

Part B2: The scientific proposal (max. 15 pages)**Section a. State-of-the-art and objectives****a1. Project overview and objectives**

This project proposes a step-change in the study of modern China reliant upon scalable data-rich history. It will deliver precise historical information at an unprecedented scale from heretofore untapped sources - as well as reshaping the analysis of existing sources - to create a new dimension in the study of the transformation of elites in modern urban China. It will deploy an array of cutting-edge digital methods with this information — including data mining, sampling, and analysis within an integrated virtual research environment. To establish the validity of this approach in historical research, this project focuses on elites in the three main and interrelated urban centers of China (Shanghai, Beijing/Tianjin, and Canton/Hong Kong). The key issue that the project will address is breaking through existing limits of access to historical information that is embedded in complex sources and its transformation into refined, re-usable and sustainable data for contemporary and future study of modern China.

Key objectives of the project include:

(1) **Analyzing urban elites in modern China at the level of actors rather than state institutions or community organizations** to determine who were the individuals and groups that emerged as *elite* in the selected cities and the modalities and processes of elite formation and (re)deployment of elite networks within and beyond China.

(2) **Analyzing the vectors, patterns and timelines of the involvement of elites in public action**, from acting in an official capacity (administrative and military elites), in self-organized associations (native-place associations, merchant organizations, religious groups) but also assuming the role of opinion leaders, directly (press, publishers) or indirectly (e.g. cinema, advertising) in the context of the different selected urban areas.

(3) **Investigating the transnationalization of urban elites as a process** that shaped the social profile, scope of actions and expertise of elites in a way that overstepped the limits of institutions and nation states.

(4) **Demonstrating how the radical change of scale and quality of historical information** that the combination of selected advanced digital methodologies — data extraction from print sources, data mining and machine learning, social network analysis, historical geographic information systems — can lead to data-rich history based on large-scale empirical data.

(5) **Establishing new digital resources for extended research by the broader community**, together with a virtual research environment (VRE) based on standardized workflows, replicable analysis procedures for large-scale textual corpora using capture of intermediate data-sets and user-friendly adaptable interface of advanced data mining tools.

In the century that followed China's forced entry into the world order shaped by the advancing Western empires (1832-1949), the nature of elite groups in China experienced a tremendous and radical transformation. The ruling group that had dominated the country for a millennium — a narrow elite of Confucian-trained civil and military scholar-officials selected through a double process of education in the Chinese classics and the highly competitive imperial examinations (Elman 2013, De Weerd 2007 & 2016) — was finally swept away in 1911. Power and social prestige shifted to socially, intellectually and ethnically more diversified groups that included not just Chinese, but foreigners who operated within interlocked transnational networks. **Within and beyond China, this period saw the rise of the first global elite that set the pace of multiple entangled histories.**

The displacement of the ruling elite of scholar-officials did not happen in a single strike. The abdication of the Qing emperor and the collapse of the imperial order in 1911 marked the end of a tumultuous historical process. On the one hand, China faced increasing internal tensions between the growing size of her population, the trap and constraints on resources of her pre-modern economy, and a lesser capacity by her overburdened bureaucracy to maintain social stability. This resulted in popular rebellions and eventually a devastating and traumatic civil war (Taiping Rebellion) that caused a tremendous sacrifice of lives, a desolation of the economy in the wealthiest area of the country, and a momentous loss of confidence among the elites in the dynasty's capacity to restore the political and social order, and to meet the challenge of Western intrusion. The threat — and the search for a solution — was perceived as primarily military. Yet the fundamental consequence of the admission of Westerners and later Japanese into China lay elsewhere. It was in the flow of people, knowledge, technologies, institutional forms, etc. that challenged the position and the *raison d'être* of the Confucian-trained elite. Incrementally, but irremediably it became irrelevant. The abolition of the imperial examinations in 1905

marked the acknowledgement that the source, structure, and function of the elites in China had definitely changed.

The scope of the project includes the three urban areas (Shanghai, Beijing/Tianjin, Canton/Hong Kong) that had the most profound impact on the course of modern Chinese history. They were the major engines of change in China, they had the highest number of educational institutions, especially foreign-run schools and universities, the largest business communities, the most diverse groups of administrators, the most intense pace of innovation in all fields. They were the space where Chinese and foreigners came into real contact, learned from each other, worked together, and sometimes fought each other. They provided the setting as much as they were the vector of production of new transnational elite groups. The China-centered and group-based approach dominant in the historical literature of the past two decades has *de facto* drawn all sorts of dividing lines among elite groups and instituted an artificial dichotomy between Chinese and foreigners at the level of elites. We propose a radically new approach.

The selected cities followed very **different historical trajectories**, especially in terms of local power and relation with the central state. Canton, the empire's commercial doorstep, was displaced by Shanghai and Hong Kong. It remained very much under the control of a succession of local autonomous power holders during most of the Republican period. Beijing, the capital city for centuries, was stripped of its prestige and source of wealth with the collapse of Qing rule in 1911 and the relocation of the capital to Nanjing in 1928. Thereafter, it was the seat of a string of short-lived military dominated governments before and after 1928. Shanghai was a tri-partite city placed under two fully autonomous foreign administrations and a succession of diffused local Chinese administrations until 1927, then under a single unified modern municipal administration. Yet it was not really unified under a Chinese administration before 1945. Tianjin had a similar though more complex setup due to the concomitant existence of up to nine foreign settlements. Beijing, Tianjin and Canton experienced short periods of full Western occupation and management. Finally, Hong Kong emerged and developed strictly under British colonial rule. All these places experienced the traumatic fate of the Japanese occupation.

We make the hypothesis that China represents a case that does not fit in the usual paradigms through which historians have studied the relationship between the dominant Western powers or Japan and more broadly the colonial world. Apart from the fact that China maintained her sovereignty, despite the indentations of the treaties, the most significant factor was the existence and persistence of a centuries-old elaborate pre-modern elite. **The project envisions elites in urban China as actors** —individual and collective actors — whose status, position, and practices were shaped by the power configurations that developed over time and whose actions through institutions and informal/formal networks in turn were a determining factor in redrawing social and political boundaries. **The project will address these key issues:** (1) If we observe elite groups at the level of actors rather than state institutions or community organizations, who were the individuals and groups that emerged as “elite”, what were the modalities and processes at work in elite formation and what shaped elite networks? (2) How did the vectors and patterns of the involvement of elites in public action, which ranged from acting in an official capacity, in self-organized associations, but also assuming the role of opinion leaders, directly or indirectly vary in the selected urban centers? (3) What was the impact of economic globalization on the role the economic elites and elite networks played in the transformation of the selected Chinese cities and in the emergence of public spheres at a time when elite groups became much more complex and diversified; (4) Elite groups were at the source of disputes that sometimes led to outright conflicts between competing groups or between state and society. How did the various competing elite groups face, mobilize, manage and interact during such events of short duration or during the major socio-political upheavals (e.g. 1911 revolution, May Fourth Movement, Sino-Japanese conflict, etc.) that gripped the whole country?

The present project emphasizes **feasibility**. It delimits very precisely the corpus of press, directory, and biography-related materials in three core urban areas and combines the proposed **macroscopic approach to selected case studies**. This two-pronged approach is meant to maintain a solid connection between both approaches and to address head on the crucial issue of scalability in data-rich history (Bode 2017).

The **groundbreaking nature** of the proposed project can be defined as follows: **Substantially, it will redefine previous research** on the Chinese elites based on narrow samples by introducing a groundbreaking **change of scale** in the historical information for the study of elite formation, elite networks and the role of elites in the broad realm of political construction, economic exchange, and social control in the three main and interrelated urban centers of China. It will create the conditions for a more sophisticated and refined interpretation of the process of transformation of urban elites in modern China based on rich empirical data. **Methodologically**, it takes up the challenge of **solving three crucial bottlenecks** in digital methods for historical research: (1) automatic retrieval of the historical information embedded in print materials and their transformation into structured data (WP1-1); (2) retrieval of historical information in vast digital corpora

beyond data mining through text analysis and machine learning (WP1-2); (3) effective processing of the vast amount of collected data into scalable structured data sets (WP1-3).

Our approach contrasts with existing digital workflows in the humanities which, although they increasingly produce digital collections employing standards-based core metadata representations, have been unable to capture research activity that can be registered as annotation metadata in a sustainable way; nor have there been means to make traceable and preserve the application of virtual instruments in terms of datasets prior to interpretation for publication. Our **implementation strategy in this project** - which benefits from recent but significant developments such as IIF and Mirador - will be to establish new digital research collections comprising extracted texts and actors linked to annotated source pages, as well as acquiring intermediate datasets and results from virtual research tools applied to these collections. These research collections will use *freizo* and the Open Annotation Data Model (OADM) to improve portability and reduce dependencies on specific technologies - making them economically sustainable. Significantly, they will also be exported by *freizo* to the Invenio library management system pioneered by CERN, which has now been adopted by major libraries and which supports MARC21 connection with the wider research community. The research collections will enhance communication between internationally distributed partners and provide long-term protection of research investment, as well as establishing a sustainable digital resource for dissemination and future researchers.

The project will also **enable the PI to build on a large amount of existing work** conducted with a wide range of methods (from archival work to data mining) on large corpora of sources of a very different nature (archives, press, etc.) that resulted in a **significant amount of pilot and advanced projects, validated methods and data sets**. This existing investment will be converted for compatibility and inter-operability with the new research collections. More specifically, the project will break down into five major interrelated work packages (WP1/1-3, WP2/1-2) and one work package devoted to meetings, training, and publicizing, as discussed below. These WPs will be further divided into 33 tasks (Tsk).

a2. State of the art

This section presents a brief review of studies on urban history in modern China and more specifically on elites and elite organizations in historical research, as well as a discussion of works and methods in social network analysis and historical databases in relation with the study of Chinese society.

The study of elites by historians represents a rich historiography, but it does not in itself constitute a specific field of history. The term elite(s) has most often been used loosely about socially or politically leading groups to address issues of modernization or state-society relations (Coste 2014, Korom 2015). Conversely, there is a well-established field of “elite theory” in sociology and political science that provides conceptual frameworks and methodologies to define and study elites and power in contemporary societies (Hartmann 2007, Domhoff 1967, Higley 2006, Lachmann 2010, Hertz & Imber 1995). These works point to two intertwined dimensions of power: the capacity of a group, class, or community to combine organizational forms, resources and social morale to achieve its goals; yet also the ability of any group, class, or community to be successful in competition or conflicts with other groups, classes or communities. **How do we define elites in this project?** The study of elites conventionally conveys the notion of “dominance” and “power”. This is also the major streak in the use of the term “elites” in the historiography of modern China. This project adopts the notion of power as “the ability to achieve desired social outcomes” (Domhoff, 2012, 23) and of elites as “those individuals, from whatever class, who held leadership posts within civil society across a range of political and voluntary activities” (Trainor 1993).

