Final report / Abschlussbericht: DFG Forschungsstipendium

1. General information

DFG sign / Geschäftszeichen: SCHI 1065/2-1

Applicant: Dr. Maximilian Schich

Institute: Center for Complex Network Research, Northeastern University, Boston
Chair: Prof. Albert-László Barabási

Subject:

Complex Networks in Art Research – Exemplary Proofs of Concept

deutsch: Komplexe Netzwerke in den Kunstwissenschaften – Beispielanwendungen zur Etablierung eines Forschungszweiges.

Report and funding time-frame: April 1, 2009 – March 31, 2012

List of most important project publications:

1. *Schich, Maximilian:* ***Revealing Matrices.*** *in: Beautiful Visualization. Looking at Data through the Eyes of Experts. ed. Julie Steele and Noah Iilinsky. Sebastopol, CA: O'Reilly 2010 pp. 227-254.*
   URL: http://revealingmatrices.schich.info

2. *Schich, Maximilian:* ***Netzwerke komplexer Netzwerke in der (Kunst)Wissenschaft.*** *in: Die Dynamik sozialer und sprachlicher Netzwerke. ed. Job, Barbara; Mehler, Alexander & Sutter, Tilmann. Wiesbaden: Springer VS 2012. [in print]*
   Preprint-URL: http://www.schich.info/pub/2011/Schich_2011_Bielefeld_Preprint.pdf

3. *Schich, Maximilian; Coscia, Michele:* ***Exploring Co-Occurrence on a Meso and Global Level Using Network Analysis and Rule Mining.*** *MLG'11 Proceedings of the Ninth Workshop on Mining and Learning with Graphs. ACM KDD San Diego, CA, USA.*
   URL: http://www.cs.purdue.edu/mlg2011/papers/paper_22.pdf

4. *Blumm, Nicholas; Ghoshal, Gourab; Forró, Zalan; Schich, Maximilian; Bianconi, Ginestra; Bouchaud, Jean-Philippe & Barabási Albert-László:* ***Dynamics of Ranking Processes in Complex Systems.*** *Physical Review Letters [to appear Sept. 4] (2012)*
   URL: http://prl.aps.org/vtoc/PRL/v109

Schich, Maximilian; Meirelles, Isabel; Edmonds, Ernest (eds.):
Special Section: Arts, Humanities, and Complex Networks. [ongoing]
Leonardo Journal 43:3, June 2010, pp. 212, URL: http://www.mitpressjournals.org/toc/leon/43/3
Leonardo Journal 44:3, June 2011, pp. 239-267, URL: http://www.mitpressjournals.org/toc/leon/44/3
Leonardo Journal 45:1, February 2012, pp. 77-89, URL: http://www.mitpressjournals.org/toc/leon/45/1
Leonardo Journal 45:3, June 2012, pp. 275-286, URL: http://www.mitpressjournals.org/toc/leon/45/3
in total analogous to
Schich, Maximilian; Meirelles, Isabel; Malina, Roger (eds.): ***Arts, Humanities, and Complex Networks [Kindle Edition],*** *Cambridge, MA: MIT-Press June 2012 URL: http://ahcncompanion.info/ [enclosed in electronic form only]*

5. *Schich, Maximilian; Lehmann, Sune; Park, Juyong:* ***Dissecting the Canon: Visual Subject Co-Popularity Networks in Art Research.*** *5th European Conference on Complex Systems (online proceedings), Jerusalem, September 3, 2008.*
   URL: http://www.jeruccs2008.org/node/114 [solves an aim after application, before funding]

PDF-ePrints of all publications are enclosed. Publications 1, 2, 3, and 4 will be openly accessible via the Heidelberg ART-Dok-Server at http://archiv.ub.uni-heidelberg.de/artdok/ or will be self-archived at http://www.schich.info; 5 and 6 are copyright Leonardo/ISAST and MIT-Press, with the eBook available for $7.99.

2. Work and results report

2.1 Initial questions and project aims

The *Complex Networks in Art Research* project started with a set of datasets that were collected in the fields of art history and archaeology within the last 60 years. Key aims of the project were to establish an emerging branch of research and work towards a new big picture of arts and humanities by creating a number of practical proofs of concept. Relying on both qualitative expertise and rigorous quantification, the project aimed to avoid predefined limitations such as the disection of history into eras or too specialized inquiry that generalizes from a handful of objects. Key phenomena as recorded in a variety of datasets, such as art historical canons or subject classification, were hypothesized to be inscribed in a large universe of complex networks. The analysis of multiple dimensions of that universe would entangle a wide range of fields into a single approach, dealing with objects, people, locations, eras, and their combinations.

From the beginning, the project aimed to combine methods taken from art history, physics, and computer science to investigate properties and dynamics of relevant complex networks. Planned particular subjects initially included *Archäologische Bibliographie* (Schwarz et al. 2008), the *Census of Antique Works of Art and Architecture Known in the Renaissance* (Census 2005) as well as the so-called *THERMAE dataset* (Schich 2009). Further image, object, and library datasets were planned to be integrated following a broader proof of concept. Key questions targeted the structure and emergence of canon in (sub)disciplines as well as the co-popularity of objects and subject themes.

As a main result for art research and archaeology, the project aimed to provide a prototypical mapping of emerging subject areas. Network science methods were furthermore foreseen to provide a new base for the evaluation of arts and humanities database projects. Last not least, the project was seen as a case study for similar investigations in the growing *Giant Global Graph* of the *Semantic Web*.

2.2 Development of project work

including deviations from the initial concept, scientific setbacks, problems of project organization or technical procedure.

2.2.a Achievement summary

All initial aims, as mentioned above in section 2.1, were met in a flexible but systematic way. Specific outcomes as summarized here are explained in more detail below in section 2.3.a-e.

Published project results include **three practical proofs of concept that combine** research in the **arts and humanities with** the disciplines of **information visualization, computer science, and physics** respectively. A **fourth aspect** is an ongoing exercise of community building that **fosters the establishment of the emerging multidiciplinary research area of *Arts, Humanities, and Complex Networks*** by means of satellite symposia at relevant conferences, an ongoing special section in *Leonardo Journal*, and a growing eBook at *MIT-Press*.

Beyond the published results, I also continued to collaborate with the *Zuccaro* project at *Bibliotheca Hertziana*, the *Max-Planck-Institute* in Rome,[1] started a collaboration working with substantial amounts of provenance data from the *Getty Research Institute*, as well as large person datasets from a number of data providers (see section 2.3.e for the latter two). The results of all these additional efforts will be published in follow-up publications, acknowledging DFG funding as appropriate.

As foreseen, my stay in Boston was very fruitful with regard to academic connections, including my personal embedding in research beyond my home fields of art history and archaeology. A strong indicator for the rising attention towards my approach is the growing share of invited talks that I was allowed to present at venues spanning multiple disciplines, academia and industry, as well as locations in Europe, the US and beyond.[2]

In terms of my career goal of establishing a professorship dealing with complex phenomena in art research I was also successful. In the final month of my fellowship, I was offered and accepted a position as an endowed ***Associate Professor of Art and Technology*** at the ***University of Texas at Dallas***, starting in January 2013.[3] In the meantime, between the expiration of my DFG fellowship and my move to Texas, I am currently hired by Prof. Dirk Helbing, the coordinator of the *FuturICT EU Flagship* proposal, to work on my person data project at his *Chair of Sociology (in particular modeling and simulation)* at *ETH Zurich*.[4]

2.2.b Extended project time

The initial funding time frame of the project from April 2009 to March 2011 was extended to March 2012, due to the birth of our daughter Valeria. While we only learned of this possibility two months before birth, I'd like to express my gratitude to *German Research Foundation* for providing such an internationally unique incentive and service.

1 For preliminary results visualizing the *Social Network of Roman Baroque* see *http://zuccaro.schich.info* (user: zuc, password: zuc); For previous related work at Bibliotheca Hertziana see the 2008 project report at *http://archiv.ub.uni-heidelberg.de/artdok/volltexte/2009/712/*.
2 For a full list of my public speaking activity including selected videos see *http://www.schich.info/en/publications.htm#Talks*.
3 For ATEC's vision check out *http://goo.gl/aMzeX*; for the current state see *http://www.utdallas.edu/atec/about/*.
4 See *http://www.futurict.eu/* and *http://www.soms.ethz.ch*.

2.2.c Deviations from the planned procedure:

After careful pondering, *RGK-Europabibliographie* was not investigated as foreseen. *Archäologische Bibliograhpie* provided enough data for the planned analysis (see section 2.3.b), and the two decade timeframe of *RGK* classification was too short to look at subject evolution. The desicion to focus on one bibliography was also made with an eye to the overall coverage of the project: Given the funding timeframe, I could either focus on a relatively standard collaboration bridging IT and librarianship, or provide a variety of proofs of concept, that would include cutting-edge visualization, computer science, and physics applications, as well as foster the convergence of the *Arts, Humanities, and Complex Networks* community (see sections 2.3.a-d).

