Interactive Global Histories

For a new information environment to increase the understanding of historical processes

Best Conference Paper Award - Kyoto 2013 Culture and Computing

Andrea Nanetti
School of Art, Design and Media
Nanyang Technological University
Singapore
nanetti.andrea@gmail.com

Siew Ann Cheong | Mikhail Filippov
School of Physical and Mathematical Sciences
Nanyang Technological University
Singapore
cheongsa@ntu.edu.sg | mikhfilippov@gmail.com

Abstract—The paper presents the results of an experimental case study on intercontinental trade, diplomacy, conflicts and other interactions among cities, nations and continents during Late Middle Age and Early Renaissance (1205-1533 CE). This study is based on Andrea Nanetti's ongoing research project Engineering Historical Memory (EHM) and conducted at Nanyang Technological University by Andrea Nanetti, Siew Ann Cheong, and Mikhail Filippov. The two main aims are to test: 1) ontologies to organize texts, images, and sounds in a relational database suitable to develop a systemic approach to the study of complex interactions among key subjects of the historical landscape; and 2) coherent narratives from growing historical data and metadata that can be tested at the same level of rigor as scientific hypotheses and theories. The vision is that the generation of such narratives, supported by a new coherent ontology, automatically and in a scalable way, can revolutionize the practice of historical studies.

Keywords—global histories; complex interactions; new ontologies; information environment; heritage science; storytelling

I. THE BIG PICTURE

The theoretical starting point of this frontier research is the conclusion that Umberto Eco wrote at the end of his essay From the Tree to the Labyrinth, which gives the title to his collected studies on the history of Semiotics published in 2007. Here, we look at these conclusions as a key issue for the development of new ontologies and ICT tools not only to organize data and metadata in the semantic web but also to contribute to the advancement of knowledge in heritage science [1]:

If cultures survive, it is also because they have been able to lighten themselves placing many concepts in latency, and guaranteeing its members a kind of vaccination from the Vertigo of the Labyrinth... But... facing always and again the Vertigo of the Labyrinth is often the price to pay to put into crisis the laziest ontologies of ours.

In the past, human societies embedded their knowledge in complex interactions of written, pictorial, sculptural, and architectural records, oral memories, and performed rituals. These were their media and in this form they transmitted to us their arts and science. Now, information-communication technologies open a new frontier for the advancement of data sharing. But they are not yet ready to support a new and substantial advancement of knowledge. Indeed, the exponentially growing volume of data is a solution and a problem at the same time, as Sandra Rendgen recently highlighted in an iconic way [2]:

Data are the new raw material. Today, infinite amounts of new information can be accessed in seconds and across large distances. However, raw data in themselves are of negligible value—they need to be filtered and evaluated. That's why professional data and information management will be a central cultural tool in the decades to come.

Our society is very close to a fully digital and sustainable access to all information encapsulated in monuments, museums, galleries, libraries, live performances, and archives all over the world and in any language. But to distill data into knowledge, a new generation of scholars needs to discuss, test, and implement ICT tools and solutions in correlation with the centuries-old results of each single discipline, which contributes to the domain of heritage science (in humanities, social sciences, architecture, life sciences, engineering, and computer sciences).

In a computer and information sciences perspective, at the heart of our experimental solution is a virtual interdisciplinary and interactive laboratory, where all sorts of digitized materials—from tomographic analyses of heritage buildings and artifacts to digitally recorded musical and dance performances through all types of texts and images—can be linked to each other in non-hierarchical and non-linear ways.

In this virtual place, using English as lingua franca, heritage science scholars of all countries in collaboration with ICT professionals are invited to test organizational and sharing solutions for literary, documentary, art historical, cartographical, archaeological, photographic, video, oral, and musical data sets. Computer scientists can help in Artificial Intelligence and Augmented Reality issues, by lending to heritage science their expertise on data representation, compression, manipulation, and sharing.