The project approaches the study of elites not just as specific groups of individuals framed in distinct organizations or occupations, but as individuals embedded in multiple and evolving networks, of which formal organizations represented one important but not unique facet. The choice not to start from a hard (but elusive) definition or from pre-established groups constitutes a risk, but it also opens up new perspectives beyond the prosopographic approach, especially in view of the “multiple ecologies” within which elite groups evolved (Abbott 2005, 2016). It is a feasible and realistic approach with the methods and the change of scale in the production and processing of historical information discussed in full in Section B (Methodology).

There is no identifiable field of elite studies in the history of modern China. The seminal works by Chang Chung-li (1955) or Ho Ping-ti (1962) on the literati elites or those of Bergère (1967) and Elvin (1967) on the “gentry/bourgeoisie”, or Bergère’s use of prosopography (Bergère 1986) inspired very few followers. Yet the term “elite” has gained increased currency among China historians. It has been applied to specific *Chinese* groups (merchants, literati, intellectuals) and used more “in passing” than as a genuinely operative concept. In the study of modern Chinese cities, the focus of historians has moved from broad social constructs (bourgeoisie, workers) to social categories and institutions based on the actual structure of Chinese society (Review essays: Henriot 1996; Stapleton & Liu 2006; He 2014). Within this historiographical trend, the

attention to elites has grown out of a general reflection on the transformation of society at the local level, mostly at the provincial level (Chang 1978, Schoppa 1982, Rankin 1986, Esherick 1990, Xiao-Planes 2001), the decline of the imperial state and the advent of the 1911 revolution (Esherick 1976, Rhoads 1975), and the much-debated issue of civil society/public sphere (*Modern China* 1993).

A significant number of major works has centered on the role of social elites, especially the economic or business elites, (Bergère 1986, Kwan 2001, Xu 2001, Sheehan 2003, Gabbiani 2011, Han 2016), on their organizations (Fewsmith 1985, Rowe 1984 & 1989, Goodman, 1995; Belsky 2005) or cultural practices (Wu Jen-shu 2005, 2007; Paulès 2010). Even when the term appears in the title (Goodman 1995, Brook 2000, Brook 2007), elites was not used as a central heuristic concept. In the literature on Chinese cities since the 1980s, the China-centered perspective has also led to a historiographical bifurcation, with the bulk of studies devoted to various groups in Chinese society and only a short string of works on individual foreign communities and their organizations (Bickers 1999, Ristaino 2001, Gao 2013). The explicit use of network can be found only in a number of works by historians and sociologists on Chinese merchants, mostly Cantonese, in South China and in Southeast Asia (Lee 1991, Lai 1994, Choi 1998, Buiskool 2009, Lim 2010 Tagliacozzo 2011). They examine whole communities (e.g. Chaozhou merchants) and do not specifically study elite groups. None of these studies is based on the systematic collection of data and the use of network analysis.

In the study of elites in Europe (see selected [bibliography](#)) — country/nationwide studies, studies in a specific locale (most often cities), and professional groups — few works have adopted a specific methodological approach such as prosopography or network analysis (Charles 1992). Conversely, there is a vast body of works based on prosopography (especially in ancient and medieval history, such as [Prosopography of the Byzantine Empire](#) or [Prosopography of Anglo-Saxon England](#), designed usually as the study of specific groups, both elite and non-elites, selected on the basis of their profession or position (Stone 1971, Carney 1973, Charle 1980). The field of global history has also produced a rich literature and debates on epistemic communities, expats, transnational professionals, international civil servants and of course, transcultural exchange processes and the development of internationalism (Herren & Löhr 2014, Herren 2014; Hoerder 2005, 2011, Sachsenmaier 2011). In the perspective of this project, transboundary actors and networks are of increasing importance and especially true for the urban context, addressed as transnational meeting points for diaspora communities (Goebel 2017, Conrad 2017).

In the field of modern Chinese history, previous attempts at building biographical databases (*Chinese Biographical Database*) have died of technological obsolescence. The [China Biographical Database](#) (CBDB) project at Harvard University is probably the closest to this project in terms of data, yet with many caveats. Its content (360,000 individuals) covers primarily early imperial China (7th-19th century), mostly from a limited corpus of official documents processed manually. Its major drawback is the rigid proprietary software application on which it is based (Microsoft Access) and the use of pre-defined categories in designing the database. Whereas the range of collected data in prosopography was previously limited in scope and depth, modern computing and databases have lifted such constraints (Cornwell et al. 2015, Dedieu 2005, 2013; Bradley & Short 2005, see also scholarly networks¹). Furthermore, conventional prosopography can now harness the potential of Social Network Analysis (SNA) (Brioiest 2006, Verbruggen 2007). In the early 1990s John F. Padgett's work (Padgett 1993, Wetherell 1998) introduced SNA in historical research, but a new awareness has emerged only recently due to the impact of the Internet and the rise of social networks apparatus (Facebook, Twitter, etc.) in contemporary societies (Brughmans 2016). While SNA has been taking a new hold in historical studies, requiring careful consideration as Claire Lemerrier argues forcefully (Lemerrier 2004, 2005, 2010a, 2010b, 2015), only rare works have implemented SNA based on the data mining of historical records (Camp 2016).

In summary, current historiography on modern Chinese urban society is characterized by a marked historiographical imbalance that excludes for the most part foreigners while the China-centered approach has prevented historians from examining one of the central features of social development in the main Chinese cities, namely **the rise of a global elite**. The most striking result is the juxtaposition of group- or community-focused studies, with hardly any attempt to examine whether, how and to what extent contacts and connections emerged, developed, and changed between the various components of the elites and how this shaped power in these cities and in China. The project argues for the need for a new approach. First, it is now possible **to dig in a wide array of sources heretofore beyond the reach of the human mind** thanks to the massive digitalization of core collections of historical documents (e.g. newspapers, local gazetteers, biographical dictionaries, etc.) in Chinese, Japanese, English and French and the use of cutting-edge digital methods. Second, by **embracing the long century** that saw the collapse of two political regimes and periods of massive

¹ For France, see [Réseaux et Histoire](#) and [Modélisation des réseaux en histoire](#), at the European level see [Historical Network Research](#) and [The Connected Past](#), in the United States, see in particular [Visualizing Historical Networks](#).

upheavals and wars, the project will investigate individual trajectories in conjunction with the dynamics of elite networks and their structures. Third, the project seeks to **examine elite networks in their various configurations and articulations** as they emerged and operated within and beyond the selected cities, China, the Western and Japanese empires, and the power nations (Great Britain, France, United States, Japan) themselves.

a3. Output objectives

The main outputs will unfold in three formats: i. Research articles in peer-reviewed journals, research monographs, and dissertations; ii. Digital scholarship in the form of both scholarly presentations and research collections making a very large repository of informed and validated historical data on urban elites available to the research community using acknowledged standards (including OADM, IIF and MARC21); iii. An integrated workflow of tools for automatically parsing comprehensive OCR outputs, extraction and mining data, annotating source materials, refining and correcting data-mined data, and acquiring and preserving intermediate data-sets from third-party tools (GIS, SNA, etc.). This will enhance communication between internationally distributed partners and provide a sustainable digital resource for future researchers, as well as graduate training in historical digital methods. The project will provide *in real time* to the scholarly community and to the general public:

- A quantitative and qualitative analysis of the dynamics of elite networks based on micro-data for individual members of the political, economic and social elite, and economic or political organizations. **This would be the first quantitative analysis of elite networks based on extensive historical data in urban China for the period 1830-1949.**

- A comprehensive historical database with longitudinal individual data covering elite networks for at least three generations (from ca. 1830 to 1949). The database will be constructed with data collected from multilingual information in European and Asian languages covering a wide range of self-organizing networks of the Chinese elites, along and in interaction with non-Chinese networks, mostly Westerners and Japanese.

- Conception and development of an innovative machine learning process in data mining of unknown entities through massive digital corpora, with user interface to meet the needs of non-programming humanists (in history and beyond history).

The present project will offer a renewed interpretation of the historical experience and complex process of elite formation in modern China when the country experienced a momentum of drastic and rapid change. The Chinese historical experience has a direct bearing on our understanding of modern elites in general. It is not indifferent at a time when the formidable transformation of the PRC in the last fifteen years has witnessed the emergence of a new global elite. **Methodologically, the project will create the conditions for the implementation of a virtual research environment beyond the fields of history.** It will elaborate tools and methods applicable to other corpora and usable in various disciplines of the humanities and social sciences.

Section b. Methodology

b1. Description and justification of methods

Our over-arching question in this project is simply **what can digital methods and their combination offer to historical research?** To answer this to ourselves and a broad audience of historians the first step is to imagine and develop the methodologies to create, refine, and analyze large historical data sets. The second step is to apply these methodologies to specific corpora and to concrete historical objects to understand how these can be implemented by describing and explaining the tools that are essential to answer the research questions we are interested in. This underpins the logic for having two major applied projects. **Data-rich history implies producing and using large-scale historical data, namely the “big data”** of historians, at an unprecedented scale by using the cutting-edge methods discussed below. To achieve this objective, the project will implement three fundamental sets of digital methods: data mining and text analysis, Social Network Analysis (SNA), and Geographical Information System (GIS) which we discuss fully in WP1-1 & 1-2 and WP2-1.

To collect historical information from the selected vast source corpora, the project will implement **data mining and text analysis**. They include an array of methodologies that address through different approaches the collection of information from structured and unstructured texts. They are based on algorithms that range from searching and matching pre-identified terms from a dictionary to their equivalent in pre-determined corpora (Gregory 2012-16, Communication and Empire 2017), to analyzing the content of texts based on natural language processing (word frequency, collocation, etc. Unsworth 2011, Piotrowski 2012), to procedures of machine learning based on semantic analysis and annotating texts to determine patterns that serve in turn to generate increasingly accurate and focused detection of events in large corpora (Bellot 2011,

Hakkani-Tür 2011, Prévôt 2009). These methodologies are applicable to modern print sources, digitalized historical sources and scholarship, as well digital-born materials such as historical scholarship of the last two decades. These methodologies are particularly suitable to explore and process the vast digital corpora that constitute the documentary basis of the project.

In analyzing historical data, the project will enable two major methodologies especially well-suited for our approach of elite networks, **Social Network Analysis (SNA)** and **Geographical Information Systems (GIS)**. SNA is a tool for the exploration and analysis of complex data that sits at the intersection of network science and data visualization. One of the main heuristic values of visualization is to highlight what the naked eye does not perceive, even through statistical analysis, and suggest unexpected research hypothesis. SNA consists of four basic assumptions that together give coherence to a larger approach. First, actors in all social systems are viewed as “dependent rather than independent”, thus any modification of an actor/part of the network may affect the whole system. Second, the relations among actors channel information and other resources. Third, the structure of those relations among actors both constrain and facilitate action. Finally, the patterns of relations among actors define economic, political and social structure (Wetherell 1998). SNA shifts the gaze from the individual, group or organization to the ties or linkages — and their transformation over time — between two or more persons, groups and institutions (Wellman & Berkowitz 1988). It can highlight nexus of relationships, hence structuring patterns that the actors involved in them may not have been aware of and that the historian himself/herself might miss entirely.