After visualizing the *Census of Antique Works of Art and Architecture Known in the Renaissance*, including publication in a prominent *O'Reilly* book and the ongoing *scimap.org* exhibition (see section 2.3.a), I conctacted the *Census project* at *HU-Berlin*, regarding the good news and hoping for potential future collaboration. I never received an official reply.

Temporal irregularity (burstiness) was also investigated. Again, both the *Census* and *Archaeologische Bibliographie* turned out too limited in terms of time-frame, temporal precision, and coarse graining of temporal steps. In the *Thermae dataset*, I found a convincing burstiness effect compared to randomized versions of the data, using *Monte Carlo simulations*. However, due to the data spanning several centuries, the burstiness was heavily intermingled with the way more prominent pattern of exponential growth in data density. This observation provided one of the leads to for a completely different project using much larger person datasets (see section 2.3.e).

Collaborative filtering was discarded in favour of *Hierarchical Link Clustering HLC*, which in 2010/11 was at the forefront of community finding algorithms, and had the additional benefit of being developed locally at BarabásiLab (see section 2.3.b).

Co-popularity was not only analyzed in the *Census* and *Thermae dataset* as foreseen (see section 2.2.d), but also in *Archäologische Bibliographie*, which is an order of magnitude larger (see section 2.3.b).

2.2.d A previously solved aim: Combining Image Matrices and Community Finding

The inclusion of existing image information in the network analysis of art historic databases, i.e. the third task in my initial fellowship application, was actually addressed in the time between the application and the start of my stay in Boston, where I could use part of my 2008 grant from the *Innovation Fund of the President of Max-Planck-Society*. The original Roman Baroque data of my Max-Planck project turned out to be limited, for which reason I went ahead with the *THERMAE dataset* analysis, as planned in my DFG proposal. As a consequence I produced a conference paper titled **Dissecting the Canon: Visual Subject Co-Popularity Networks in Art Research** in collaboration with Sune Lehmann and Juyong Park of BarabásiLab. It was published in the online proceedings of the *European Conference on Complex Systems ECCS2008* in Jerusalem [6]. Subsequently I also presented the work at *Bibliotheca Hertziana* in September 2008, where it provoked a very unusual 90 minute discussion in a room of 80 art historians, with reactions ranging from *brilliant* to *if you were a banker you'd be a locust*. The work furthermore won a Best Poster Award at the *Challenges and Visions in the Social Sciences* workshop, again 2008 in *ETH Zurich*. Finally, during my DFG fellowship in 2010, I collaborated with Yong-Yeol Ahn to improve *image matrix clustering*, going on to present an extended version at the *MIT h+digital* conference, targeting a wider audience. The talk video with the title *Image Matrices – Learning from Klosterneuburg* was watched over 5800 times on *MIT TV* and my own *Vimeo channel*.[5] After the *MIT* talk, the *Getty Research Institute* invited me to start a large follow-up project (see section 2.3.e). The publication of this extended work is planned as a chapter of an upcoming book publication. Further development is planned in the future, when further large image classification datasets become available.[6]

2.2.e Multidisciplinarity as a career risk:

Spending a DFG research fellowship as an art historian at a multidisciplinary physics lab is a highly inspiring and breathtaking post-doc experience. The rare privilege was underscored in my mind for the first time when I realized to be the only humanities fellow among over a hundred DFG and other fellows at a German house meeting in New York City during the first weeks of my fellowship. Looking at co-download statistics between disciplines,[7] I furthermore realized that my choosen combination of fields is way below the threshold of being depicted in an official map. I found myself skiing trees in the backcountry, far from the groomed runs and save boundaries of regular arts and humanities scholarship. While I learned to ski the trees, I constantly met more people that share the vision, fun, and fascination of going from STEM to STEAM, aiming to integrate Science, Technology, Engineering, Arts, and Math. On the other hand I also encountered huge risks in the form of institutional particularism and reviewers that would refute complexity research in the humanities without further reason or readiness for open debate when possible.

Further risks became clear on the way: As multidisciplinary projects take time, I was happy to accept the new DFG parent year extension; this however, took me over the age threshold for a an Emmy Noether application. Applying for other research group fundings, such as Max-Planck (with Bibliotheca Hertziana 2010), ERC (with Bibliotheca Hertziana 2010 and ETH Zurich 2011) or Volkswagen Dilthey (with University of Heidelberg 2011), I unambigously encountered a

5 See *http://techtv.mit.edu/videos/7864-image-matrices-learning-from-klosterneuburg* (5444 views) and *http://vimeo.com/12115271* (361 views).
6 Desirebale German candidates include results of the *http://www.hyperimage.eu* and *http://www.artigo.org/* projects.
7 See Bollen et al. PLoS ONE 2009 (Clickstream Data Yields High-Resolution Maps of Science), *doi:10.1371/journal.pone.0004803*.

minority of categorically refuting super-particularist reviewers that pull the application below the funding threshold, due to the perceived lack of publications in a particular area or their disregard for essential parts of my cv that factually set the base for what I am doing. Projects *opening new areas in longer term and possibly risky projects* that are *multidisciplinary* as well as *high-risk/high-gain* are in reality only desired on paper, but often severely punished in review.

Potentially interesting faculty positions in the growing field of Digital Humanities, both in Germany and the US also turned out be inaccessible. More than once I was told bluntly both inofficially and in interviews that art historians do not belong in social science and literature departments where the Digitial Humanities positions would be based.

On the upside, the experienced risks of multidisciplinarity are set off by a rare breed of supporters, such as in my case Ralf Biering, Vinzenz Brinkmann, Horst Bredekamp, Max-Planck, Albert-László Barabási, DFG, Dirk Helbing, Roger Malina, and ATEC visionaries of the University of Texas at Dallas.

2.3 Description of results achieved

and discussion with regard to the relevant state of the art, possible application perspectives, and follow-up investigations.

The field of *Digital Humanities* is often critizised for either it's narrow focus on literature and/or not delivering striking results from quantification, essentially being stuck in IT service and observation instead of explanation.[8] The proofs of concept produced in this project bring evidence that the study of *complexity* and *networks* in the arts and humanities can indeed create vigorous insights using large datasets, combining the qualitative with the quantitative and the general with the special.[9] In fact, the examples show that visualization sometimes beats textual description, that computer science *pipelines* can reveal patterns which are opaque to single researchers, and that physics and advanced math can provide better general models of cultural processes than our educated intuition. The following sections introduce proofs of concept along these lines, including some follow-up projects. Selected literature is listed at the end of the report.

2.3.a Ubiquity of Complex Networks in Art Research – A Visualization Proof of Concept

In April 2009, a week after I arrived at BarabasiLab, I received an email from Tim O'Reilly, *Google*'s Chris DiBona, and *Nature*'s Timo Hannay, inviting me *to join 200 people from around the world who are doing groundbreaking work in diverse areas of science and technology*. The group was to convene for the *SciFoo2009* [10] unconference at the Googleplex in Mountain View. Using the opportunity, I conceived and ran a session titled the *coral reef of culture* with complex systems scientist Brian Uzzi and Oxford art historian Martin Kemp. Going on to discuss some of my preliminary visualizations with *O'Reilly* editor Mike Henderson, the idea formed that it is useful and desireable for a wider audience to map and explain databases in their entirety. As a consequence the necessary task of mapping all inherent complex networks in an arts and humanities database, preceding their analysis with network science methods, became a proof of concept of it's own.

The initial result of the following effort was a map of the *Census of Antique Works of Art and Architecture Known in the Renaissance* (fig. 1), whose genesis is explained in a chapter of O'Reilly's *Beautiful Visualization*, a highly popular collection of case studies by world class practitioners in information visualization.[11] The *Census* map also passed peer-review and has been added to the touring *scimaps.org* exhibit, whose next stop is the *National Academy of the Sciences* in Washington D.C.[12]

A second follow-up article uses simple words to expand the concept of database mapping. It highlights arts and humanities databases as *networks of complex networks* that are embedded in an even larger *network of datasets*, which currently becomes more and more explicit in *Linked Open Data* and other forms of the *Semantic Web*. The second article was submitted and accepted in March 2011 and will soon appear in a German *Springer VS* volume upon invitation of the sociologist editors Barbara Job, Alexander Mehler, and Tilman Sutter. The English version is part of an ongoing book project (see section 2.3.e), and there is also an open request for a Chinese version in the *The Journal of Chinese National Art*. The following two abstracts introduce both the *O'Reilly* and *Springer VS* articles.

> The **Revealing Matrices** book chapter uncovers some nonintuitive structures in curated databases arising from local activity by the curators as well as from the heterogeneity of the source data. Our example is taken from the fields of art history and archaeology, as these are my trained areas of expertise. However, the findings I present here—namely, that it is possible to visualize the complex structures of databases—can also be demonstrated for many other structured data collections, including biological research databases and massive collaborative efforts such as DBpedia, Freebase, or the Semantic Web. All these data collections share a number of properties, which are not straightforward but are important if we want to make use of the recorded

8 See for e.g. Thomas Thiel FAZ 24.7.2012 (Digital Humanities–Eine empirische Wende für die Geisteswissenschaften?) *http://www.faz.net/-gsn-71khe* or Martin Mueller CSCDC Reflections 8.2.2012 (Stanley Fish and the Digital Humanities) *http://cscdc.northwestern.edu/blog/?p=332*.