While historians toil to ensure that the data are properly captured, interpreted, and visualized, there is also a need for our platform to select, test, and provide the right tools for a massive deciphering effort. Information scientists can help historians to explore this myriad of growing data,
generate hypotheses on the courses of histories, test these hypotheses, and thereafter record their intellectual process in ways capable of standing up to rigorous reviews in the future.

Human interactions represent the main advantage of this information environment, applicable to distill data into knowledge in the digital era. With this vision, Andrea Nanetti leads and shares an ongoing research project called Engineering Historical Memory (EHM) to increase the understanding of historical processes through the identification of case studies and development of new ontologies able to provide scalable solutions to exploit rapidly growing data.

II. INTERACTIVE GLOBAL HISTORIES AS A SCALABLE CASE STUDY FOR ENGINEERING HISTORICAL MEMORY

Engineering Historical Memory (EHM) (http://www.engineeringhistoricalmemory.com) is an experimental methodology and an ongoing research project for the organization of historical data in the digital age, that Andrea Nanetti theorized when he was Visiting Scholar at Princeton University in 2007, to develop and test new sets of shared conceptualizations and formal specifications for content management systems in the domain of heritage science. What sets it apart from other approaches is a focus on developing and applying computationally intensive techniques (e.g., pattern recognition, data mining, machine learning algorithms derived from other disciplines, and visualization solutions) to achieve this goal. It entails the creation and advancement of databases (relational, graph, and hybrid), algorithms, computational, statistical, and complexity techniques and theories to solve formal and practical problems arising from the study, interpretation, conservation, and management of cultural heritage data.

EHM, as an experimental methodology, is based on a few simple questions: What shall the historian do having all data available in a digitalized form and available in any language? What the implications will be for studying cultural heritage when all research materials will be digitized and searchable through metadata in any language? When we know the events, what kind of interpretation can be built on top of them? Are universal interpretations possible, supported by relations between events?

The preliminary sets of formal specifications and results of tests on highly cross-linked historical data have been published by Andrea Nanetti in 2008 (local urban historical memory transmission) [3], 2010 (world views and networks) [4], and 2011 (regional man-heritage-landscape systems) [5] when he was in the Faculty of Cultural Heritage Conservation at the University of Bologna (Italy).

In a global history perspective—as Murray Gell-Mann pointed out at the international conference A Crude Look at the Whole (Singapore, Nanyang Technological University, 4-6 March 2013) in his keynote lecture on A Crude Look at the Whole: A Reflection on Complexity [6]—the vision is to increase the understanding of historical processes improving, rather than simply criticizing and marginalizing, the complex way opened by the British historian Arnold J. Toynbee between 1934 and 1961 with his 12-volume magnum opus A Study of History, in which the author presents the development of all major world civilizations [7].

The Nanyang Technological University research team of EHM was born following this inspiration under the auspices of the NTU Complexity Program Director, Jan W. Vasbinder. Its initial aim is to provide a new information environment that allows us to better appreciate famous and discussed theories like the one proposed in Guns, Germs, and Steel by Jared Diamond (whose geographical determinism does not explain for example how conquerors are conquered by the culture of the conquered, like the Romans in Greece, the Ptolemaic dynasty in Egypt, and the Mongols in China) [8]; or the one by Daron Acemoglu and James A. Robinson that does not impute to either climate, geography, or culture, but to institutions, the fall of nations [9]; or even the monumental research on the geography of human genes by Luigi Luca Cavalli-Sforza, Paolo Menozzi, and Alberto Piazza, for which Diamond writes [10]:

This remarkable book approximates a history of everything about everybody because the authors begin their accounts of each continent with a convenient summary of the continent’s geography, ecology, and environment, followed by the prehistory, history, languages, physical anthropology, and culture of its peoples.

III. 1205-1533 AS AN ADVANTAGED TIME PERIOD FOR THE SCALABILITY OF THE CASE STUDY

The starting date 1205 is emblematic for late medieval world history (Fig. 1): it was the year, just after the Fourth Crusade and the conquest of Constantinople (1201-1204), when the Venetians and the Western barons started to settle in the territories of the former Eastern Roman Empire and Genghis Khan made his first small-scale raids into Western Xia. The year 1533 is more related to the contingency of the main historical source used is this case study: the Venetian Diaries.