The project does not adopt SNA as a way to “reveal networks”, but as **a powerful tool to explore complex data to probe into underlying social mechanisms and raise new questions**. It takes the existence of networks as a working hypothesis to delineate their contours and components from empirical data and to understand their formation and their specific role among social and political elites in modern Chinese cities. This project will precisely build on the massive amount of collected historical data to break through the limitations of quantity and quality of historical data that until now constituted a major impediment in using SNA in history.

A **GIS** is a type of database in which every item of data is linked to a location. It enables a wide range of possibilities such as integrating a wide range of attributes based on their location, mapping and visualizing the data, allowing space-based queries and analysis (Gregory 2003). GIS represents **one of the most innovative digital methods that has reinvigorated research methods in history** (Knowles 1999, 2004, 2008) and more broadly in the humanities (Gregory 2013, Bodenhammer 2010). In the project, in addition to building a gazetteer based on extracted place names in their spelling variations and connections, GIS provides the most suitable technique **to place actors and networks in space, to examine movement, circulations, and migrations**, and to represent topical data associated to place names extracted from the corpora processed in WP1-1 & 1-2. Spatial networks, although they study formal physical networks, provide robust methodologies on networks and complex data that the project can adapt and transpose (Oliver 2016, Yang & Shekhar 2017).

b2. Implementation of methods and work packages

WP1 Methodologies: From texts to historical "big data"

This WP defines ambitious but concrete new approaches to i) **automated extraction** of high quality research data from both unstructured and semi-structured sources—those already digitized as well as from challenging publications on paper previously addressed only through manual transcription; ii) **enhanced formats of data mining** based on computational linguistics and innovative application of machine learning algorithms; iii) **long-term sustainability** of materials in digital form, and of intermediate results arising within research procedures (in contrast and in addition to published results).

A key dimension in this project is to collect precise historical information on the individuals, corporations and agencies who in the course of their lives or activities, emerged as significant actors and joined/formed elite groups, as well as the institutions (companies, associations, clubs, etc.) to which persons belonged and the events in which they participated. The **first primary block of documentation** (WP1-1) consists in the numerous annual directories — concretely the databases of the time — published from the mid-nineteenth century to W.W. II. Each volume amounts to hundreds of pages of dense printed words that contain very rich and complete information on the individuals that lived and worked in China and Asia. The English-language publications (*Asia Directory and Chronicle*, *Hong List*, *Who's Who*, to name a few) document with precision who was doing what (position, activity), where (location, movement), and when (exact date, fuzzy dating). The Japanese and of course, the Chinese have produced a wealth of directories, yearbooks, *Who's Who*, professional *minglu* (register) of all kinds. The cross-compilation of these documents often allows to solve pending issues of imprecise data (dates).

A pilot project by scholars at the Basel Institute for European Global Studies, Westminster University (Data Futures) and IrAsia (P.I.) has established a **successful proof of concept** through a process of intelligent

digitization to automatically extract data on hundreds of thousands of actors while keeping full traceability and connection to the original source, down to the relevant fragment of information on page, and export as structured data into a processing platform (WP1-3). The pilot project makes it possible to reconstitute in its entirety the recorded foreign presence in China and in the whole Asian region over a century (Herren 2017). Processed through a GIS it documents in great detail the progressive extension of the Western presence in the region, which can then be analyzed by place, occupation, companies, groups of actors, etc. (see Figure 1). This preliminary mapping of the Western presence also confirms the **centrality of the three selected urban centers of the project**. Because data extraction is implemented without human intervention, it constitutes a major **technological breakthrough** to retrieve the formidable wealth of information that lay dormant in similar essential works in other languages (Chinese, Japanese). This approach was applied successfully to Chinese-language treatises in another small unfunded project. Considerable progress has been made in developing these techniques and the key challenge here is to enhance them further into a generic model applicable and adjustable to various similar materials. This is a path-breaking step beyond this project.

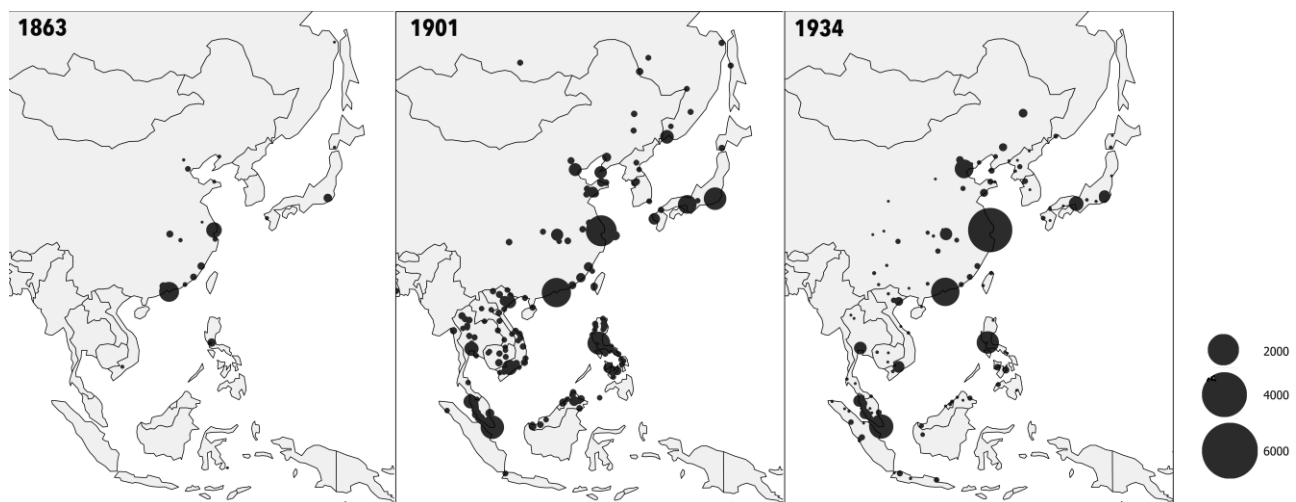


Figure 1. *The expansion of Western companies in Asia based on the data extracted from a sample of the Far East Foreigners Directory series. Caption: number of employees.*

The **second block of primary documentation** we use is the press (WP1-2). Newspapers and periodicals provide the most complete set of “observations” on everyday life. They reported on social or political activities, the meetings of associations, ceremonies, sport meetings, the coming and going of ships and passengers, etc. They documented the place, time, and nature of actions at the most elementary level. They caught actors in their tracks, from the moment they appeared “on print” to the time when they vanished or the medium itself disappeared. In the case of China, historical newspapers and periodicals have been turned into fully searchable digital collections at an amazing pace and scope. The P.I. has identified **no less than 20 periodicals**, in Chinese and English, available in digital repositories for the three selected urban centers. This amounts approximately to the volume of British newspapers Ian Gregory has explored and mined successfully in a previous ERC project in spatial humanities (Gregory 2012-16). Because the press started only in the mid-nineteenth century in China, with uneven developments and coverage in the selected urban areas, the project will use **alternative sources**, especially the rich biographical sections of local gazetteers (*difangzhi*), the *Wenshiziliao* (Materials in history and culture), the Chinese civil records (*Qingdai jinshenlu jicheng*), the entire collections of which are available in digital format, except the Chinese civil records that exist only in print and will be processed with the same methodology as WP1-1.

This WP1-2 will implement in parallel **two novel methodologies** to retrieve historical information from periodicals and other sources. The first one, used mostly in literary studies, but also in a few history projects (Gregory 2016; 2017; Cordell 2017) relies on the established **techniques of computational linguistics for the analysis of large corpora of texts** (Oltamari 2010, Gibbs and Cohen 2011, Rockwell 2016). These methods are relevant to our sources in two major respects. They can produce frequency counts of words, distributional analysis of word occurrences, diachronic models of word usage, which enable historians to search for specific expressions, terms and to assess their importance and relevance across corpora. Natural Language Processing also features semantic-tagging – the automatic grouping of words into categories based on meaning. It allows to specify a topic (war, tax, etc.) at a higher level of generality to perform collocation analysis (Prevot et al. 2009, Sajous et al. 2013). The key challenge here is to apply these techniques not just to Chinese, but to successive versions of the Chinese language, from near classical (1830s-1870s) to near contemporary (1930s-40s). We will start by selecting sample years (one per decade) to develop and program these techniques based

on our existing tools and take the most well-known Shanghai newspaper, *Shenbao*, as a test bed. We expect the initial phase to take 6-8 months. As the set of techniques reaches completion, we will apply it to the other selected corpora. This will spread over an active phase of 24 months for both English and Chinese corpora, and a follow up phase (6 months) depending on research questions.

The project will also put in the hands of historians **machine learning tools** to recognize patterns of interest in texts such as a conjunction of events. Machine learning is a set of very powerful tools, from the domain of artificial intelligence, which can be very effective for analyzing data, but requires a level of expertise generally not developed by humanities researchers (Fabre et al. 2012, Favre & Liu 2013). In short, the exploration of texts relies on algorithms that explore a text through successive iterations based on human annotations (Favre 2016; Prévôt 2015). These annotations serve to determine patterns (e.g. relationships) in the text that the “machine learns” and uses as guides to enrich and sharpen its capability to target and retrieve the most relevant information. The project will **create tools** which can help researchers to quickly produce high quality annotations of the phenomena they are interested in, through an active learning loop and a set of consistency measurements, to build effective predictors (Ji & Favre 2013). We propose to jointly tackle human annotation (the teaching part) and algorithms (the learning process) in a coherent and interoperable package, with **tools to verify/validate what machine learning algorithms have discovered on their own to address the particular challenges of historical data**. This will also open the way to explore the whole corpus of secondary literature in Chinese history for information on the individuals and actors identified in WP1-1. We expect the initial phase to take 6-8 months, followed by an active phase of 24 months for both English and Chinese corpora.