9 Reviewers often desire a decision between *nomothetic* and *ideographic* approaches. *Complexity* integrates both as non-random general patterns emerge from a large number of specifics. Beyond *mean field* generalizations and intelectual particularism both sides are inseparable.

10 Cf. *http://singularityhub.com/2009/07/14/the-hub-goes-to-scifoo-2009-it-was-awesome-scifoo09/*

11 See Nathan Yau's review on his influential visualization blog at *http://flowingdata.com/2011/03/09/review-beautiful-visualization-looking-at-data-through-the-eyes-of-experts/*

12 See *http://scimaps.org/maps/map/the_census_of_antiqu_129/*

data or if we have to decide where and how our energies and funds should be spent in improving them. In conclusion, the chapter illustrates that enriched and refined data model matrices are very useful for database project evaluation, exposing many nonintuitive data properties that are hard to uncover by simply using the database or looking at the commonly used indicators of quality. As data becomes more accessible in the form of Linked Data, RDF graphs, or open dumps of relational tables, the presented methods can be applied by funders or the projects themselves, within a very short time frame in a mostly automated process.

The visualizations shown in the chapter present the first comprehensive big picture of the entire CENSUS database, where we can see the initial data model definition as well as the emerging complex structure in the collected data. By looking at the visualizations, we found out that many of the numbers given in the project description were incomplete or even misleading. Some of the new numbers may be smaller than the initially presented ones, but as we have learned from our analysis, sometimes a little less is more—and more is different (Anderson 1972).

The follow-up article on **Networks of Complex Networks in (Art) Research** [german: *Netzwerke komplexer Netzwerke in den (Kunst)Wissenschaften*] explains the relevance of complex networks in arts and humanities data in simple words. The article shows that virtually any structured database from simple indices to all-encompassing datasets of datasets such as the Linked Open Data Cloud can be understood as networks of complex networks. In addition the article goes beyond structured data, explaining how the Google Books project, and other text digitization and markup projects for that matter, will converge towards semi-structured and structured data, i.e. another network of complex networks. First, Google extracts n-grams (i.e. word-combinations) from books, it then attaches meanings, such as geo-coordinates to the n-grams, and finally it connects these concepts in a network with multiple node and link types. In conclusion the article makes a strong case against ignoring complex network phenomena within the realm of arts and humanities.

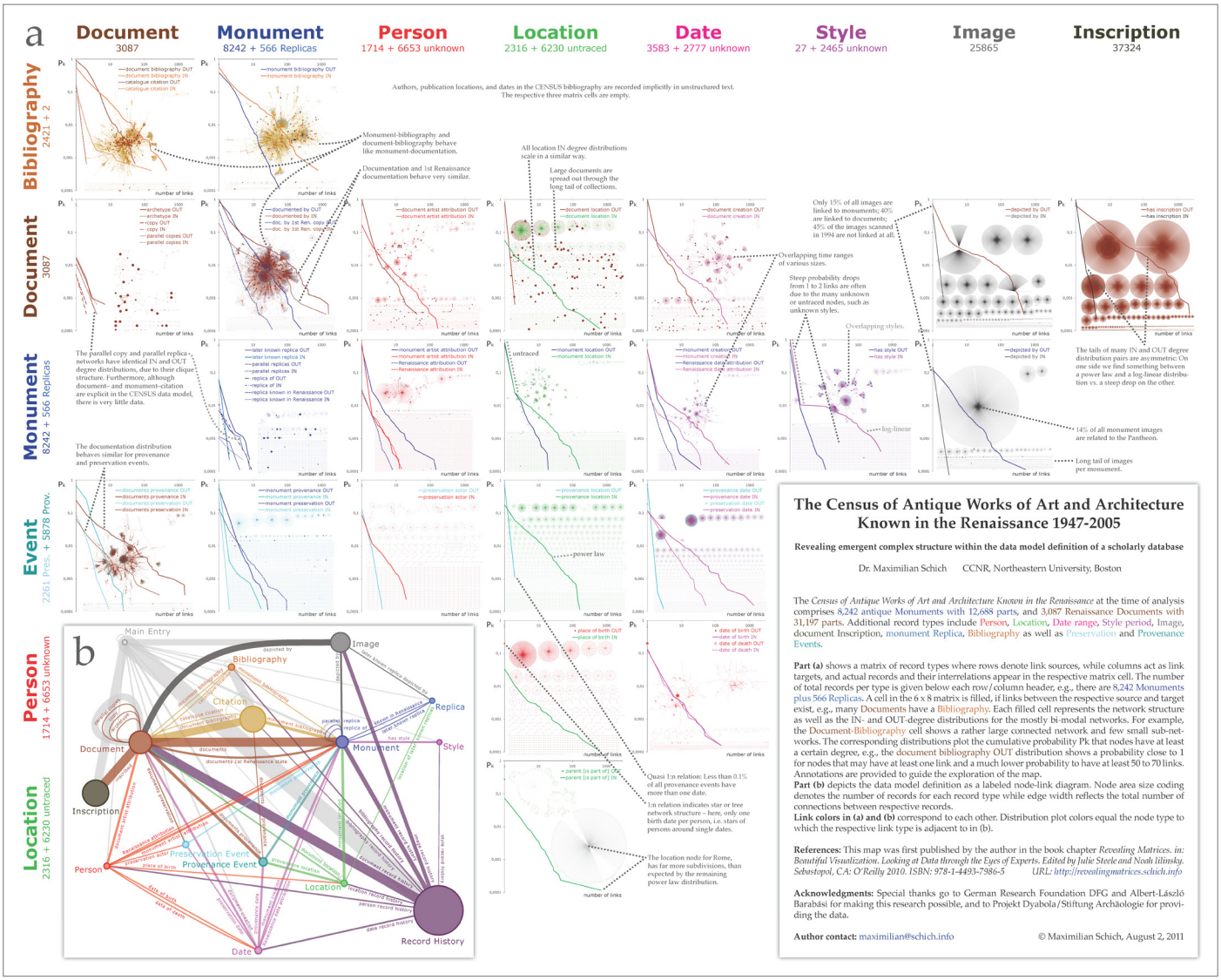
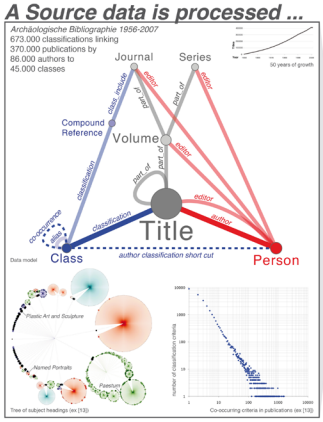


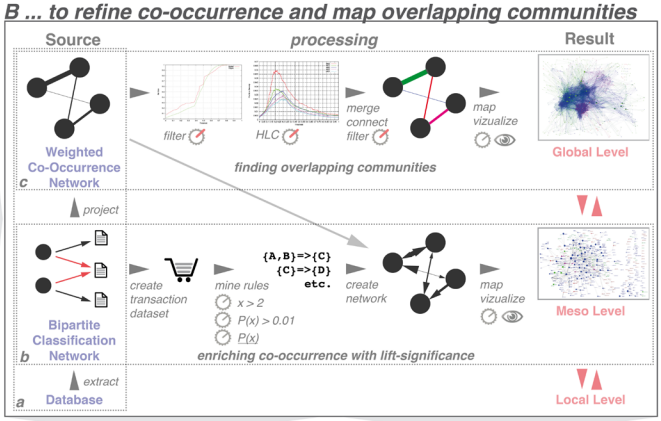
Figure 2.3.a. A network of complex networks: The CENSUS map depicts the entire content of a large database within its own data model definition. Emerging complex structure is revealed, facilitating fair evaluation. The original size is 76 x 61 cm at 300 dpi.

An Emerging Big Picture of Classical Archaeology as a Complex System

Maximilian Schich
CCNR, Northeastern University
110 Forsyth Street, Boston MA 02115, USA
+1 (617) 817-7880 – maximilian@schich.info

Michele Coscia
KDDLab, University of Pisa
Largo B. Pontecorvo 3, 56125 Pisa, Italy
+39 (050) 2213-3136 – coscia@di.unipi.it





Abstract

In this poster we present a big picture of classical archaeology, depicting a complex ecosystem of about 45.000 subject themes and circa 88.000 authors. Using data from *Archäologische Bibliographie* [14], a database that records and classifies relevant scholarly literature since 1956, we construct networks of subject co-classification and author co-interest, further analyzing their complex interplay and evolution over 50 years. Using complex systems science methods, such as HLC community finding [2], Vespignani graph filtering [3], and association rule mining [1] – a pipeline introduced in a previous proof of concept [*] – we show that it is indeed possible to navigate the complex system in question from the granular to a global perspective over a number of levels, revealing a clear and significant difference of the factual emerging structure to *a priori* definitions of classical archaeology, as they become evident in the now half a century old *tree of subject headings* in *Archäologische Bibliographie*, or the typical shelf configuration of an archaeological library.