The Morosini codex (1205-1433) is the first successful example of Venetian historical diary dealing with the whole
known world and represents the model for the following Venetian vernacular historiography, which will lead to Girolamo Priuli's work for the years 1494-1512 and to the most famous 58-volume *Diarii* (1496-1533) by Marin Sanudo the Younger [12]. Providing information on all the empires and cities having marketplaces in the known world, it represents one of the most important international historiographical texts for late medieval European and Mediterranean history. It deals with thousands of first-hand political and economical information taken mainly from merchants' (news)letters and the official deliberations of the Venetian councils. During the second half of the nineteenth and the twentieth centuries, in the Venetian diaries many scholars have been “fishing” by excerpts those documents on which the academies of Italy, Slovenia, Hungary, Croatia, Bosnia-Herzegovina, Serbia, Montenegro, Greece, Turkey, Russia, Cyprus, Syria, Lebanon, Israel, Egypt, Libya, Tunisia, Algeria, Morocco, Spain, and France, met together studying the history of the societies that appeared at the Mediterranean and at the Black Sea window from Medieval Times onwards, through the eye of Venice. Last, but not least the Mediterranean Basin is the longest and best-studied record of the ways in which human activities have transformed the world.

The time period 1205-1533—besides its historical importance for having reshaped the world systems on land and sea both in the West (with the rise of nations, Humanism, Renaissance, geographical discoveries, etc.) and in the East (with the Mongol-led Yuan Dynasty, the Empire of the Great Ming, etc.) [13]—is significantly interesting from a data point of view as well, because there is just enough but not too much historical evidence for feasible tests of new ontologies to organize texts, images, and sounds in a relational database suitable to develop a systemic approach to the study of complex interactions between key subjects of the historical landscape.

For this same reason of providing just enough but not too much historical evidence, the choice of this time period facilitates also the scalability across the entire range of global histories in the experimental drawing of interaction layers among Europe, Africa and Asia. Indeed, from a global perspectives, the results of this case study can help investigations in other epochal topics, whether less documented, like the diffusion of technologies related to art from the Bronze to the Middle Ages (e.g. pottery kilns, ceramics/porcelain, silk, glass, paper, lost-wax casting bronze productions, burial practices, etc.), or overwhelmed by documentation like the European expansion on the oceans in Modern Times. In this way, this case study creates a new paradigm to bypass the current theoretical impasses of global heritage/art histories as well, focusing on artifacts.

### IV. DATASET AND ONTOLOGY OF THE CASE STUDY

Andrea Nanetti between 2010 and 2012 digitized the 1400-1423 portion of the Morosini codex and created a relational database as co-PI of a strategic research project directed by Gherardo Ortalli at the University of Venice Ca' Foscari. The database provides both data (text-image links for the Morosini codex) and metadata for a first set of primary contents (linked geographical locations, governmental entities, and attributions), with a customized Web interface for data input and retrieval. Since then, a college of experts (Andrea Nanetti, John Melville-Jones, Alan Stahl, Sergej P. Karpov, Giovanni Caniato, and Andrea Rizzi) is responsible for the translation into English and the constant update of the critical edition in the database.

In 2010 and 2011, around the idea of the database took place a series of seminars led by Andrea Nanetti at the European Center for Living Technology of the same University of Venice, using the ideas about organization thinking developed in the first chapter of the ISCOM book by David Lane (one of the dissemination activities of the EU project ISCOM ‘Information Society as a Complex System’, 2002-2006) [14] to develop models of 'kinds of world', which could be used to distinguish Morosini's from 'earlier' (or just 'different') kinds. That is a way of doing inference on the kind of material that the Morosini's diary represents, making rigorous what historians often treat in a completely anecdotal and suggestive way, without using "all" the material available, but only selecting fragments that illustrate their particular interpretations. We realize that all this probably sounds very mysterious—in part because there would be a lot of thinking to do before being very clear on the implications of the research started in Venice, in part because to write what have been thought in Venice would be a lot of thinking to do before being very clear on the implications of the research started in Venice, in part because to write what have been thought in Venice would take a whole lot more pages than probably it warrants in this paper. And, above all, in the theoretical field there is no need to showcase any new work of genius.