The combined use of methods in automatic extraction (WP1-1) computational linguistics, data mining and machine learning (WP1-2) will produce large quantities of “raw data”, by which we mean data that still require verification, cleaning, and refining for further use in historical analysis. What is common to any form of quantitative work represents here a genuine challenge in this project due to the scale of collected data. To address the potential bottleneck that processing such a vast quantity of data represents, WP1-3 will rely on **an advanced data platform (Actoz/freizo)** that brings together in a fully operational workflow two complementary database structures that perform separately two essential operations. This combined platform shall not be seen as just a repository — a place to store information for later retrieval — but as **a major methodological tool** (Dedieu 2014, Evan Davis 2017). The pilot project on the *Asia Directory and Chronicle* has produced a platform capable of processing massive quantities of data through techniques that brought down processing time by users (historians, Ph.D. students) from months for a given task to a matter of days or even hours. The **first component** provides tools for organizing, cleaning, and refining data through automatic and semi-automatic procedures. It operates on the basis of unique identifiers for actors, place, documents, etc., whatever their form in the original source (spelling, language, etc.). These identifiers also serve as automatic connectors between a recorded entity (e.g. actor) in the database and any new information on the same entity that will accrue from harvesting the sources. The **second component (freizo)** of the platform enables four functions: (1) it can incorporate and integrate data from any data silo in a unified format and build new comprehensive collections (e.g. importing data from existing databases, or producing a collection based on a particular element, like images, in a given corpus); (2) it guarantees data long-term sustainability independent from the original applications under which the data was created; (3) it records data sets at intermediate stages of research with complete metadata and provides full scalability of data collections; (4) it transforms the data into formats adapted for third-party applications (e.g. network analysis, GIS, visualization tools, etc.). The PI has used this combined platform in recent projects with data sets that contained tens of thousands of individual complex records (cadaster, industry, entertainment). A show case demonstrating the steps and procedures implemented through the platform is available on the [project web page](#). We expect the programming phase to build crosswalks to ingest the data from data mining and export routines to third-party applications to take 3 months. The process of cleaning and refining the data will be spread over a period of 30 months.

In this work package (all three WP1), the PI will be assisted by a computer engineer specializing in data mining, a postdoctoral fellow in computational linguistics (2 years), a postdoctoral fellow (0.2FTE) in history with experience in digital methods and PhD students (history) who will be trained in the implemented digital methods. The developments in WP1-1 and W1-3 will involve the two scholars, JP Dedieu and P. Cornwell, who designed the tools used in these work packages. The expertise for WP1-2 will be based on combining the PI’s deep knowledge of the nature of sources and registers of language with, L. Prévôt’s high expertise in language annotation, B. Favre’s innovative work in machine learning applied to a set of historical texts in French and a data set with millions of live chat exchanges, and P. Bellot’s extensive similar experience with the vast corpora of *Open Edition* scholarly literature.

The combined use of these methods will enable the collection of data on a vast number of historical figures and organizations, the cross-examination of different aspects of their lives/trajectories, and the integration of

their biographical information, which had hitherto been scattered in millions of pages of various historical sources. This constitutes a **path-breaking change of scale** in making historical information available for research.

b3. Data-rich history: large-scale data in historical research

We propose to address the massive character of the collected data and engage in historical analysis in two steps. The first one is to analyze the **data through the lens of social network analysis (SNA)**, in combination with GIS to enable spatialization and timelines to reintroduce time in the analysis (WP2-1). The second one is to take **selected groups as case studies** based on a wider range of qualitative documentation (WP2-2).

WP2-1 China elites: From large-scale database to data-rich history

The project makes the hypothesis that **China represents a case that does not fit in the usual paradigms** through which historians have studied the relationship between the dominant Western powers or Japan and more broadly the colonial world. It argues that China provides a particularly salient case for the study of the transnational flows of people that created, within shifting power configurations, the conditions for the emergence of overlapping, interlocked or juxtaposed elite networks and spheres of power. We propose to adopt **a perspective based on empirical data** on what the elites actually did. This approach will provide substantial support to revisit the debate on the nature of Chinese society with large scale data. The previous debate has revolved around two main historical objects: organizations, in particular merchant organizations, and public opinion as construed by the press, mostly periodicals run by literati/intellectuals. The debate, however, left out entirely the foreign presence and role in China. **Our approach does not preordain who were the actors that mattered.** More importantly it bridges the dominant Chinese/foreigners divide. Furthermore, while much has been written about the role of *guanxi* (relations) in Chinese society, in works inspired by prosopography (Bergère 1986, Henriot 1993), *guanxi* were substantiated only through a limited number of parameters (provincial origin, education, etc.). The project will examine these parameters along with a broad range of joint activities and shared experiences, both in the public and private spheres.

The project shifts the focus methodologically by mining vast corpora of mediums that were precisely the vectors of the alleged civil society and the very places where public spheres may have been constructed. Qualitatively, it not only permits retrieval of any piece of information from the news articles, announcements, and advertisements that constituted the three main components of any periodical, but creates new — and critically open and sustainable — research collections for scrutiny and scholarly endeavor by future scholars. It will represent a leap forward in our capacity to identify individual and collective actors precisely in relation to specific events, even the most ordinary ones (start of a business, a publication, wedding, etc.).

This WP will first construct the database **documenting the formation and evolution of self-organizing networks**. The research will focus on the major political and social organizations as well as the figures that dominated the major cities selected for this project. It will rely on the rich list of names collected from the secondary and primary sources in Chinese, Japanese, and Western languages processed in WP1-1. The identification of this first circle — running in the tens of thousands of names — will provide the initial lists (individuals, organizations) to proceed with data mining in the major digital corpora (e.g. newspapers) based on named entity recognition and semantic analysis to enlarge the scope of the social landscape and provide the information for the identification/construction of elite networks (WP1-2). The database will include the members of the Chinese and foreign political, business, and social elite residents in the selected cities. It will **trace the trajectories of these individuals/networks** from the native place and their circulations until they reached their place(s) of settlement. Second, the WP will provide a **spatial visualization** of the places, at various scales (city, country, continent, etc.), where the members of the elite groups lived, stayed, visited, etc. in order to identify spatial clusters and patterns of circulation/migration and measure such flows. It will aim at identifying spatial and social clusters within the cobweb of networks that may highlight the determinants of collaboration/competition among individuals, groups, organizations, the degree of relevance and impact of the expansion/contraction of networks on social order, identity politics, and the advent/management of crisis in the cities and the regions of contact around them.

Rather than assuming the existence of elite networks and selecting obvious candidates, our interest lies precisely in **establishing which networks emerged over time**, which factors prevailed in network formation, to what extent these networks were sustainable and formalized networks, how they evolved and transmuted with political, social, and economic change. We propose to examine where and how Chinese, Western, and Japanese elites interacted in the economic, political and intellectual spheres, how far and how deeply their networks interlocked, how power actually operated both in times of crisis and in regular times, how foreign and Chinese co-operated and competed in the economic realm. What we intend to examine is the extent of

cooperation, friction, and competition in local networks, the function of networks as vehicles for power and legitimacy, and their role in social management (welfare, mobilization).

Shanghai, Beijing/Tianjin, Canton/Hong Kong were made up of communities of sojourners. Previous studies have established beyond any doubt the importance of native place ties in social, economic and political life. The major organizations that brought together the different communities at various degrees of granularity (one county, a group of counties, a whole province, etc.) – *huiguan* (“guild”), *gongsuo* (“trade association”), *tongxianhui* (native-place association) – constituted formal networks that reached deep into local society and had the strongest influence in shaping urban social and economic life. In a similar way, foreigners flocked to China in increasing numbers and variety. The largest communities almost throughout the whole period were the British, the Japanese, the French and the Americans. These four groups will obviously be our main target groups, in part because they are much more documented, in part due to their influence in all walks of life in the selected cities, in part because they belonged to larger empires within which many held successive positions, and finally because their status underwent tremendous change and debasement when the Sino-Japanese war broke out in 1937.

The **elite networks** that connected these various groups **extended far beyond the city’s borders**, nationally and internationally, through the extensive networks that merchants, administrators, military officers, and other groups formed within the formal and informal Western and Japanese empires. Within China, the project will examine the spatial scope of elite networks to take into account the multiple connections these networks maintained between Shanghai and the major areas from which sojourners originated, namely the Jiangnan region, the Beijing-Tianjin area, and of course Canton/Hong Kong. The Chinese elites, however, circulated also in Japan, Southeast Asia, Australia, and Western countries (e.g. the founders of Chinese department stores in Shanghai were overseas Chinese). The members of the foreign elites came by definition from abroad, but they very often made more than one stop in the course of their migration to (and from) China. They moved through the space of the various empires as well as within the domestic space of China.

The main methodology in WP2-1 is Social Network Analysis (SNA). The PI takes SNA as *a tool for analysis* to delve into the massive amount of data, not as a device to “reveal networks” (Cerrutti 2008, Lemerrier 2015). The P.I. has applied this approach to two different sub-sets of data on Shanghai (entertainment; real estate market over a century) that proved beyond doubt that network analysis provides for example (real estate) considerable insights into the spatial logics, firm strategies, formal and informal networks of land-owners, etc. (Figure 2). **SNA will serve to investigate** (1) the venues and arenas of social, political, economic, and cultural transaction such as associations, clubs, companies, etc. to examine the forms of relationships and modes of elite formation; (2) the modalities of constitution, transformation, and extension of self-organizing networks among elite groups based on strong empirical data; (3) the patterns of mobilization and action in the public sphere with a focus on significant events, not as they are known as such by historians, but as they were unfolding; (4) the time-space of elite formation across China and beyond (circulations and migrations) to ground network analysis on actual terrain and questions of periodization in modern Chinese history.

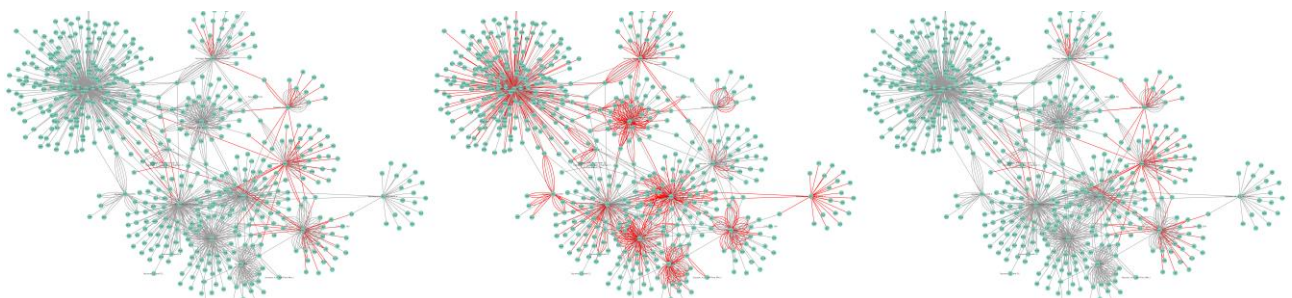


Figure 2: *The Sassoon family and associates’ network of land ownership in Shanghai (left to right): i. 1853-83. ii. 1903-24. iii. 1933-36. The green dots show the land plots. The red lines show ownership. Initially a small and loose network of four major individuals, it evolved into large and tight network around four new individuals with maximum ownership in the 1920s. By 1936, most Sassoons had disposed of their properties, following the 1932 Shanghai Battle. By 1937, all properties had virtually disappeared.*

One of the weaknesses of SNA for historians is the issue of actually measuring the nature of relationships beyond the simple fact of coming from the same class and university, of belonging to the same board of administration, etc. Precisely, the *Actoz/freiz0* database records not just “raw tabular data”, but **cross-referenced relationships linking actors, places and times which are traceable in the source materials**. The accumulation of such indices — the multiplicity of activities in which actors were involved, including, for

individuals, qualitative personal, even subjective, information from documents such as diaries, memoirs, etc. — will constitute a reasonable ground for delineating the limits of social networks. This is where the combination of tools in computational linguistics, machine learning and the *Actoz/freizo* database will provide a powerful resource to scratch beyond the surface of historical data in predetermined sources.