The example visualization that acts as the centerpiece of our poster, first takes the observer from the level of single subject themes (A, B), over group/ego-networks of co-classified subjects, enriched by a rule-mined link-significance (C, D, E, I), all the way up to a network of subject theme communities that emerge from HLC community finding (F, G, H, I).

Finally we make clear, that to go even further up, it is not sufficient to threshold the community finding on a more coarse grained level, but that it would make sense to recursively run the community finding again on the network of subject theme communities, in order to arrive from the actors (i.e. subject themes), over cities and overlapping countries (i.e. subject theme communities), at the conceptual continents of the academic discipline.

Similar visualizations and browsable sets that can be explored by the respective scholars are producable for any given library classification such as taken from arXiv, OCLC, Europeana, and maybe in not too far a future even from Google Books. As such, the poster exemplifies the usefulness of complex systems approaches, enabling a wider audience to explore and understand the complexity they are exposed to every day.

Walk through the figures

The figure sequence makes clear how we enable meso- and global-level exploration of subject classification beyond the standard user interface of common bibliographies:

A – A data model sketch for our source dataset *Archäologische Bibliographie*, including **(below left)** a visualization of the inherent *tree of subject headings* with general subject themes, locations, persons, periods, and events; **(below right)** the fat-tail distribution for classification co-occurrence in publications (both taken from [13]); **(up right)** an indication of dataset growth 1956-2007.

B – Our data analysis and visualization pipeline includes **(a)** a one-mode projection of the publication–classification or author–classification network to classification co-occurrence, **(b)** the creation and visualization of *rule-mined directed lift-significance weights* in addition to the regular *co-occurrence link weight*, and **(c)** the creation and visualization of an overlapping community network, using Vespignani backbone filtering and Hierarchical Link Clustering (HLC). The **(inset plots)** in pipeline c show a backbone filtering phase transition, where edges disappear quicker than nodes, and HLC partition density distributions for five decades 1967-2007, both of which provide meaningful thresholds for further processing.

C – An example for *classification co-occurrence in publications with lift-significance* for the branch *Plastic Art and Sculpture*, i.e. a subset of classifications in the *tree of subject headings of Archäologische Bibliographie*. The picture, which can be seen as an instant cheat sheet for an imaginary archaeology exam, is a simple merge of two identical networks, thresholded in different ways: Heavy co-occurring links are taken into account if they contain at least 4 publications, while additional links are included if their lift-significance is at least 0.056.

D – Another example, equivalent to figure C, where *Named Portraits* mutually define themselves and connect to a coherent story from the Roman republic to the end of the empire.

E – **(above)** An era structure dendrogram of *classification co-occurrence in publications*, where algorithmically computed eras are colored in the tree, while our arbitrarily chosen decades are highlighted in the x-axis labels. **(below)** Classification co-occurrence evolution includes initially highly significant, i.e. dark, links that become less significant and wider as they accumulate literature over five decades. Exceptions point to controversial or special topic clusters, such as the 4-clique around *Skylla*, *Polyphemos*, *Pasquino group* and *rape of the palladium* in figure C that stays highly significant over several decades.

F – Details of the global *class community overlap network* for co-occurrence in publications and authors, with nodes represented as pie diagrams and edges split by color, indicating the inherent frequency of classification supertypes, i.e. subject themes, locations, periods, persons and objects.

G – Isolating links in the community overlap network F by supertype color, corresponding to subject themes, locations, periods, and persons, reveals that the are distributed in very different ways.

H – **(left)** Both classification co-occurrence in publications as well as in authors evolve over time by fleshing out structure that emerges early on in the process. **(right)** Communities belonging to various temporal snapshots, can be connected using a dedicated algorithm that reveals interesting merges and splits over time, indicating both diversification and specialization.

I – Combining global and meso-level exploration by zooming into overlapping communities containing a given classification – here the Italian site *Paestum* – uncovers its meaning even to the uneducated eye, clearly improving over a simple ego-networks that often present themselves as a massive featureless clique.

Conclusion

In this poster we have shown that subject themes in classical archaeology, as recorded in *Archäologische Bibliographie*, are granular components of a complex system, which we can explore both on a meso level (where themes are connected by co-occurrence) as well as on a global level (where theme communities are connected by theme overlap).

Figures F and G point to another, even more global level, as we can spot obvious clusters with our bare eyes. In order to extract these conceptual continents of the academic discipline, which might be more manifold in other datasets, it seems natural to run our pipeline c for another time.

Reference summary

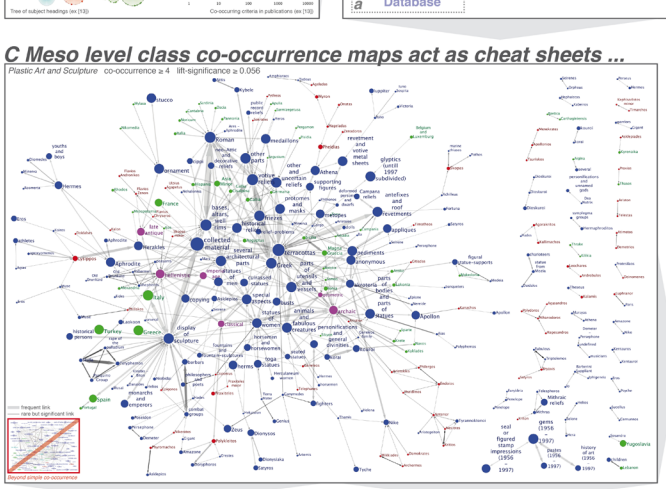
[1] Agrawal Imielinski Swami SIGMOD1993 (Mining Association Rules ...)
[2] Ahn Bagrow Lehmann NATURE 2010 (Link Communities Reveal Multiscale Complexity...)
[3] Angeles-Serrano Boguna Vespignani PNAS 2009 (Extracting the multiscale backbone ...)
[4] Berlingerio Coscia Giannotti Monreale Pedreschi PAKDD 2010 (As Time Goes By ...)
[5] Evans Lambiotte PRE 2009 (... and Overlapping Communities ...)
[6] Ferlez Faloutsos Leskovec Mladenic Grobelnik MDL ICDE 2008 (... Network Evolution ...)
[7] Fortunato PHYSICS REPORTS 2010 (Community detection in graphs)
[8] Fortunato Barthelemy PNAS 2006 (Resolution limit in community detection)
[9] Hipp Güntzer Nakhaeizadeh ACM SIGKDD 2000 (Algorithms for association rule mining)
[10] Newman PNAS 2006 (Modularity and community structure in networks)
[11] Porter Onnela Mucha NOTICES OF THE AMS 2009 (Communities in networks)
[12] Roelleke Wang ACM SIGIR 2008 (TF-IDF uncovered)
[13] Schich Hidalgo Lehmann Park BARCH ON-LINE 2009 (... Subject Co-Popularity ...)
[14] Schwarz et al., Archäologische Bibliographie. Online-Database. Munich: Biering & Brinkmann, 1956-2011 URL: http://www.dyabola.de (Update February 2008)
[15] Shannon et al. GENOME RES. 2003 (Cytoscape)

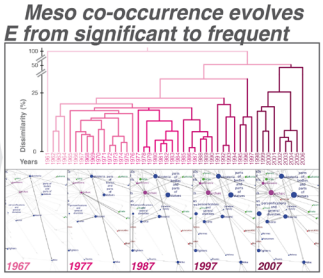
Acknowledgements

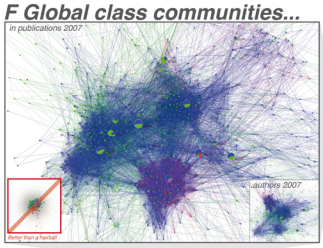
Maximilian Schich is a German Research Foundation (DFG) fellow, hosted by Albert-László Barabási; Michele Coscia is a recipient of the Google Europe Fellowship in Social Computing, and this research is supported in part by this Google Fellowship.