The main benefits and improvements that the computer can provide to historical research are not new to scholars. In the 1970s, the considerations made by the French historian Emmanuel Le Roy Ladurie, on which implications the use of the computer would have in historical studies, made already aware that [15]:

> History based on computers/information technology is not limited to a very specific category of research, but also leads to the establishment of an 'archive'. Once transferred to tape or punched cards, and after having been used by a first historian, the data can in fact be stored for future researchers, who want to find non-experimented correlations... What emerges is a new kind of archivist; a sort of engineer of history, very different from the great scholars trained at the *École Nationale des Chartes*... At the end, the duty of the historian will almost uniquely consist in thinking: which, in fact, is supposed to be his unique vocation.

And even the idea of looking at unexpected relations in the repertory of knowledge is not new, when we theorize that research is not only finalized to use what we find for what it is, but also to find new possibilities of interactions among elements that initially did not have any reciprocal relationship. This concept dates back at least to Roger Bacon (ca. 1214-1294), who switched the concept of *inventio* (invention/discovery) from the search of what was already known to the discovery of the unknown [16].

The real frontier research starts when we want to make things work in the digital era and engage heritage science people in teamwork. To do this we need to define a research domain, to propose it as a case study, and to share an explicit
formal specification of the domain terms and their reciprocal relationships. In one word, we need a new ontology, according to the definition given by Thomas R. Gruber [17]. The relationships (trade, diplomacy, conflicts, etc.) among governmental institutions (agents) in the intercontinental network (the system) constitute our first ontology for this case study of interactive global histories from 1205 to 1533 as a scalable solution to investigate change (i.e. the fundamental and nonlinear force of history).

V. AN EXPERIMENTAL DATABASE FOR AUTOMATIC NARRATIVES GENERATION

In the words of John H. Holland [18] and William Brian Arthur [19], complex systems are characterized by their dependence on contingencies, for example, event B happens because event A happened in the past, but not event C. These contingencies, also called path dependences, make the study of global histories highly bewildering, because of the concatenation of conditional probabilities [20].

Seeing history not as a linear progression of events, but as a complex, nonlinear network of contingencies gives us the correct frame of mind to respond to the issues raised by William A. Green in his article on world history periodization [21]:

Periodization is rooted in historical theory. It reflects our priorities, our values, and our understanding of the forces of continuity and change. Yet periodization is also subject to practical constraints. For pedagogical reasons, world historians must seek reasonable symmetry between major historical eras despite huge discrepancies in the availability of historical data for separate time periods and for different areas of the world.

Our proposed solution to delimit time periods is to move away from focusing on main individual events, but to look instead at intensity in the flow of events in societies’ natural nonlinear perception of time.

Marten Schefter’s research team recently used ecosystems as a showcase to point out that complex systems theory associates regime shifts as critical transitions with higher intensities of events [22]. Building on his insights, we decided to mine the complex network of intercontinental trade, diplomacy, conflicts and other interactions among cities, nations and continents during Late Middle Age and Early Renaissance (1205-1533 CE) to identify time and geography of such transitions.

To understand how narratives can be generated automatically as a key tool of the database, we turn to the basics of storytelling, which follows the pattern of ‘who’, ‘what’, ‘when’, ‘where’, ‘why’, and ‘how’. In Fig. 2(left), we show how these storytelling elements can guide our analysis of historical data. In essence, for a computer program parsing historical data to automatically generate a narrative, it must identify the key actors (‘who’), key events (‘what’), key periods (‘when’), key locations (‘where’), key factors (‘why’), and key actions (‘how’). This can be done partially through the construction and analysis of the kind of complex network shown in Fig. 2(right).