With SNA we seek to address issues such as: in cities with such different backgrounds what was the nature of power and what role did the elites play in the transformative social, political, and economic process between 1830 and 1949? In view of the transregional and transnational nature of the “urban elites”, how did networks emerge, operate, cooperate and intersect in the circulation of political, economic and social norms and practices within the vast cobweb that the activities of the elites traced in and beyond China? How can we periodize the processes of transformation and conversion of elite groups in these different urban centers? How and to what degree did elite networks extend across the nation and across empires? Were empires providers of foreign and Chinese elites, especially among the overseas Chinese communities? How did empires contribute to the transnationalization of elites in urban China? How, through which channels, and for what reasons did individual actors circulate in the course of their life? To what extent the application of new methods (SNA, computational linguistics, GIS) to massive digital corpora can contribute to a regeneration of Chinese political and social history? Beyond the use of these methods individually and separately, can we demonstrate the benefit of combining and merging the use of biographical databases, digital text analysis, historical GIS and SNA in the study of late imperial and republican Chinese elites?

The WP also **addresses the disconnect between network analysis and the spatial and temporal dimensions of social networks**. For the first dimension, the project will use GIS in combination with SNA to reintroduce space in network analysis (e.g. [Visone](#)). This is crucial to delineate the spatial gradations among social networks and to shift the emphasis on circulations, migrations, and flows across cities and countries. Can we discern spatial patterns in elite networks in China? How can we factor in space as an explanatory element and what does space reveal about network formation and the nature of networks? Can space be used as an operative agent to categorize networks? Since we use newspapers and periodicals, the circulation of news and texts across periodicals can also be read as other forms of networks over space (replication of same news, individual or collective proclamations, calls for action, etc.).

To address the second disconnect —time — we propose to re-inject networks into time, concretely, to use the time-stamped markers of networks as data points **in a timeline tool**. The objective is to study the interactions between the formation of elites and elite networks, the economic situation, the political and military crises at different scales (city, nation, global). In this way, the history of the elites will be articulated to the jolts of social, political and military history, to the shudders of the economic conjuncture, and to the whirlwinds of cultural life. This approach focuses on (i) Observing and analyzing the multiple rhythms proper to the formation of the elites in their respective spaces; (ii) Endeavoring to periodize as such the process of elite transformation. **Rather than a linear and univocal chronology, we propose a pluralistic and multiscalar approach**. Applications such as *La ligne du temps* (Quebec) or *Topotime* (Stanford) provide advanced visualization at various scales, events annotation, parallel imbedded timelines, and even depth of events. Timelines have been experimented effectively in a study of advertising in Shanghai (Armand 2017). Were all the spaces (cities) and all the facets of the elites affected in the same way, at the same moments, at the same rhythms? Was the effect of crises or prosperity immediate or delayed, even retroactive? Which actors in particular benefited from wars and crises and to what extent?

WP2-2 Case studies: quantitative/qualitative crosswalks

While we are confident that the methodologies applied to the global data will allow us to go a considerable distance towards explaining the transformative process of elite groups in China, we still believe that case studies, which can be contextualized by the broader work and make use of a wider range of sources, have much to offer in terms of delivering the nuance and detail of the mechanisms that drove change. This task seeks to redress the lens we use in WP2-1, especially the focus on periodicals as our main source, and **to implement a more qualitative approach based on network sampling**. In this task, pre-determined elite groups are used as the foci for a combined approach through the data mining of digital corpora and the close reading of historical materials. They constitute test beds for our research apparatus, both in substantive terms (social network analysis applied to elites) and in methodological terms (data mining). Except for women, the selected groups are based on their belonging to an identifiable cluster of occupation/position/status. They represent occupations that either had a long continuity, though they underwent an important renewal (administrators, military, banking, publishing) or constitute totally new occupations (industry, press, advertising). Because their members were public or prominent figures at various degrees, there is a high probability of their “presence” in contemporary media, professional directories, biographical dictionaries, and even current web resources. Moreover, we argue that the groups considered here represent networks within more or less interlocked

networks (Abbott 2016).

The **reasons for this choice** are: i. This approach links the “distant/close” reading methodologies and reintroduce a more qualitative approach; ii. These groups overlapped or were interrelated to various degrees, and involved Chinese and foreigners; iii. We can use data sets produced in previous projects, at least for Shanghai (industrialists 1932-1937; land owners 1846-1941; entertainment 1907-1967). In addition to the main sources discussed in WP1-1 & 1-2, the case studies will use qualitative materials like memoirs, diaries, private papers, company archives (e.g. BAT, Jardine, HSBC, Swire), and selected archival documents at archival institutions (Shanghai, NARA, British and French national and diplomatic archives). Qing high officials have left copious diaries compiled and edited in *Qingdai riji huichao* (Qing era diary collection, 1982). Major republican figures, such as Song Ziwen (political/banking), Zhang Jia’ao (banking), Norwood F. Allman (businessman), etc. have bequeathed all their private papers to American institutions (Hoover Institution). Major collections based on oral history (Columbia University, Academia Sinica) still await a more systematic examination. In the world of business, banks and credit information agencies produced detailed surveys of companies and their managers and investors. They offer precious information on business actors. We will pursue here the research questions outlined in WP2-1.

The first group, **the administrative (or bureaucratic) elite**, including the military, embodies formal power in cities with very different bureaucratic histories: a national imperial elite in Beijing, juxtaposed Chinese and foreign bureaucracies in Shanghai and Tianjin, a British-run colonial bureaucracy in Hong Kong, and a local Chinese bureaucracy in Canton. All went through the changes of regimes that produced shifting power configurations at various scales before China was reunited in 1949. While the ruling elite of Confucian literati-officials may have vanished from the historical stage as a social group, the individuals that composed this group did not evaporate. They were at various stages of their life and career. They were imbedded in more or less extensive family lineages, professional milieux, cultural circles, etc. Beyond the obsolete merchant/official dichotomy or the blurry notion of an increasing fusion, the project will shed light on actual individuals, on their life course and concrete actions and on the transformative process of the administrative elite.

The second group is the **business elite**, with an explicit focus on finance and industry. Both included Chinese and foreigners and, among the Chinese, significant numbers of two central communities, Ningbonese (Ningbo) and Cantonese who were not just present in all the selected urban centers, but formed groups that were deeply enmeshed in social control and power in Chinese cities. The business elite makes a particularly relevant case because it was a well-structured and informed professional group, with organizations, publications, etc. The two targeted groups, the businessmen from Guangdong and Ningbo were major players in these fields before and after the opening of China to foreign trade. But with foreigners came new players, new forms of trade and new institutions (e.g. modern banking). It aims to challenge prevailing orthodoxies such as the role and nature of compradors, hardly a category in itself and not one that should be studied in isolation, or that of a burgeoning bourgeoisie at the turn of the century, a splendid but short-lived “golden age” during and after W.W.I and a crippled bourgeoisie on the wane after 1927. The political perspective that prevailed fails to do justice to social and economic facts, that of an irresistible force that drove entrepreneurs, especially industrialists, all through the Republican period, even during periods of war, and testifies to the extraordinary resilience and strength of the business community in urban China.

The third group, **press, publishing, and advertising**, was made up of both established and emerging professions that exhibited multiple and dense connections. The rise of the press was intimately linked to the advent of modern publishing, while modern advertising after W.W.I exhibited dense crosswalks at the level of actors with the press, secondarily publishing, and beyond with the business and administrative elites. Armand has established the existence of intricate networks between actors in the press and advertisers, although she only scratched the surface in her pioneering work (Armand 2017). The press and the publishing industry had multiple connections, not the least because publishers edited both books and periodicals. It is also a broad sector in which women emerged early and eventually played a significant role, as direct actors in the press, but also as social actors about whom the press reported and even built discourses, dedicated supplements and even publications.

On **women**, we adopt deliberately the opposite approach, not to start from pre-defined “elite women”, but to explore where and how women figured in urban elite circles and politics. We make a case about women both at the level of actual role as/in elite groups and representation in the press. Women from elite background, but in general as well, experienced a drastic change at three major levels: physical, with the prohibition of bound feet; educational, with the establishment of schools for girls, then a general support for school education up to university level; social with their increasing presence in the public space. It is assumed that women figured prominently in the fields of culture and entertainment. It was indeed the case, as work on women’s journal or tabloids has shown (Yang 1999, Mittler et al. forthcoming, Huang 2004, 2005). Work on the occupational structure of Shanghai, however, has revealed the presence of women in a wide range of

occupations, many among elite groups (e.g. lawyers). Although they were a minority, the project will examine where these “elite women” were active, in which positions, which companies or institutions, etc. The increasing number of female university graduates (Chinese and foreigners) in the last decades of the Republican period changed the labor market markedly. How did this translate at the level of elites? What were their relations to men/male elites, in connection to the issue of power? Did they face difficulties to exercise power because of their sex? Is the concept of networks operative among women involved in elite groups? Was there enough density in social activity and interaction to support the emergence of women's networks? Is it an issue of substance or (mis)representation in the sources?

The main difficulties to be expected in the empirical analysis of elite networks in urban China are due to, first and foremost, the large number of individuals involved in these networks, and therefore the considerable volume of information that will be required for the project to aim at a complete coverage. The research team includes historians fully trained in quantitative methods in history, as well as digital technologies in the humanities, computer scientists fully versed in data & text mining methods and machine learning, linguists trained in corpus linguistics, Natural Language Processing, and corpus annotation. A second less obvious difficulty is related to the multilingual nature of the primary and secondary sources, both for actual close reading and data mining. The P.I., as well as the associate scholars who have agreed to contribute to the project, bring together a full ability to read in classical and modern Chinese, Japanese, French, English and most other possibly relevant European languages (German, Italian, Spanish).

Our project can only begin to unleash the potential of these approaches. This project will therefore genuinely create a step-change to historical scholarship by re-invigorating the study of the past through the use of the selected cutting-edge digital methods. As such it will turn “digital history” from an ideal project into a mature field. The project will have **a major influence on the Digital Humanities** by showing and establishing why these approaches can and will make significant contributions to knowledge. The impact will thus spread beyond these disciplines to other humanities subjects.