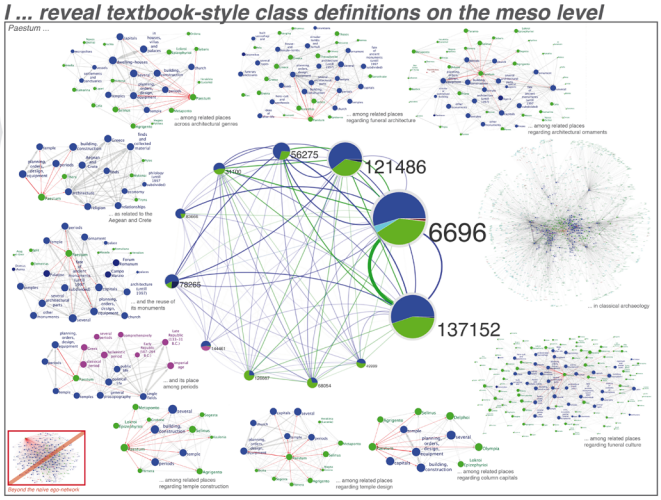
* This work underwent peer-review in multiple related disciplines. Besides to the complex systems crowd at ECCS 2011 in Vienna, Austria, it has also been presented to an audience of archaeologists at CAA 2011 in Beijing, China, while being accepted for publication at a major venue in the field of data mining at KDD MLG'11, San Diego, CA, USA. URL: http://www.cs.purdue.edu/mlg2011/papers/paper_22.pdf

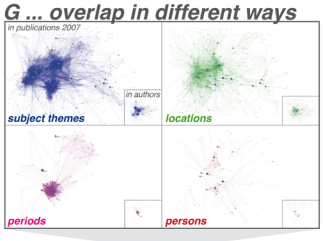
Full paper: http://goo.gl/Osrl5











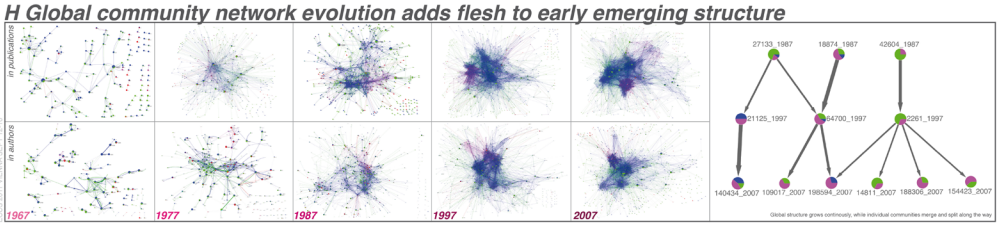


Figure 2.3.b. A big picture of classical archaeology as a complex system (depicted in an A0 poster sized 1189 x 841 cm at 300 dpi).

2.3.b Finding Patterns in Complex Overlap – A Computer Science Proof of Concept

In winter 2009/10, Fosca Gianotti and Dino Pedreschi of *ISTI-CNR* and *University of Pisa*, two specialists in *knowledge discovery and data mining*, were spending a sabattical at *BarabásiLab*. Upon introduction of my project, in particular my goal to map the big picture of classical archaeology, they put me into contact with Michele Coscia, a talented Ph.D. student in computer science with an interest in humanities topics. Funded from Italy, Michele eventually joined me in Boston, where we collaborated closely over a year to find patterns and insight in archaeologic subject classification, developing a prototypical *processing pipeline* to filter and clarify what turned out to be a dense mess of conceptual overlap. The results of our effort were published in the proceedings of the *MLG* workshop at the *ACM KDD 2011* conference in San Diego, a prestigious publication venue in computer science that combines data mining and machine learning with graph theory and complex network research.

Besides it's attractivity within computer science, the publication venue was chosen for it's short turnaround time, based on my negative experience with a related contribution to the *Annual Meeting of the International Association for Classical Archaeology AIAC 2008*: While the turn-around time in archaeology (after paper acceptance!) came close to three years, the *MLG* paper was published right after peer-review in concurrence with the KDD conference. It is noteworthy that our work passed peer-review multiple times in several related fields: Besides our *MLG* computer science paper, we also presented a long talk at the *39th Annual Conference of Computer Applications and Quantitative Methods in Archaeology CAA2011* in Beijing, as well as a peer-reviewed poster at the *European Conference on Complex Systems ECCS2011* in Vienna.

As a sidenote, we chose the poster format on purpose over submitting a talk, due to the visualization content of our work. Ironically, this may sound strange in the field of art history, where poster sessions are often some sort of a condolence for grad and undergrad participants. In complex network science and complexity science in general, poster sessions are often well organized, well located, and visible for the whole conference. Done well, a poster can reach a much higher impact and arouse way more attention than a well-hidden contributed talk in a parallel session.

The following abstract summarizes the results of our MLG paper. For a less technical audience, we translated the technical title *Exploring Co-Occurrence on a Meso and Global Level Using Network Analysis and Rule Mining* to *An Emerging Big Picture of Classical Archaeology as a Complex System*.

> In our **MLG KDD 2011 paper** we present a big picture of classical archaeology, depicting a complex ecosystem of about 45.000 subject themes. Using data from *Archäologische Bibliographie* (Schwarz et al. 2008), a database that records and classifies relevant scholarly literature since 1956, we construct networks of subject co-classification across publications and authors, further analyzing their complex interplay and evolution over 50 years. Using complex systems science methods, such as *HLC community finding* (Ahn Bagrow Lehmann 2010), *Vespignani graph filtering* (Angeles-Serrano Boguna Vespignani 2009), and *association rule mining* (Agrawal Imielinski Swami 1993) in an integrated pipeline, we show that it is indeed possible to navigate the complex system in question from the granular to a global perspective over a number of levels, revealing a clear and significant difference of the factual emerging structure to a priori definitions of classical archaeology, as they become evident in the now half a century old tree of subject headings in *Archäologische Bibliographie*, or the typical shelf configuration of an archaeologic library. Example visualizations in the paper and posters take the observer from the level of single subject themes, over group and ego-networks of co-classified subjects, enriched by a *rule-mined link-significance*, all the way up to a network of subject theme communities that emerge from HLC community finding. Finally we make clear, that to go even further up, it is not sufficient to threshold the community finding on a more coarse grained level, but that it would make sense to recursively run the community finding again on the network of subject theme communities, in order to arrive from the actors (i.e. subject themes), over cities and overlapping countries (i.e. subject theme communities), at the conceptual continents of the academic discipline of classical archaeology. Similar visualizations and browsable sets that can be explored by the respective scholars are produceable for any given library classification such as taken from arXiv, OCLC, Europeana, and maybe in not too far a future even from Google Books. As such, the MLG paper exemplifies the usefulness of complex systems approaches, enabling a wider audience to explore and understand the complexity we are exposed to every day.

2.3.c A Continuum Theory of Icons, Canons, and Cacophanies – A Physics Proof of Concept

In 2009, I presented some previous work at *BarabásiLab*, which included my observation that the most documented Roman monuments, such as the *Colosseum* or the *Laocoon*, stay on top of the frequency list for several centuries, while less popular objects appear to be less stable in rank, fluctuating around each other more and more as we go down the list (cf. Schich ECCS 2008, Schich 2009, Crandall 2009). Inspired by this insight, we embarked on a project that aimed to explain and understand the dynamics of canon in art. During the canon project we realized that the available art data is neither large nor consistent enough solve the problem. We therefore refocused on the *dynamics of ranking in complex systems* in general, looking at a variety of datasets, in which the number of ranked items and their scores span several orders of magnitude. In the end, the mathematical intricacies of the canon project involved two people in László Barabási's group for more than two years: computer scientist Nicholas Blumm and physicist Gourab Ghoshal. My own role, besides co-developing the initial hypothesis with Gourab, was to provide quantitative as well as qualitative observations, based on my arts and humanities expertise. In addition the canon project counted on contributions of

econo-physics student Zalàn Forrò, as well as the outstanding mathematical expertise of physicists Ginestra Bianconi, and Jean-Philippe Bouchaud, who is also the chairman of *Capital Fund Management*, a well-known hedge fund. The result of the canon project is a *continuum theory* that is accepted and scheduled to be published on September 4, 2012 in *Physical Review Letters PRL*, a prestigious journal in the field of physics. The following text will close the loop in a separate publication, translating the content and relevance of the *PRL* paper to a wider audience. We think it will transform the notion of canons in art and beyond:

Physics provides an important message for the arts and humanities connecting icons and canons with cacophonies of fashions and fads: Focusing on eleven striking examples from *Christ to Coke* Martin Kemp recently claimed it makes sense to separate outstanding visual icons from a broader canon of art and design (Kemp 2011). Inspired by the problem of documentation frequency in art history (Schich 2009), a group of physicists led by Albert-Lászlo Barabási now indicates that this perspective makes sense indeed (Blumm et al. 2012).

Looking at the *Dynamics of ranking processes in complex systems* we develop a theory with three different regimes or phases where ranked items are either totally rank stable, more or less score stable, or completely unstable. Rank within this circumstance denotes the list position of an item at a given point in time, while score corresponds to the quantity on which the ranking is based. In simple cases, score would be a frequency, such as the countable citations of a scientific paper. In the worst case score is a diffuse qualitative measure of importance, such as the attraction of Leonardo da Vinci's Mona Lisa or the Colosseum in Rome.

Taking into account a variety of large data-sets we find two classes of systems, all of which contain scores spanning several orders of magnitude: Among the stable systems we find English word usage over two centuries, the frequency of diseases in Medicare data, and the market capitalization of companies. Unstable systems include the frequencies of Twitter-hash-tags, site visits in Wikipedia, and somewhat surprisingly scientific citations.

In unstable systems, even the popular items, such as the hash-tag for #love, can change from top ranked at one moment to insignificant and back at another point in time. Stability on the other hand comes in two forms: Rank stability simply means the respective items don't change their rank over time. Score stability refers to items that fluctuate around a mean level or fitness, based on intrinsic quality, central location, or other external factors. As rank is a relative measure over all items in the system, it can change even while the score of an item stays constant.