To construct the complex network shown in Fig. 2(right), we first compile a list of actors and a list of relations. Depending on what scale we are interested in, the actors may be governments or individuals (or a mixture of both). We will then select relations that are appropriate for the scale of the actors. We then parse the database of historical records to find instances where two or more actors are mentioned, and map the reasons they are simultaneously mentioned to one or more of the relations we have in our list. A complex network constructed this way is time-integrated, because all relations between all actors at all times are incorporated. In such a network, the different relations between two actors at different times will be displayed all at once. This serves as a good benchmark for the time-resolved analysis that we describe below.

But before we perform the time-resolved complex network analysis, we aim to simplify the time-integrated complex network. First, the relations we listed are manually selected. To avoid missing important historical developments, we will start with a larger list of relations than necessary (i.e. some relations, though they go by different names, are actually slightly different manifestations of the same political interaction). We will therefore develop a new network-based clustering method to identify the fewer most relevant interactions. Second, actors may organize themselves into high-stable power blocs. These arrangements limit the type of interactions actors can have within a power bloc, and also the type of interactions they can have between power blocs. We again intend to employ network-based clustering to discover these power blocs.

Using the simplified time-integrated complex network after relation and actor clustering as a template, we then perform time-resolved analysis by examining the sequential activation of links, as shown in Fig. 3. This can be made into a movie with geographically situated actors on an actual (or historical) map, for play back on our web interface. Beyond aesthetic values, such a movie can help historians grasp the large-scale geopolitical changes that took place within a selected window of time. More importantly, this mode of historical analysis allows us to quantitatively identify key
periods in history, as quick successions of inter-related events, as shown in Fig. 3. These key periods are normally easy for historians to identify, but our method makes this identification process automatic.

Fig. 3. For a given time window size (1-year, 5-year, 10-year, etc), we examine the links (shown in bold) that are 'simultaneously' activated, and also the activation sequence (shown as the transitions from $t$ to $t+1$ to $t+2$).

The next parts of the narrative we can extract from the time-resolved complex network analysis are the key events. As shown in Fig. 4, we find the activation of a group of red links converging on $h$ at time $t$, followed by the activation of a single blue link between $h$ and $b$ at time $t+1$, followed by the activation of a group of green links diverging from $b$ at time $t+2$. The event associated with the activation of the single blue link can be thought of as a key event or a tipping point, because had it not occurred, the converging group of events and the diverging group of events would never be causally connected.

Fig. 4. In this figure, there is a sudden proliferation of inter-related events at $t+1$, relative to $t$ and $t+2$. We can therefore identify $t+1$ as a key period in history.

Finally, our time-resolved complex-network approach to historical analysis allows us to automatically identify the key actors driving a specific key event. At present, expert historians painstakingly piece together the key events and thereby examine the main actors in such key events. With our method, this identification can be automated. More importantly, we can adopt different definitions of key actors. We can, for example, define a key actor to be one participating in many events within a key period. Alternatively, we can define a key actor to be one participating in a key event. The two definitions are clearly different, but they can yield the same key actors, or different key actors.
Fig. 5. In this figure, we show the activation of red links converging on \( h \) at time \( t \), followed by the activation of a single blue link between \( h \) and \( b \) at time \( t + 1 \), followed by the activation of green links diverging from \( b \) at time \( t + 2 \). We can identify the event associated with the activation of the single blue link as a key event in this historical period.

Coming back to the automatic narrative generation shown in Fig. 2 (left), we see that our complex-network method will automatically discover the ‘who’, the ‘what’, and the ‘when’. Since actors and events are always geographically situated, we expect knowing the key actors and key events will automatically point us to the key locations (the ‘where’). Therefore, we are left with the ‘why’ and the ‘how’ to complete the story. We believe that once partial narratives can be generated in the way we described above, historians will have a much easier time figuring out the two remaining elements of the complete narrative. In fact, we envisage that our complex-network method can be extended in future work to also automatically generate candidates for the ‘why’ and the ‘how’, by going beyond actors to generate a complex network of key events, that is a coarse-grained version of the complex network of events that is dual to the complex network of actors (cf. Fig. 5).