WP3 Meeting, training, and publicizing

Beyond the project itself, **we believe it is necessary to take further measures in training, networking and communication to widen the impact of our methods and results.** This will happen in several ways: We will host two expert meetings (Tsk3a) in the first stage of the project. We will invite around 15 international scholars with broad expertise in their field from Europe, North America and Asia. Each meeting will be themed according to the two substantive work-packages above. Also in year 5, we will host a major international conference. The two-day workshops (Tsk3b) are meant to provide an introduction to the methods and applications used in the project to scholars and graduate students (open and free access). The two week-long summer schools (Tsk3c) will prepare participants to engage in an in-depth exploration of the fundamental applications and methods of the project. They will allow the participants to bring their own sources or data and use them with our tools. The summer schools will take place either in the early or late part of the summer and each will take a maximum of 20 participants. The project will provide the basic tuition costs and lunches, but participants will be expected to pay for their own accommodation, meals and travel. The two-day workshops (Tsk3b) will take place in years 2 and 3, the summer schools in year 3 and 4 at Aix-en-Provence or partner locations. The project will also rely on a web site and a digital platform for communication and for training (Tsk3e). While the web site will serve essentially for communication purposes, the digital platform will represent the main interface to access the produced data and research results. The data sets will be made available on line for public consultation and use, in particular through the Pilot on Open Research Data in H2020. To help with the use of digital methods (data mining, SNA, GIS), we plan to design graduate training modules (Tsk3d) based on the data of the project to contribute to developing digital skills in historical research and the humanities.

Gantt chart

List of work packages (full Gantt chart on *GanttProject*):

WP0 Project management (PI)

WP1 Methodologies: From texts to historical "big data"(divided in 3 sub-WPs)

1a Project preparation and staff: recruitment (PI) – **WP1-1** (1-1a Collecting and: digitizing print – (1-1b Development of: generic techniques – (1-1c Apply to: English-language directories – (1-1d Apply to: Chinese and Japanese – **WP1-2** (1-2a Collecting and: preparing databases –(1-2b Designing and: testing standard – (1-2c Designing and: testing machine – (1-2d Apply to: English-language newspapers –(1-2e Apply to: Chinese newspapers – (1-2f Designing generic user interface – **WP1-3** (1-3 Develop generic import crosswalk – (1-3b Processing, cleaning and refining – (1-3c Develop dictionaries: and gazetteers – (1-3d Develop export routines

WP2-1 Urban elites: From large-scale database to data-rich history

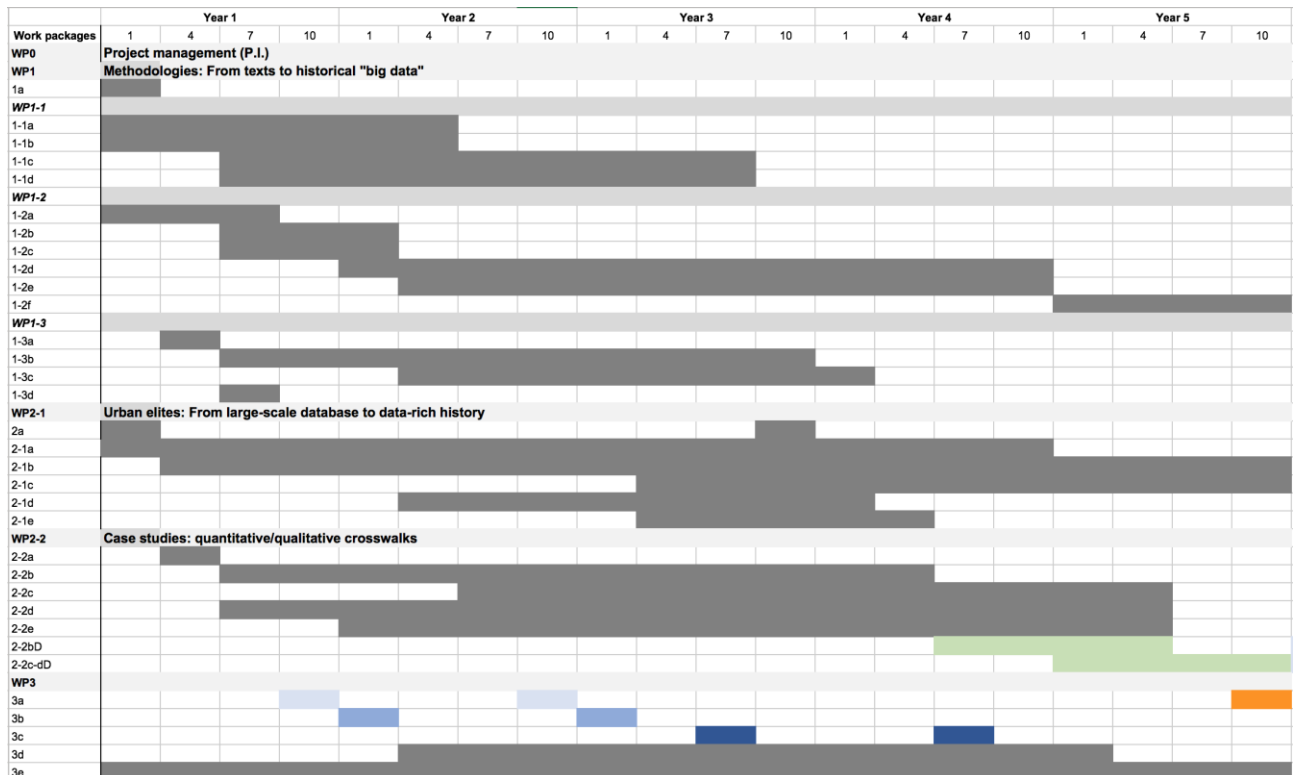
2a Project preparation and staff recruitment – (2-1a Data repository: quantitative and qualitative analysis – (2-1b SNA: Network formation, patterns – (2-1c SNA: local case studies – (2-1d The spaces of networks – (2-1e Timelines of elite networks

WP2-2 Case studies: quantitative/qualitative crosswalks

2-2a Project preparation and staff: recruitment (PI – (2-2b Administrative elite – (2-2c Business elites (banks/Industry; Ningbo, Cantonese, foreigners) – (2-2d Press, publishing, and advertising – (2-2e Women in elite circles – (2-2bD Write up: Dissertation – (2-2c-dD Write up: Dissertation

WP3 Meeting, training, and publicizing

(3a Expert workshops – (3b International conference – (3c Two-day workshops – (3d Training modules in DM/SNA - (3e Web platform

**Section c. Resources (including project costs)***Host institution and research environment*

Aix-Marseille University (AMU) is one of the leading French Universities (Excellence University). It provides an outstanding academic environment for this project, with strong support to interdisciplinary approaches and cross-center research collaboration. In particular, the project will also rely significantly on two leading research centers in linguistics ([LPL](#)) and in computing ([LIE/LSIS](#)). It will be hosted in the Centre for Asian Studies ([IrAsia](#)), a multidisciplinary research center with a very strong commitment to comparative studies of modern and contemporary Asia.

Description of the team and organization

A Gantt chart shows the way that the major WPs are sub-divided into 33 tasks and the distribution of faculty and staff time between them (see above).

The requirements of the project combine a data-driven preoccupation using advanced techniques of data mining with the close reading and analysis of historical documents. The P.I has extensive experience in historical research and the use of the digital technologies implemented in the project. He will devote at least 0.45FTE of his time on the project and it will be his paramount commitment for the next five years. He will be based continuously in an EU country, except for research trips outside of Europe. He will conduct this project with a core team of 2+2 post-doctoral researchers (3+2 years), 3 PhD researchers (4 years), an engineer in computing and a data scientist. In the first two years of the project, the P.I. will recruit one postdoc in computational linguistics (WP1-2) and one postdoc in history, with expertise in digital methods (WP1-2 and WP2). In the second phase when most of the quantitative sources are already available and the qualitative

sources will be on the way, two new postdocs in history with expertise in digital methods will be recruited to work mostly on WP2-2 (3c and 3e). The project will require substantial programming to develop/adapt research tools and produce user interface. This explains the recruitment of an engineer in computing for the duration of the project (WP1). A data scientist will be necessary to address the massive amount of data to be processed with the main methods (SNA, GIS, timelines) implemented in the project. The project will also fund three PhD studentships in the 3b, 3d, and 3e case studies. The exact topics of the PhD projects will be decided in discussion with the students. They will be kept as separate but parallel projects within the main WP2-2. The first studentship (3b) will start in month 7. The other two studentships (3d, 3e) will start on month 16. In both cases, 36 months at most will be devoted to research and 12 months to writing.

The project will involve a group of external associates in history (R. Bickers, Bristol U., M. Herren, EIB, B. Mittler, U. of Heidelberg, Wu Jen-shu, Academia Sinica, Dominic Sachsenmaier, U. of Göttingen, X. Paulès, EHESS, L. Gabbiani, EFEO, JP Dedieu, CNRS) and computing (P. Cornwell, Data Futures) who have agreed to contribute to training PhD students and post-doctoral fellows and conduct their own research on several of the case studies identified in the project, and participate in workshops and conferences.

Budget:

The interdisciplinary nature of the project and its emphasis on digital methods and large-scale data requires recruiting postdocs and engineers to provide the necessary manpower for the successful completion of the various WPs. The project will also fund three PhD studentships. Altogether, these salaries represent 50% of the total budget, with an additional 17% for the P.I. and other senior scholars. The budget for travel (mostly research trips to libraries and archives and participation to conferences) amounts to 5% of all expenses. Durable equipment (computers, software) and consumables represent 2,5%. Other direct costs (subscriptions to digital databases, expert meetings, two-day seminars, and summer schools) amount to 5,5%.

| Cost Category | | | Total in Euro | |
|---|--|---|---------------|------------|
| Direct Costs | Personnel | PI | 326250,00 | |
| | | Senior Staff | 89150,00 | |
| | | Postdocs | 500000,00 | |
| | | Students | 320000,00 | |
| | | Other | 332000,00 | |
| | <i>i. Total Direct costs for Personnel (in Euro)</i> | | | 1677300,00 |
| | Travel | | | 123500,00 |
| | Equipment | | | 52000,00 |
| | Other goods and services | Consumables | 10000,00 | |
| | | Publications (including Open Access fees), etc. | | |
| | | Other (please specify) | 137100,00 | |
| | <i>ii. Total Other Direct Costs (in Euro)</i> | | | 322600,00 |
| 1 – Total Direct Costs (i + ii) (in Euro) | | | 2000000,00 | |
| 2 – Indirect Costs (overheads) 25% of Direct Costs (in Euro) | | | 500000,00 | |
| 3a) – Subcontracting Costs (no overheads) (in Euro) | | | | |
| 3b) – Other Direct Costs with no overheads (in Euro) | | | | |
| Total Estimated Eligible Costs (1 + 2 + 3) (in Euro) | | | 2500000,00 | |
| Total Requested Grant (in Euro) | | | 2500000,00 | |

| | |
|--|------------|
| For the above cost table, please indicate the % of working time the PI dedicates to the project over the period of the grant: | 45% |
|--|------------|