In fact, due to the broad heterogeneity of scores, rank stable items, like icons in art history, turn out to be extremely rare and, if there at all, are confined to the most outstanding scores at the top of the list (Ghoshal Barabási 2011, Blumm et al. 2012). Further down the list, ranks start to fluctuate more and more. This must be disturbing to art historians and humanists, as canons are traditionally imagined as thresholded bodies of most important items (Brzyski et al. 2007): Due to the growing fluctuations in rank as we go further down, any such threshold must be arbitrary.

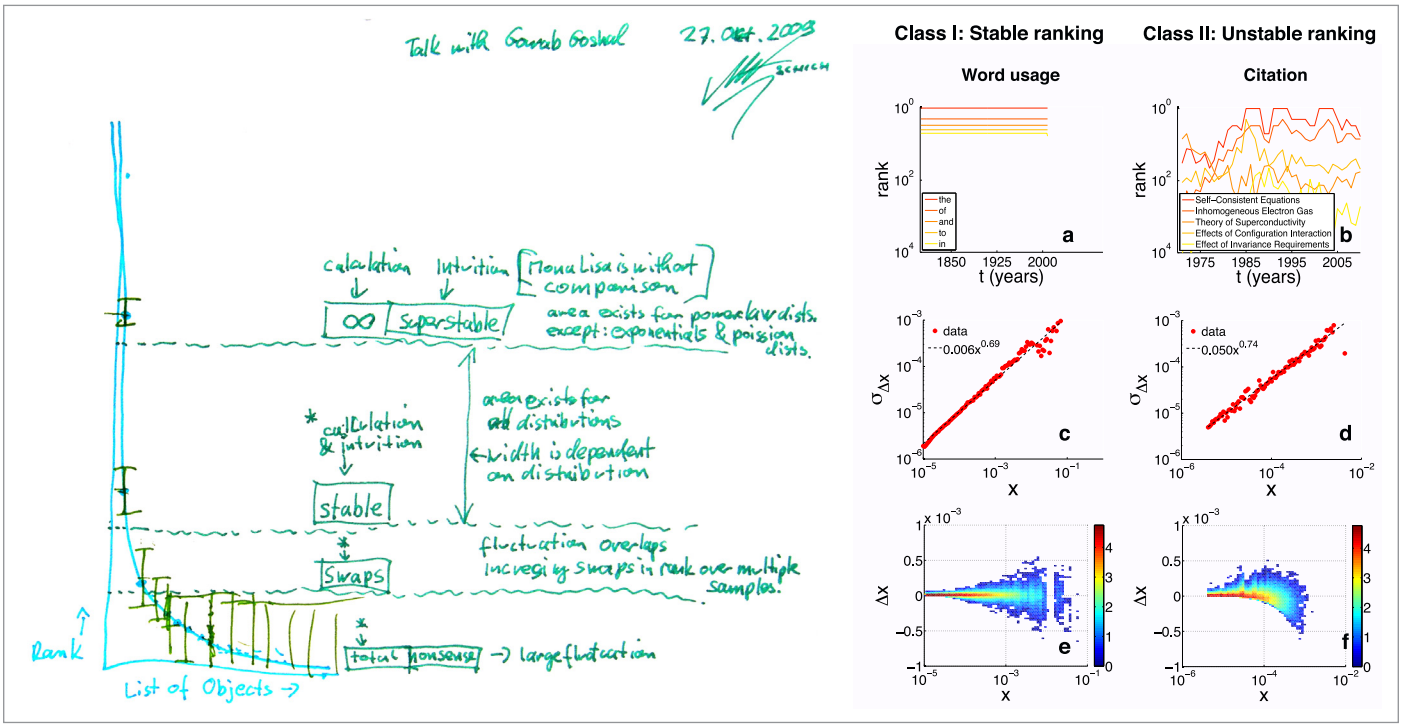


Figure 2.3.c. Redefining our concept of canon: (left) Records from an October 2009 conversation that integrated my qualitative and quantitative observations in art data with Gourab Ghoshal's insights from numerical simulations into a preliminary hypothesis of universal ranking properties; (right) Observed ranking dynamics in two large systems from our 2012 PRL paper presenting in a continuum theory of ranking in complex systems.

Score stable items as found in our analysis provide a welcome alternative definition of canon as they fluctuate around a mean fitness. The difference is that score stable items can populate the entire score distribution across several orders of magnitude. As a consequence, even a very rare item can belong to the canon if it has a consistent probability, no matter how small, of appearing within the scope or frame of the system.

One of the most inspiring parts of the PRL paper is that stable systems can transition into unstable systems by the induction of noise, much as when a solid melts due to thermal agitation. This indicates that in art history a large scale quantification of diffuse qualitative measures may be able to explain the transition between classic and contemporary art; Scholars may be able to redefine the canon of their discipline less by focusing on the intrinsic quality of neglected objects, than by simply raising the noise within their chosen frame of reference; Finally, investors in contemporary art or fashion may be interested to actively dampen the existing noise in order to stabilize the emerging canons of the future.

2.3.d Arts, Humanities, and Complex Networks – A Community Building Proof of Concept

Upon my arrival in Boston in April 2009, I approached László Barabási with the idea of a symposium series that would act as a radio beacon, fostering the emerging branch of complex network research in the arts and humanities by bringing together a large variety of practitioners that would not meet otherwise. With both László's recommendation and financial backing, I started to collaborate closely and enthusiatically with Roger Malina and Isabel Meirelles, temporarily taking in further co-organizers Gourab Ghoshal, Sune Lehmann, Riley Crane, Tijana Stepanovic, and Cristián Huepe.[13] Within the effort, a dedicated special section in *Leonardo Journal* as well as a growing *MIT-Press eBook* were born [5].[14] Isabel Meirelles and myself went on to become Editorial Advisors in the board of Leonardo Publications.[15] The published outcome provides a growing one-stop overview of our emerging research area, in lieu or as a prliminary step towards a dedicated scientific journal. A companion website, official talk videos, a recent *MIT-Press* podcast, as well as some fantastic reviews propagate the message further.[16] The next incarnation of our event series will take place at *NetSci2013* in Copenhagen.

The *Arts, Humanities, and Complex Networks eBook* is a first anthology of articles to foster the emerging convergence of arts, humanities, and complex networks in *Leonardo Journal*. Considering its current exponential growth, the book cannot and does not present an exhaustive account of all relevant aspects of the field. That said we firmly believe that this selection highlights many key questions. The articles cover a kaleidoscope of different approaches, ranging from vigorous humanistic inquiry and pure natural science to free artistic expression. Our goal is to provide a useful overview of a large variety of specializations as well as new and interesting collaborations. The anthology brings together a selection of articles from four symposia, three on *Arts, Humanities, and Complex Networks* and one on *High Throughput Humanities*. The latter was a satellite event at the *European Conference on Complex Systems 2010* in Lisbon/Portugal; the former is an ongoing satellite series at *NetSci – the International School and Conference on Network Science*, with articles selected from events that took place in Boston/MA 2010, Budapest/Hungary 2011, as well as Evanston/IL 2012. Two further articles were selected from direct submissions to *Leonardo Transactions*.[17] We have loosely organized the selected articles in a number of self-emerging groups, including **Networks in Culture**, **Networks in Art**, **Networks in the Humanities**, **Art About Networks**, and **Research in Network Visualizations**. The groups overlap and combine a great variety of perspectives. As we approach the common goal to measure, visualize, and understand complex networks in the arts and humanities, we look forward to documenting the further growth of this diversity of approaches in future incarnations of the eBook.

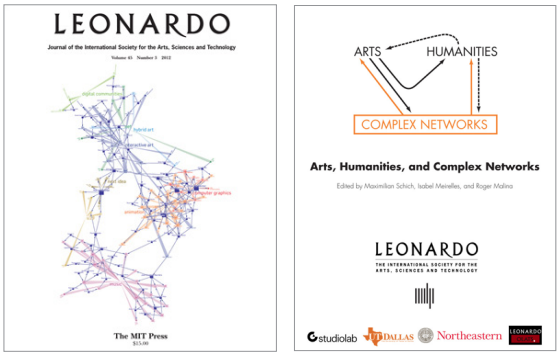


Figure 2.3.d. Covers of Leonardo Journal and our MIT-Press eBook

13 For the symposia websites see *http://artshumanities.netsci2010.net*, *http://artshumanities.netsci2011.net*, *http://artshumanities.netsci2012.net*, as well as *http://hth.eccs2010.eu*.

14 Note: The eBook is enclosed in electronic Kindle format and requires the Amazon Kindle reader (URL); All articles are accessible in PDF format via Leonardo Journal, and as such are properly peer reviewed and indexed in the ISI web of science.

15 See *http://www.leonardo.info/isast/journal/editorial/edboards.html*.

16 For abstracts and videos see *http://www.ahcncompanion.info*; MIT-podcast see *http://www.mitpressjournals.org/page/podcast_episode17_LEON* (over 500 listeners); Reviews see *http://www.netz-werker.blogspot.ch/2010/09/eccs-2010-lisbon.html*, dito *.../2011/06/arts-humanities-and-complex-networks-at.html* and *http://meredithtromble.net/aas/2012/07/22/arts-humanities-and-complex-networks/*.

17 Submissions and peer review are usually done using the EasyChair conference management system. Alternatives are direct submissions via email to *artshumanities.netsci@gmail.com* or less direct via *http://www.leonardo-transactions.com/*.