VI. VISUAL INTERACTIVITY AS A KEY INQUIRY TOOL

To engage the scholars on the Web and apply crowdsourcing principles the key factor is interactivity, possibly visual. That is why the project aims to be a user-friendly visual framework, where scholars as authors/editors are able to upload and organize their historical data, extract relevant information and make inferences from the data uploaded by their colleagues as well. If the general spirit is shared with Wikipedia, Wikimapia (where users and guests mapped over 20'000'000 objects on an interactive global map), and other collaborative databases, what makes the difference here is the application of the highest standards of scholarly publication and ethical issues in each different typology of historical source.

As a very basic example, the above mentioned Venetian diaries need to be fully indexed, parsed, and translated into English by a college of experts to become highly attractive and reliable for the historians’ international community as a whole. But the scholars have to be able to interact with the critical edition, translation, and data parsing as authors/users in an ongoing flow of addenda & corrigenda to ensure increasing collaboration to the project (e.g. the review by Sergej P. Karpov of the Morosini codex edition [23]).

The project is developing and testing tools (technologies and processes) to be readily adopted by users to visualize high volumes of data through maps, timelines, tag clouds, and/or interconnected graphs on different scales. Because, not only highly qualified scholars, but also artists, students, and all sorts of other users will create and share their narratives, by tagging, connecting and recognizing links among elements of the historical landscape.

The tag cloud can act as a link between relational and graph database solutions, suggesting case-by-case tests according to the primary differences between the two kinds of databases. Relational database structure allows keeping regular, relatively simple, structure. Graph databases, in turn, provide visualization and simplify information extraction through maps, timelines, etc. In a graph database, the relationships are stored at the individual record level, while in a relational database the structure is defined at a higher level (e.g. the table definitions).

A relational database is much faster when operating on huge numbers of records and using minimum storage space for it. At the same time, in a graph database, each record is examined individually during a query in order to determine the structure of the data, what make sense when there are a lot of variations in the relationships between records. According to these characteristics either structure could be used to provide necessary service: to store heterogeneous data types, as a platform for crowd sourcing, to mine data and metadata for coherent narratives.

Historian's theories select events and explain the processes that give to those events a meaning. Here historians can develop and test their theories of change, that are the main instruments by which they provide explanations to how and why history has evolved; being the highest
scientific and theoretical challenge of the project at this stage to offer an automatic narrative generation engine as part of the suite of tools on Interactive Global Histories.

VII. CONCLUSIONS

L’aiauola che ci fa tanto feroci
(Dante Alighieri, Paradise XXII, 151)

Such a case study with good coverage of various historical periods and geopolitical regions is able to generate from noisy and incomplete sets of historical data and metadata coherent narratives that can be tested at the same level of rigor as scientific hypotheses and theories, as more data and metadata become available. The generation of such narratives automatically and in a scalable way will revolutionize the practice of history as a discipline. An interconnected world can be investigated to understand the mechanisms of convergence/divergence between local communities and international networks, placing a crucial question: how the same networks/people can bring new wealth and development, or war and poverty? Which are the dynamics of sustainability in international mechanisms?

Last but not least, even if this historical database is mainly a platform for scholarly inquiry, being an immersive visual research experience it can also serve as a repository of highly qualified materials for entertainment and edutainment industries as it is designed to encourage and support new endeavors in media innovation and develop scalable solutions that can be readily adopted by information technology companies as well.

ACKNOWLEDGMENT

The master plan designing, Web strategies, commercial and industrial solutions of EHM are developed and administrated by Meduproject Pte Ltd. (a company established in 2002 by Andrea Nanetti as academic spin-off of the Department of History and Methods for Cultural Heritage Conservation of the University of Bologna, after having been awarded in 2001 a prize in the first Italian business plan competition devoted to projects with high content of knowledge and having been financially supported by the Italian National Agency for New Technologies, Energy and Environment) with research teams and fundraising in Venice, People's Republic of China, Australia, USA, Greece, Turkey, Russia, Egypt, India, and Singapore.

REFERENCES


[16] U. Eco, op. cit., p. 44.