References

- Abbott, Andrew. 2005. "Linked Ecologies: States and Universities as Environments for Professions." *Sociological Theory* 23(3):245–74.
- Abbott, Andrew. 2016. "1. Écologies liées : À propos du système des professions." in *Les professions et leurs sociologies : Modèles théoriques, catégorisations, évolutions, Colloquium*, edited by P.-M. Menger. Paris: Éditions de la Maison des sciences de l'homme.
- Armand, Cécile. 2017. "'Placing the History of Advertising': Une histoire spatiale de la publicité à Shanghai." doctoral dissertation, ENS de Lyon, Lyon.
- Bellot, Patrice. 2011. *Recherche d'information contextuelle, assistée et personnalisée*. Paris: Hermès Science publications-Lavoisier.
- Belsky, Richard. 2005. *Localities at the Center: Native Place, Space, and Power in Late Imperial Beijing*. Cambridge, Mass.: Harvard University Asia Center.
- Bergère, Marie-Claire. 1968. *La bourgeoisie chinoise et la révolution de 1911*. La Haye: Mouton.
- Bergère, Marie-Claire. 1986. *L'âge d'or de La Bourgeoisie Chinoise, 1911-1937*. [Paris]: Flammarion.
- Bickers, Robert A. 1999. *Britain in China: Community, Culture and Colonialism 1900-1949*. Manchester; New York; New York: Manchester University Press.
- Bode, Katherine. 2017. "The Equivalence of 'Close' and 'Distant' Reading; or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly* 78(1):77–106.
- Bodenhamer, David J., John Corrigan, and Trevor M. Harris. 2010. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Bloomington: Indiana University Press.
- Bradley, John and Harold Short. 2005. "Texts into Databases: The Evolving Field of New-Style Prosopography." *Literary and Linguistic Computing* 20(Suppl):3–24.
- Briost, Pascal. 2006. "Analyse Des Réseaux et Prosopographie. Des Outils Pour l'étude Des Cercles Intellectuels de Aa Renaissance Anglaise." Pp. 267–79 in *Liens personnels, réseaux et solidarités en France et dans Les îles britanniques (XIe-XXe siècle)*, edited by D. Bates. Paris: Publications de la Sorbonne.
- Brook, Timothy. 2005. *Collaboration: Japanese Agents and Local Elites in Wartime China*. Cambridge Mass.: Harvard University Press.
- Brook, Timothy and Andre Schmid. 2000. *Nation Work Asian Elites and National Identities*. Ann Arbor: University of Michigan Press.
- Brughmans, Tom, Anna Collar, and Fiona Coward. 2016. *The Connected Past: Challenges to Network Studies in Archaeology and History*. Oxford University Press.
- Buiskool, Dirk A. 2009. "The Chinese Commercial Elite of Medan, 1890-1942: The Penang Connection." *Journal of the Malaysian Branch of the Royal Asiatic Society* 82(297):113–29.
- Camp, Matje van de. 2016. "A Link to the Past: Constructing Historical Social Networks from Unstructured Data." Doctoral dissertation, International Institute of Social History.
- Carney, T. F. 1973. "Prosopography: Payoffs and Pitfalls." *Phoenix* 27(2):156–79.
- Cerutti, Simona. 2008. "Histoire pragmatique, ou de la rencontre entre histoire sociale et histoire culturelle." *Tracés. Revue de Sciences humaines* (15):147–68.
- Chang, Chung-li. 1967. *The Chinese Gentry: Studies on Their Role in Nineteenth-Century Chinese Society*. Seattle: University of Washington Press.
- Chang P'eng-yuan. 1978. "Political Participation and Political Elites in Early Republican China: The Parliament of 1913-1914." *Journal of Asian Studies* 37(2):293–313.
- Charle, Christophe. 1980. *Prosopographie des élites françaises: (XVIe-XXe siècles) : guide de recherche*. [Paris]: Centre national de la recherche scientifique, Institut d'histoire moderne et contemporaine.
- Charle, Christophe. 1992. "The Present State of Research on the Social History of Elites and the Bourgeoisie: A Critical Assessment." *Contemporary European History* 1(1):99–112.
- Choi, Yuk Ping. 2001. "Social Networks and the Economic Adaptation of Immigrants: The Chaozhou and Fujian in Hong Kong." Doctoral dissertation, University of Oxford.
- Conrad, Sebastian. 2017. *What Is Global History?* Princeton: Princeton University Press.
- Cordell, Ryan. 2015. "Scale as Deformance." Retrieved August 5, 2017 (<http://ryancordell.org/research/scale-as-deformance/>).
- Cornwell, Peter, Dan Granville, Alexandra Eveleigh, Eric Decker, and Christian Henriot. 2015. "Improving Compliance with Evolving Standards Using Computed Transformation of Digital Collections." Sidney.

- Coste, Laurent, Stéphane Minvielle, François-Charles Mougel, and Gironde) Centre d'études des mondes moderne et contemporain (Pessac. 2014. *Le concept d'élites en Europe de l'Antiquité à nos jours*. Pessac: Maison des sciences de l'homme d'Aquitaine.
- De Weerd, Hilde. 2007. *Competition over Content: Negotiating Standards for the Civil Service Examinations in Imperial China (1127-1279)*. Cambridge, Mass.: Harvard University Asia Center : Distributed by Harvard University Press.
- De Weerd, Hilde. 2013. "Communication and Empire." Retrieved August 27, 2017 (<http://chinese-empires.eu/>).
- De Weerd, Hilde. 2016. *Information, Territory, and Networks: The Crisis and Maintenance of Empire in Song China*. Harvard University Asia Center.
- Dedieu, Jean-Pierre. 2005. "Prosopographie rénovée et réseaux sociaux. Les agents de l'Etat en France et en Espagne de la fin du XVIe au début du XIXe Siècle. Dix ans de publications." Pp. 281–96 in *Jahrbuch für europäische Verwaltungsgeschichte*. Berlin: H-Soz-Kult.
- Dedieu, Jean-Pierre. 2014. "Three Pillars of Historical Wisdom: Atomization, Data Building and Flexibility." Retrieved April 2, 2016 (<https://halshs.archives-ouvertes.fr/halshs-00973443/document>).
- Dedieu, Jean-Pierre and Alvaro Chaparro. 2013. "De la base au réseau. L'apport des bases de données à l'étude des réseaux." *Les Cahiers de Framespa : Nouveaux champs de l'histoire sociale* s. p. (en ligne).
- Domhoff, G. William. 1967. *Who Rules America?* Englewood Cliffs, N.J.: Prentice-Hall.
- Domhoff, G. William. 2012. "An Invitation to a Four-Network Theory of Power: A New Viewpoint Compatible with Elite Theory." *Einladung Zu Einer Netzwerktheorie von Macht: Ein Neuer Ansatz Zur Verbindung von Machttheorie Und Elitetheorie*. 37(1):23–37.
- Elman, Benjamin A. 2013. *Civil Examinations and Meritocracy in Late Imperial China*. Cambridge, Mass.: Harvard University Press.
- Elvin, Mark. 1967. "The Gentry Democracy in Shanghai 1905-1914." Oxford University, Oxford.
- Esherick, Joseph. 1976. *Reform and Revolution in China: The 1911 Revolution in Hunan and Hubei*. Berkeley: University of California Press.
- Esherick, Joseph and Mary Backus Rankin, eds. 1990. *Chinese Local Elites and Patterns of Dominance*. Berkeley: University of California Press.
- Evan Davis, Matthew. 2017. "The Database as a Methodological Tool." Retrieved (<https://digitalmedievalist.wordpress.com/2017/08/10/the-database-as-a-methodological-tool/>).
- Fabre, Benoît et al. 2016. "Summarizing Behaviour: An Experiment on the Annotation of Call-Centre Conversations."
- Fabre, Benoit and Shuhua Liu. 2013. "Understand the Global Economic Crisis: A Text Summarization Approach." *Intelligent Systems in Accounting, Finance and Management* 20(2):89–110.
- Fewsmith, Joseph. 1985. *Party, State, and Local Elites in Republican China: Merchant Organizations and Politics in Shanghai, 1890-1930*. Honolulu: University of Hawaii Press.
- Gabbiani, Luca. 2011. *Pékin à l'ombre du mandat céleste: vie quotidienne et gouvernement urbain sous la dynastie Qing, 1644-1911*. Paris: Éd. de l'EHESS.
- Gao, Bei. 2013. *Shanghai Sanctuary: Chinese and Japanese Policy toward European Jewish Refugees during World War II*. Oxford: New York : Oxford University Press.
- Gibbs, Frederick W. and Daniel J. Cohen. 2011. "A Conversation with Data: Prospecting Victorian Words and Ideas." *Victorian Studies* 54(1):69–77.
- Goebel, Michael. 2017. *Anti-Imperial Metropolis: Interwar Paris and the Seeds of Third World Nationalism*. Cambridge UK ; New York NY: Cambridge University Press.
- Goodman, Bryna. 1995. *Native Place, City, and Nation : Regional Networks and Identities in Shanghai, 1853-1937*. Berkeley: University of California Press.
- Gregory, Ian. 2012. "Spatial Humanities | TEXTS, GIS & PLACES." Retrieved August 8, 2017 (<http://www.lancaster.ac.uk/fass/projects/spatialhum.wordpress/>).
- Gregory, Ian and Paul Atkinson. 2017. "Finding Child Welfare in Victorian Newspapers." *Journal of Interdisciplinary History* 48(2):159–86.
- Gregory, Ian, Paul Atkinson, Andrew Hardie, and Amelia Joulain-Jay. 2016. "From Digital Resources to Historical Scholarship with the British Library 19th Century Newspaper Collection." *Journal of Siberian Federal University* 9(4):994–1006.
- Gregory, Ian N. and Alistair Geddes. 2014. *Toward Spatial Humanities: Historical GIS and Spatial History*. Bloomington: Indiana University Press.
- Hakkani-Tür, Dilek, Tur Gokhan, Benoît Fabre, and Shriberg. 2012. "Finding the Structure of Documents." Pp. 29–49 in *Multilingual Natural Language Processing Applications: From Theory to Practice*, edited by D. Bikel and I. Zitouni. Indianapolis: IBM Press.