2.3.e Follow-up projects – *Getty Provenance Index Visualization, a Topography of Desire, and a textbook.*

While at BarabasiLab from 2009 to 2012, I have accumulated a number of interesting research questions, which I plan to address in my future research. Considerable amounts of time and effort in the role of first author have already been spent on three particular projects: The *Getty Provenance Index Visualization* project, a *Topography of Desire* based on large person data, as well as a wider audience book publication. For space reasons I mention all these projects in very short paragraphs. However, as a considerable amount of related work has been spent during my fellowship, I intend to use the remaining DFG funding to cover open-access publication costs.

The *Getty Provenance Index Visualization* project, is a collaboration with Piotr Adamczyk (Metropolitan Museum) and Lev Manovich (UC San Diego / calit2) upon invitation of and with Christian Huemer (Getty Research Institute). We already started to analyze several hundred thousand sales transactions in the British, French, Belgian, and Dutch art markets from 1801-1820, eventually to broaden our focus to the entire *Getty Provenance Index*. Currently we are invited to add a book chapter in an upcoming volume edited by Hans Migroet, Neil DeMarchi, and Sophie Raux, three central figures in art market research. Furthermore we plan to produce an open-access article that will map the choosen markets in all their network dimensions. A draft manuscipt and figure sequence is already in existence.

A *Topography of Desire* (working title) looks at networks of cultural centers based on several large person datasets, recording 20.000 to 160.000 noted people and artists, where we know both the birth and death locations over a time range of 2500 years. Data providers include public sources such as *Wikipedia* and *Freebase.com* as well as proprietory sources such as *Verlag Walther DeGruyter*, the *Getty Research Institute*, plus optionally many more. Collaborators include physicists Yong-Yeol Ahn, Albert-László Barabasi, Dirk Helbing, Alexander Mirsky, Chaoming Song, and cinematic information designer Mauro Martino. In terms of results we aim for a quantification of cultural history, both building on and extending upon existing research on human mobility, urban scaling, and historic demography. I am currently summarizing results in a preliminary manuscript, working in the group of Dirk Helbing at ETH Zurich.

A wider audience book publication on *Complex Networks in Art History* (working title) and or a respective review article will be the outcome of my initial teaching engagements at *UT Dallas* in 2013. I have developed the draft syllabus during my *DFG fellowship* in Boston, in preparation of a variety of talks as well as more integrated teaching engagements with 3 to 14 hours of lecture. With *DFG* approval, I thought twice at the *Department for Image Science* at *Danube-University* in Krems/Austria. In summer 2012 I am invited to teach at the *Université del Eté* in Yverdon-les-Bains, Switzerland, as well as at the excellence-funded *Marsilius-Kollegg* at *University of Heidelberg*, Germany.

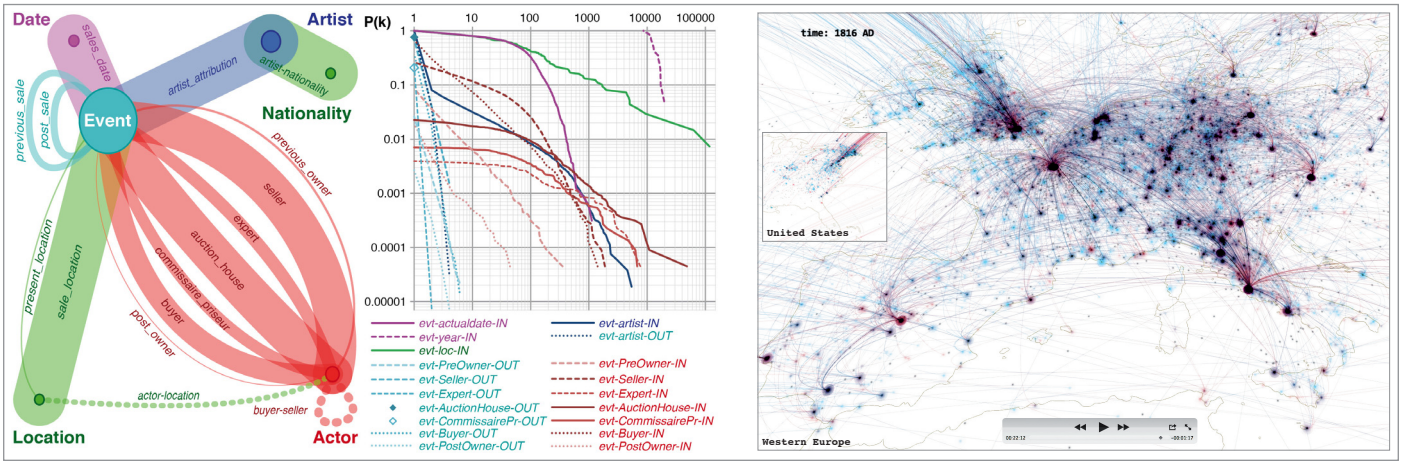


Figure 2.3.e. Initial follow-up results: (left) The Getty Provenance Index is a network of complex networks: data model and corresponding degree distributions; (right) Birth–death records of noted people reveal a dynamic topography of desire: Europe and the US in 1816.

2.4 Statement if results are commercially exploitable, if such exploitation is done or to be expected.

As stated above in section 2.3.a, the visualization of entire databases as *networks of complex networks* enables much better evaluation of database projects in the arts, humanities, and beyond. This can be exploited and may even be comercially viable in the biomedical field, where customers may be interested in the full structure of datasets auch as *Flybase*, etc.; an additional application is the visualization of data in the *Linked Data Cloud* or other *Semantic Web* datasets.[18] A respective funding application that exlores this avenue further is in the planning phase.

As mentioned in section 2.3.c, the results of the *PRL* paper on *ranking dynamics in complex systems* may have consequences for art investment and consulting. No such exploitation is done or planned so far. The first author Nicholas Blumm recently joined *Ginger.io*,[19] a company that aims to predict upcoming diseases, based on the patient behaviour and other indicators measured by cell phone sensors. As the ranking dynamics of disease frequency in Medicare data are one example of our PRL paper, there may be an indirect application.

18 See for e.g. *http://flybase.org/*, *http://linkeddata.org/*, and *http://ckan.org/instances/*.
19 See *http://ginger.io/*

2.5 Who has contributed to project results (cooperation partners in Germany and abroad, project collaborators etc.).

Collaborations are also mentioned in context above. Local collaborators contributing to results at Northeastern University include Albert-László Barabási, Ginestra Bianconi, Nicholas Blumm, Gourab Ghoshal, Sune Lehmann, and Isabel Meirelles; remote collaborators include Jean-Philippe Bouchaud (Capital Fund Management), Michele Coscia (Uni Pisa), Riley Crane (MIT Media Lab), Ernest Edmonds (Leonardo Transactions), Zalan Forrò (ETH Zurich), Cristián Huepe (Northwestern University), Roger Malina (Leonardo/ISAST), and Tijana Stepanivic (Museum Ludwig Budapest).

Without stuffing the list, I'd like to acknowledge further collaborators that contributed to preliminary as well as ongoing work: Piotr Adamczyk (Metropolitan Museum), Yong-Yeol Ahn (University of Indiana Bloomington), Dave Crandall (University of Indiana Bloomington), Dirk Helbing (ETH Zurich), César Hidalgo (MIT Media Lab), Christian Huemer (Getty Research Institute), Lev Manovich (UC San Diego), Alexander Mirsky (LMU Munich), Martin Kemp (Oxford University), Juyong Park (National University of Seoul), Diego Puppin (Google Cambridge), Martin Raspe (Bibliotheca Hertziana), Mauro Martino (Northeastern), Georg Schelbert (Bibliotheca Hertziana), and Chaoming Song (Northeastern).

Cooperation partners providing data include *Stiftung Archaeologie* and *Verlag Walther DeGruyter* in Munich and Berlin as well as the *Getty Research Institute* in Los Angeles.

2.6 Qualification of young academics related to the project (e.g. diploma, doctorates, habilitations etc.).

I was not involved in any formal Ph.D. advising within the project. However, collaborative papers in computer science and physics are often counted towards cumulative master's and Ph.D.s: Michele Coscia counts his co-authorship in publication [3] towards his Ph.D. in computer science at University of Pisa. Nicholas Blumm counts his authorship in [4] towards his Ph.D. in computer science at Northeastern University. Zalàn Forrò counts his contribution in [4] towards his master's thesis in physics at ETH Zurich.

2.7 Selected Literature:

The following selection provides an abbreviated overview over essential literature cited in the project publications [1] to [6]. Full citations in line with the standards of the choosen disciplines are given in the project publications themselves. Other sources and publications are cited directly in the text and footnotes above.

