre Bildwerte aber völlig unterschiedlich sind (siehe Abbildung 9 und Abbildung 10), zeigen, dass eigentlich kein absoluter Bezugsrahmen existiert, der es erlauben würde die digitalen Bilder der Datenbank miteinander zu vergleichen und daraus Rückschlüsse auf das Verhältnis der reproduzierten realen Gemälde zueinander zu ziehen. Dazu müssten die digitalen Bilder der zu untersuchenden Gemälde systematisch unter konstanten Laborbedingungen erstellt werden.²⁹

27 <http://www.census.de> [zuletzt aufgerufen am 28.03.2016].

28 Maximilian Schich hat dies beispielsweise in seiner Dissertation zur Antikenrezeption getan: Schich, M. (2009). *Rezeption und Tradierung als komplexes Netzwerk: der CENSUS und visuelle Dokumente zu den Thermen in Rom.*

29 Siehe hierzu auch den Blogbeitrag von Trüper, L. (2015). *Das Bild als Datensatz: Ein Kommentar zu neuen Methoden der quantitativen Bildanalyse*, <http://blog.studiumdigitale.uni-frankfurt.de/reprometh/blog/2015/11/24/das-bild-als-datensatz-ein-kommentar-zu-neuen-methoden-der-quantitativen-bildanalyse/> [zuletzt aufgerufen am 28.03.2016].



Abbildung 9

Digitalisat der Lucca-Madonna („Version 1“)



Abbildung 10

Digitalisat der Lucca-Madonna („Version 2“)

Fazit

Bei den durchgeführten Visualisierungen hat sich durchgehend gezeigt, dass die Fragestellungen, denen mit der Hilfe von Visualisierungstools nachgegangen werden soll, mit der vorhandenen Datenbasis abgeglichen werden müssen. Wollte man sich nicht mit der Untersuchung der Nutzung der Datenbank des kunstgeschichtlichen Instituts der Universität Frankfurt in der Lehre zufrieden geben, so war die Datenlage und -dichte der Datenbank in diesem Fall sicher nicht ausreichend, um im Hinblick auf allgemeine kunsthistorische Fragen valide Visualisierungsergebnisse zu liefern. Der sinnvolle Einsatz von Software-Tools zur Visualisierung quantitativer Daten in den Geisteswissenschaften – das hat die praktische Beschäftigung mit Datenbank und Visualisierungstools gezeigt – muss auf einer systematisch, im Hinblick auf eine gewisse Repräsentativität eines Forschungsgegenstandes angelegten Datenbank, aufbauen. Und bereits bei dieser Anlage von Datensätzen – weit, bevor Visualisierungssoftware überhaupt zum Einsatz kommt – müssen vom Anwender Entscheidungen getroffen werden, die bereits wichtige »Stellschrauben« für die weitere Verarbeitung darstellen: Welchen Umfang soll der Datenbestand haben?, welche »Objekte« sind in die Datenerfassung einzubeziehen?, anhand welcher Eigenschaften sollen diese »Objekte« erfasst werden? Um nur einige in diesem Zusammenhang relevante Fragen zu stellen. Hier gilt es, unbedingt sensibel zu bleiben bzw. als KunsthistorikerIn ein Bewusstsein für die eigenen Forschungsdaten zu gewinnen.

Dennoch hat die Beschäftigung mit der Analyse und Visualisierung geisteswissenschaftlicher Daten gezeigt, dass Visualisierungen durchaus in der Lage sind, auf interessante Muster in großen kulturellen Datenmengen aufmerksam zu machen, und dadurch bereits aufgestellte Hypothesen bestätigen oder widerlegen zu können, oder auch zu weiteren Fragestellungen zu führen. So kann die Datenvisualisierung in den Geisteswissenschaften einerseits der Demonstration quantitativer Argumente dienen, zum anderen dient sie aber auch als Instrument der Forschung, das dem Forschenden überhaupt erst den Zugang zur Interpretation quantitativer Daten erlaubt.

Dieses Argument ließe sich auch noch weiter ausbauen, z. B. die Richtung, dass die Visualisierungswerkzeuge weniger ein »wissenschaftliches« Mittel zur Interpretation sein können als vielmehr ein Instrument im vorhergehenden Schritt der Sichtung, Beobachtung und Strukturierung mit enormen Potential jenseits der individuellen Erfahrung/Leistung. Das würde den Rahmen dieses knappen »Laborberichts« allerdings sprengen und möge an anderer Stelle fortgesetzt werden.

Literatur

BASTIAN, MATHIEU; HEYMANN, SEBASTIAN; JACOMY, MATHIEU (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, 8, 361–362.

MANOVICH, LEV (2009). Cultural analytics: visualising cultural patterns in the era of “more media”, http://manovich.net/content/04-projects/063-cultural-analytics-visualizing-cultural-patterns/60_article_2009.pdf [zuletzt aufgerufen am 28.03.2016].

MANOVICH, LEV (2012). How to compare one million images?. In: *Understanding Digital Humanities*. Palgrave Macmillan UK, 249–278.

PIAS, CLAUS (2003). Das digitale Bild gibt es nicht – Über das (Nicht-)Wissen der Bilder und die informatische Illusion, in: *zeitenblicke* 2, Nr. 1 [08.05.2003], <http://www.zeitenblicke.historicum.net/2003/01/pias/index.html>.

SCHICH, MAXIMILIAN; SONG, CHAOMING; AHN, YONG YEOL; MIRSKY, ALEXANDER; MARTINO, MAURO; BARABÁSI, ALBERT-LÁSZLÓ; HELBING, DIRK (2014). A network framework of cultural history. *Science*, 345(6196), 558–562. Schich, Maximilian (2009). *Rezeption und Tradierung als komplexes Netzwerk: der CENSUS und visuelle Dokumente zu den Thermen in Rom*.

TRÜPER, LENA (2015). Das Bild als Datensatz: Ein Kommentar zu neuen Methoden der quantitativen Bildanalyse, <http://blog.studiumdigitale.uni-frankfurt.de/reprometh/blog/2015/11/24/das-bild-als-datensatz-ein-kommentar-zu-neuen-methoden-der-quantitativen-bildanalyse/> [zuletzt aufgerufen am 28.03.2016].