- Han, Seunghyun. 2016. *After the Prosperous Age: State and Elites in Early Nineteenth-Century Suzhou*. Cambridge, Mass.: Harvard University Asia Center.
- Hartmann, Michael. 2007. *The Sociology of Elites*. London; New York: Routledge.
- He, Yimin. 2014. "Chinese Urban History Studies Face the Twenty-First Century." *Chinese Studies in History* 47(3):73–99.
- Henriot, Christian. 1993. *Shanghai, 1927-1937: Municipal Power, Locality, and Modernization*. Berkeley: University of California Press.
- Henriot, Christian. 1996. "Cities and Urban Society in China in the Nineteenth and Twentieth Centuries: A Review Essay in Western Literature | Christian Henriot - Academia.Edu." *Newsletter for Modern Chinese History* (21):151–75.
- Herren, Madeleine. 2014. *Networking the International System Global Histories of International Organizations*. Retrieved August 27, 2017 (<http://dx.doi.org/10.1007/978-3-319-04211-4>).
- Herren, Madeleine. 2017. "[Global Information at a Glance: Power, Law, and Commerce through the Lens of Asia Directories](https://europa.unibas.ch/de/forschung/forschungsfeld-gesellschaft-und-geschichte/global-information-at-a-glance/)." Retrieved August 27, 2017 (<https://europa.unibas.ch/de/forschung/forschungsfeld-gesellschaft-und-geschichte/global-information-at-a-glance/>).
- Herren, Madeleine and Isabella Lohr. 2014. *Lives beyond Borders: A Social History, 1880-1950*. Leipzig: Leipziger Universitätsverlag.
- Hertz, Rosanna and Jonathan B. Imber. 1995. *Studying Elites Using Qualitative Methods*. Thousand Oaks, Calif.: Sage Publications.
- Higley, John and Michael G. Burton. 2006. *Elite Foundations of Liberal Democracy*. Lanham: Rowman & Littlefield Publishers.
- Ho, Ping-ti. 1962. *The Ladder of Success in Imperial China: Aspects of Social Mobility, 1368-1911*. New York: Columbia University Press.
- Hoerder, Dirk. 2011. *Cultures in Contact World Migrations in the Second Millennium*. Durham: Duke University Press.
- Hoerder, Dirk, Adrian Shubert, and Christiane Harzig. 2005. *The Historical Practice of Diversity: Transcultural Interactions from the Early Modern Mediterranean to the Postcolonial World*. New York, N.Y.; Oxford: Berghahn Books.
- Huang, Nicole. 2004. "Fashioning Public Intellectuals: Women's Print Culture in Occupied Shanghai (1941-1945)." Pp. 325–45 in *In the shadow of the rising sun : Shanghai under Japanese occupation*, edited by C. Henriot and W.-H. Yeh. Cambridge ; New York: Cambridge University Press.
- Huang, Nicole. 2005. *Women, War, Domesticity: Shanghai Literature and Popular Culture of the 1940s*. Leiden; Boston: Brill.
- Ji, Heng et al. 2013. "Open-Domain Multi-Document Summarization via Information Extraction: Challenges and Prospects." Pp. 177–201 in *Multi-source, Multilingual Information Extraction and Summarization, Theory and Applications of Natural Language Processing*. Springer, Berlin, Heidelberg. Retrieved August 27, 2017 (https://link.springer.com/chapter/10.1007/978-3-642-28569-1_9).
- Knowles, Anne Kelly. 1999. *Historical GIS: The Spatial Turn in Social Science History*. Durham, N.C.: Duke University Press.
- Knowles, Anne Kelly. 2004. *Past Time, Past Place: GIS for History*. Redlands, Calif.: ESRI Press.
- Knowles, Anne Kelly and Amy Hillier, eds. 2008. *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*. Redlands, Calif.: ESRI Press.
- Knowles, Anne Kelly, Robert E. Kennedy Library, and Data studio. 2014. *Visualizing History with GIS*. San Luis Obispo, Calif.: [Robert E. Kennedy Library].
- Korom, Philipp. 2015. "Elites: History of the Concept." Pp. 390–95 in *The International Encyclopedia of Social and Behavioral Sciences*. Leiden: Elsevier.
- Kwan, Man Bun. 2001. *The Salt Merchants of Tianjin : State-Making and Civil Society in Late Imperial China*. Honolulu: University of Hawai'i Press.
- Lachmann, Richard. 2010. *States and Power*. Cambridge, UK; Malden, MA: Polity Press.
- Lai, Chi-Kong. 1994. "Cantonese Business Networks in Late Nineteenth Century Shanghai: The Case of the Kwang-Chao Kung-So." *Essays in Economic & Business History* 12:145–54.
- Lee, Pui Tak. 1991. "Business Networks and Patterns of Cantonese Compradors and Merchants in Nineteenth-Century Hong Kong." *Journal of the Hong Kong Branch of the Royal Asiatic Society* 31:1–39.
- Lemercier, Claire. 2003. "[Business Networks and Meso-Level Institutions in Nineteenth-Century Paris](https://hal-sciencespo.archives-ouvertes.fr/hal-01053333/document)." Retrieved April 2, 2016 (<https://hal-sciencespo.archives-ouvertes.fr/hal-01053333/document>).
- Lemercier, Claire. 2005. "Analyse de réseaux et histoire." *Revue d'histoire moderne et contemporaine* no52-2(2):88–112.
- Lemercier, Claire. 2010. "[Réseaux et groupes d'influence – bilan historiographique](https://hal-sciencespo.archives-ouvertes.fr/hal-01053333/document)." Retrieved April 2, 2016

- (<https://halshs.archives-ouvertes.fr/halshs-00522888/document>).
- Lemerrier, Claire. 2015. "Formal Network Methods in History: Why and How?" in *Social Networks, Political Institutions, and Rural Societies*, edited by G. Fertig. Turnhout: Brepols.
- Lemerrier, Claire and Emmanuelle Picard. 2010. "[Quelle approche prosopographique ?](#)" Retrieved April 2, 2016 (<https://halshs.archives-ouvertes.fr/halshs-00521512/document>).
- Lim, Jason. 2010. *Linking an Asian Transregional Commerce in Tea Overseas Chinese Merchants in the Fujian-Singapore Trade, 1920-1960*. Leiden; Boston: Brill.
- Mittler, Barbara, Michael Hockx, and Joan Judge, eds. forthcoming. *A Space of Their Own: Women and the Periodical Press in China's Long Twentieth Century*. Cambridge: Cambridge University Press.
- Oliver, Dev. 2016. *Spatial Network Data: Concepts and Techniques for Summarization*. Cham: Springer.
- Oltramari, Alessandro, Laurent Prevot, and Stefano Borgo. 2010. "Interfacing Ontologies and Lexical Resources." Pp. 185–200 in *Ontology and the Lexicon. A Natural Language Processing Perspective*. Cambridge: Cambridge University Press.
- Padgett, John F. and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology* 98(6):1259–1319.
- Paulès, Xavier. 2010. *Histoire d'une drogue en sursis l'opium à Canton, 1906-1936*. Paris: Éd. de l'École des Hautes Etudes en Sciences Sociales.
- Piotrowski, Michael. 2012. *Natural Language Processing for Historical Texts*. San Rafael, Calif.: Morgan & Claypool.
- Prévôt, Laurent, Pierre Magistry, and Chu-Ren Huang. 2015. "Un État Des Lieux Du Traitement Automatique Du Mandarin." *Faits de Langues* 46(2):51–70.
- Prévot, Laurent, Laure Vieu, and Nicholas Asher. 2009. "Une Formalisation plus Précise Pour Une Annotation Moins Confuse: La Relation d'Élaboration d'entité." *Journal of French Language Studies* 19(2):207–28.
- Rankin, Mary Backus. 1986. *Elite Activism and Political Transformation in China: Zhejiang Province, 1865-1911*. Stanford University Press.
- Rhoads, Edward J. M. 1975. *China's Republican Revolution: The Case of Kwangtung, 1895-1913*. Cambridge, Mass.: Harvard University Press.
- Ristaino, Marcia R. 2001. *Port of Last Resort: The Diaspora Communities of Shanghai*. Stanford, Calif.: Stanford University Press.
- Rowe, William T. 1984. *Hankow: Commerce and Society in a Chinese City, 1796-1889*. Stanford, Calif.: Stanford University Press.
- Rowe, William T. 1989. *Hankow: Conflict and Community in a Chinese City, 1796-1895*. Stanford, Calif.: Stanford University Press.
- Sachsenmaier, Dominic. 2011. *Global Perspectives on Global History Theories and Approaches in a Connected World*. Cambridge; New York: Cambridge University Press.
- Sajous, Franck, Emmanuel Navarro, Bruno Gaume, Laurent Prévot, and Yannick Chudy. 2013. "Semi-Automatic Enrichment of Crowdsourced Synonymy Networks: The WISIGOTH System Applied to Wiktionary." *Language Resources and Evaluation* 47(1):63–96.
- Schoppa, R.Keith. 1982. *Chinese Elites and Political Change: Zhejiang Province in the Early Twentieth Century*. Cambridge, Mass.: Harvard University Press.
- Sheehan, Brett. 2003. *Trust in Troubled Times: Money, Banks, and State-Society Relations in Republican Tianjin*. Cambridge, MA: Harvard University Press.
- Stapleton, Kristin and Haiyan Liu. 2006. "Chinese Urban History: State of the Field." *China Information* (20):391–427.
- Stone, Lawrence. 1971. "Prosopography." *Daedalus* 100(1):46–79.
- Tagliacozzo, Eric and Wen-Chin Chang. 2011. *Chinese Circulations: Capital, Commodities, and Networks in Southeast Asia*. Durham: Duke University Press.
- Trainor, Richard H. 1993. *Black Country Élités: The Exercise of Authority in an Industrialized Area, 1830-1900*. Oxford; New York: Clarendon Press ; Oxford University Press.
- Unsworth, John. 2011. "Computational Work with Very Large Text Collections." *Journal of the Text Encoding Initiative* (Issue 1). Retrieved April 2, 2016 (<https://jtei.revues.org/215>).
- Verbruggen, Christophe. 2007. "Literary Strategy during Flanders' Golden Decades: Combining Social Network Analysis and Propopography." Pp. 579–601 in *Prosopography approaches and applications: a handbook*, edited by K. Keats-Rohan. Oxford: Oxford University Press.
- Wellman, Barry and Stephen D. Berkowitz. 1988. *Social Structures: A Network Approach*. Cambridge; New York: Cambridge University Press.
- Wetherell, Charles. 1998. "Historical Social Network Analysis." *International Review of Social History* 43(Supplement S6):125–144.

- Wu Jen-shu 巫仁恕. 2013. *Youyou Fangxiang : Ming Qing Jiangnan Chengshi de Xiuxian Xiaofei Yu Kongjian Bianqian* 優游坊廂: 明清江南城市的休閒消費與空間變遷 (*Urban Pleasures : Leisure Consumption and Spatial Transformation in Jiangnan Cities during the Ming-Qing Period*). Taipei: the Institute of Modern History, Academia Sinica.
- Xiao-Planes, Xiaohong. 2001. *Education et politique en Chine: le rôle des élites du Jiangsu, 1905-1914*. Paris: Ecole des hautes études en sciences sociales.
- Xu, Xiaoqun. 2001. *Chinese Professionals and the Republican State : The Rise of Professional Associations in Shanghai, 1912-1937*. Cambridge ; New York: Cambridge University Press.
- Yang, KwangSoo and Shashi Shekhar. 2017. *Spatial Network Big Databases Queries and Storage Methods*. Cham: Springer.
- Yang, Mayfair Mei-hui, ed. 1999. *Spaces of their own: women's public sphere in transnational China*. Minneapolis: University of Minnesota Press.