Ubiquity of Complex Networks [1] & [2]

Anderson SCIENCE 1972 *(More is different)*; **Bartsch** PEGASUS 2008 *(... Census-Datenmodell und ... Vorgänger)*; **Bertin** IDJ 2001 *(Matrix Theory of Graphics)*; **Bizer Heath Berners-Lee** JOURNAL ON SEMANTIC WEB & INFORMATION SYSTEMS 2009 *(Linked data – The story so far)*; **CENSUS** 1997–2005 & CENSUS BBAW 2006 *(The Census of Antique Works of Art and Architecture Known in the Renaissance)*; **Chen** ACM TDS 1976 *(The entity-relationship model...)*; **Henry** 2008 *(Exploring large social networks with matrix-based representations)*; **Leicht D'Souza** arXiv 2009 *(Percolation on interacting networks)*; **Nesselrath** 1993 *(Erstellung einer wissenschaftlichen Datenbank)*; **Newman** CONTEMPORARY PHYSICS 2005 *(Power laws, Pareto distributions and Zipf's law)*; **Saxl** 1957 *(Continuity and variation in the meaning of images)*; **Schich** 2009 *(Rezeption und Tradierung als komplexes Netzwerk)*; **Schich Barabási** 2009 *(... from the Renaissance to the 21st century)*; **Segaran** O'REILLY 2009 *(Connecting data)*; **Shannon** et al. GENOME RES. 2003 *(Cytoscape).*

Patterns in Complex Overlap [3]

Agrawal Imielinski Swami SIGMOD 1993 *(Mining Association Rules ...)*; **Ahn Bagrow Lehmann** NATURE 2010 *(Link Communities Reveal Multiscale Complexity...)*; **Angeles-Serrano Boguna Vespignani** PNAS 2009 *(Extracting the multiscale backbone ...)*; **Berlingerio Coscia Giannotti Monreale Pedreschi** PAKDD 2010 *(As Time Goes By ...)*; **Evans Lambiotte** PRE 2009 *(... and Overlapping Communities ...)*; **Ferlez Faloutsos Leskovec Mladenic Grobelnik** MDL ICDE 2008 *(... Network Evolution ...)*; **Fortunato** PHYSICS REPORTS 2010 *(Community detection in graphs)*; **Fortunato Barthelemy** PNAS 2006 *(Resolution limit in community detection)*; **Hipp Güntzer Nakhaeizadeh** ACM SIGKDD 2000 *(Algorithms for association rule mining)*; **Newman** PNAS 2006 *(Modularity and community structure in networks)*; **Porter Onnela Mucha** NOTICES OF THE AMS 2009 *(Communities in networks)*; **Roelleke Wang** ACM SIGIR 2008 *(TF-IDF uncovered)*; **Schich Hidalgo Lehmann Park** BARCH ON-LINE 2009 *(... Subject Co-Popularity ...)*; **Schwarz** et al., Archäologische Bibliographie. Online-Database. Munich: Biering & Brinkmann, 1956-2011 URL: http://www.dyabola.de *(Update February 2008)*; **Shannon** et al. GENOME RES. 2003 *(Cytoscape).*

Icons, Canons, and Cacophanies [4]

Batty NATURE 2006; **Bianconi Barabási** EUROPHYSICS LETTERS 2001; **Blumm** et al. PHYSICAL REVIEW LETTERS 2012 *(Dynamics of ranking processes in complex systems)*; **Bohannon** SCIENCE 2011; **Borghesi Bouchaud** QUALITY & QUANTITY 2007; **Bouchaud Mézard** PHYSICA A 2000; **Brockmann Geisel** PHYSICAL REVIEW LETTERS 2003; **Brzyski** et al. 2007 *(Partisan Canons)*; **Cancho Solé** PROCEEDINGS OF THE ROYAL SOCIETY OF LONDON B 2001; **Capocci Rao Caldarelli** EUROPHYSICS LETTERS 2008; **Cha Haddadi Benevenuto Gummadi** ICWSM 2010; **Chen Xie Maslov Redner** JOURNAL OF INFORMETRICS 2007; **Crandall** WWW 2009 *(Mapping the World's Photos)*; **Dorogovtsev Goltsev Mendes** REVIEWS OF MODERN PHYSICS 2008; **Eisler Bartos Kertész** ADVANCES IN PHYSICS 2008; **Ghoshal Barabási** NATURE COMMUNICATIONS 2011 *(Ranking stability and super-stable nodes in complex networks)*; **Golder Macy** SCIENCE 2011; **Hidalgo Blumm Barabási Christakis** PLoS COMPUTATIONAL BIOLOGY 2009; **Kemp** 2012 *(Christ to Coke. How Image Becomes Icon)*; **Mantegna Stanley** NATURE 1995; **Menezes Barabási** PHYSICAL REVIEW LETTERS 2004; **Michel Shen Aiden Veres Gray Pickett Hoiberg Clancy Norvig Orwant** et al. SCIENCE 2011; **Radicchi Fortunato Markines Vespignani** PHYSICAL REVIEW E 2009; **Salganik Dodds Watts** SCIENCE 2006; **Schich** 2009 *(Rezeption und Tradierung)*; **Sornette Deschatres Gilbert Ageon** PHYSICAL REVIEW LETTERS 2004; **Volkov Banavar Hubbell Maritan** NATURE 2003;

Arts, Humanites & Complex Networks [5]

For an overview see the table of contents in the attached eBook.

Image matrix & Community Finding [6]

Ahn Bagrow Lehmann NATURE 2010 *(Link Communities Reveal Multiscale Complexity...)*; **CENSUS** 1997–2005 & CENSUS BBAW 2006 *(The Census of Antique Works of Art and Architecture Known in the Renaissance)*; **Fortunato** PHYSICS REPORTS 2010 *(Community detection in graphs)*; **Gombrich** 1979 *(The Sense of Order)*; **Kubler** 1962 *(The Shapes of Time)*; **Lehmann Schwartz Hansen** PHYS REV E 2008 *(Biclique communities)*; **Palla Derény Farkas Vicsek** NATURE 2005 *(Uncovering ... overlapping community structure ...)*; **Porter Onnela Mucha** NOTICES OF THE AMS 2009 *(Communities in networks)*; **Russell Torralba Murphy Freeman** JOURNAL OF COMPUTER VISION 2008 *(LabelMe: a database and web-based tool for image annotation)*; **Schich** PCT 2008 *(Method for producing scalable image matrices)*; **Schich** 2009 *(Rezeption und Tradierung als Komplexes Netzwerk)*; **von Ahn Liu Blum** CHI 2006 *(Peekaboom: A Game for Locating Objects in Images)*; **Wu Huberman** ACM EC 2008 *(Popularity, novelty and attention).*

3. Final Report Abstract

During the DFG Research Fellowship "Complex Networks in Art Research – Exemplary Proofs of Concept" the applicant Dr. Maximilian Schich (*http://www.schich.info*) and his collaborators (*http://www.barabasilab.com*) were able to bring evidence that the study of complexity in the arts and humanities creates vigorous insights using large datasets, combining the qualitative with the quantitative and the general with the special. In three innovative proofs of concept arts and humanities are combined with the disciplines of information visualization, computer science, and physics, while a successful community building effort adds a another aspect.

The first proof of concept uses an information visualization approach to depict an entire art research database, the "Census of Antique Works of Art and Architecture Known in the Renaissance", in a single map, revealing a network of complex networks that spans objects, people, locations, time ranges, and events. Emerging structure becomes visible within the predefined data model, innovating database project evaluation and providing a starting point for follow up projects in a large variety of disciplines.

The second proof of concept introduces a computer science processing pipeline to create and explore a new big picture of classical archaeology, based on 50 years of "Archäologische Bibliographie". Source data is processed to refine regular subject co-occurrence and to map densly overlapping subject communities. Meso level results include algorithmically calculated cheat cheets for students, stories such as the evolution of portraits, and textbook-style class definitions. Global level structure replaces the usual featureless hairball of naive network visualizations.

The third proof of concept presents a physics theory that explains dynamics of ranking processes in complex systems. Initially inspired by observations in art research data, the collaborators identify stable systems such as the frequency of words in English books, diseases in Medicare, or marketcapitalization of companies, versus unstable systems such as Twitter hash tags, Wikipedia site visits, and scientific citations. A newly developed continuum theory not only predicts the stability of the ranking process, but shows that a noise-induced phase transition is at the heart of the observed differences in ranking regimes, connecting stable icons and canons with cacophanies of fashions and fads.

The fourth aspect of the project is a community building effort that brings together a large variety of practitioners and approaches dealing with „Arts, Humanities, and Complex Networks“. Results of this effort are an ongoing symposium series at "NetSci – The International School and Conference on Network Science", an ongoing special section in "Leonardo Journal", as well as a growing Kindle eBook at MIT-Press.

As a consequence of the whole project, the applicant was offered and accepted a position as an endowed Associate Professor of Art and Technology at the University of Texas at Dallas, starting in January 2013 (cf. *http://goo.gl/aMzeX* and *http://www.utdallas.edu/atec/about/*). In the meantime the applicant was hired by Prof. Dirk Helbing, the coordinator of the FuturICT EU Flagship proposal, to work at his Chair of Sociology (in particular modeling and simulation) at ETH Zurich (cf. *http://www.futurict.eu/* and *http://www.soms.ethz.ch*). Ongoing follow-up projects include cooperations with the Getty Research Institute, the invesitgation of large scale person data, and a wider audience book publication.