Approaching the Arabic Press of the Late Ottoman Eastern Mediterranean through Digital History

by Till Grallert

The essay explores the use of digital history for the systematic study of the periodical press in the late Ottoman Eastern Mediterranean (1906 – 1918) as a discursive field. It evaluates the methodological and practical challenges of digital history as rooted in the socio-technical infrastructures of the Global North when applied to the Global South. It does so using a case study of four Arabic journals from Baghdad, Beirut, Cairo, and Damascus. After outlining the need for building a corpus and the challenges presented by this effort, this article explores a digital corpus of a total of circa 2.65 million words through (social) network analysis and stylometric authorship attribution.

This essay discusses the challenges and promises of digital history – broadly understood as historiography aided by computational approaches to research questions and based on digitized sources – for the study of historical societies of the Global South. I do so through the history of the Arabic periodical press in the late Ottoman Eastern Mediterranean. The latter is a lose moniker for the predominantly Arabic speaking provinces of the Ottoman Empire along the eastern shore of the Mediterranean between the mountains of Anatolia in the north, Mesopotamia in the east, the deserts of the Arabian Peninsula in the south, and the Libyan Desert in the west, between the mid-nineteenth century and the collapse of Ottoman rule during World War I. I explore the history of the region’s Arabic periodical press through the spatial metaphor of an “ideosphere” as a reference to the realm of human ideas in its entirety, only some of which become manifest and thus traceable in concrete intellectual production. With ideosphere, I argue that one has to transcend individual periodicals and engage in the systematic study of the periodical press as a discursive field and at scale in order to better understand both the intellectual history of the Eastern Mediterranean at a crucial historical juncture and periodical production itself. It is important to note that most of the challenges

of digital history discussed here are in no way limited to the specific case study. Instead, I present particularly pronounced variations of a theme that will ring true for all historians at a moment when the question whether something is digital or not has become increasingly meaningless and when the computational has become hegemonic as everything has already always been mediated through technology.²

Early Arabic periodicals, such as Butrus al-Bustānī’s al-Jīnān (Beirut, 1876–1886), Ya’qūb Ṣarrūf, Fāris Nimr, and Shāhīn Makārīyūs’s al-Muqṭaṭaf (Beirut and Cairo, 1876–1952), Muḥammad Kurd ‘Alī’s al-Muqtabas (Cairo and Damascus, 1906–1918/19) or Rashīd Riḍā’s al-Manār (Cairo, 1898–1941) are at the core of formative discourses that still reverberate through the Arabic-speaking Middle East: the Arabic (cultural) renaissance (nahda), Arab nationalism, and Islamic reform movements. The periodical press of the late Ottoman Eastern Mediterranean has received scholarly attention since the early twentieth century.³ Yet, core questions concerning the intellectual history of the periodical press and the social history of periodical production are still unanswered. Arabic approaches to the periodical press have been, for a long time, both encyclopaedic and anecdotal as well as biased by an almost exclusive focus on Cairo and Beirut. Many of these works were compiled by authors who themselves were journalist and adhere to a specific political view of Arab nationalism and demonize the Ottoman Empire, particularly during the reign of Sultan ‘Abdūlhamīd II (1876–1909).⁴ Non-Arabic scholarship adopted similar geographic and political biases and commonly perceives of periodicals as a source for intellectual, social and political history but only rarely as a subject in its own right. There is the noteworthy exception of Ami Ayalon’s extensive writings on the press in the Arab Middle East,⁵ but comprehensive, synthesising approaches are severely

limited by the almost complete absence of systematic studies on individual periodicals. Even scholarly approaches to the press as a source remain largely anecdotal with a focus on opinion pieces and editorials in a small sample of canonical journals from Beirut and Cairo. The main reasons are scale and scattered collections, which, however, remain unacknowledged. Consider two recent examples: While Florian Zemmin acknowledges that Rafiq al-’Azm published in *al-Hilāl*, *al-Muqtaṭaf*, *al-Ahrām*, and *al-Muqattam*, he does not comment on his selection of four articles from *al-Manār*, which is most likely due to the latter’s availability in digital form. Zemmin also claims that al-’Azm was “a core contributor to al-Manar” without elaborating the criteria for his evaluation. Similarly, Fruma Zachs claims to illustrate a “debate” by picking two articles – one from 1886 and published in Cairo, the other from 1894 and published in Beirut – without commenting on this choice and its implications. What is needed, therefore, are transparency about methods and the actual sources at hand as well as empirically grounded approaches.

In consequence, we still need to answer the question of what are the core nodes (authors, periodicals, other texts) in the ideosphere of the late Ottoman Eastern Mediterranean and how did these networks develop over time? Answers partially depend on another set of open questions: Who authored the majority of articles that did not carry a byline? Can we confirm the common – and untested – assumption that the proprietor or editor-in-chief mentioned in a journal’s imprint authored all the anonymous texts themselves? From the underlying research question also follow larger questions with potentially severe implications for the intellectual history of the Middle East. Is the

---


9 Zemmin, *Validating Secularity in Islam*, p. 76. A computational query on the same digital remediation available to Zemmin reveals that al-’Azm authored only 13 out of more than 4,300 articles.

geographic bias of Cairo and Beirut justified if we look at more than the “easily” accessible handful of monthly journals? How would we need to re-write the intellectual history of the final decades of the Ottoman Empire and the Arabic nahḍa if we included the myriad of papers and their contributors from places as far as Algiers, Basra or Aleppo? This essay computationally explores the question of authorship and references to other periodical titles and the resulting intellectual, social and geographic networks. It presents a first foray into computational approaches to these questions by adopting methods broadly summarised as “distant reading,” namely social network analysis and stylometry for authorship attribution.\footnote{Franco Moretti, Distant Reading, London 2013; Matthew Jockers, Macroanalysis. Digital Methods and Literary History, Urbana 2013.} I start by confronting hyperbolic promises of mass digitization and computational methods as a hegemonic episteme rooted in late twentieth-century, English-speaking capitalism from the margins – that is, the study of a historical multilingual society whose material heritage has been looted, destroyed and neglected; a society, whose written textual heritage resists digitization efforts by being dependent on non-Latin scripts, which cannot reliably be extracted from facsimiles due to the limitations of available OCR technologies; a society, whose contemporary heirs between Mosul, Basra, Aleppo, Homs, the two Tripolis, Ma‘ān, Gaza, and Khartoum cannot draw on the vast resources in wealth and socio-technical infrastructures of the Global North (or Gulf countries).

I argue that the hegemonic digital paradigm of socio-technical infrastructures built upon English and Latin script contributes to a neo-colonial divide between the abundance of digitized cultural artefacts of the Global North and the invisibility of almost anything beyond.\footnote{C. f. Roopika Risam, New Digital Worlds. Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy, Evanston, IL 2019, https://doi.org/10.2307/j.ctv7tq4hg; Paul Gooding, Historic Newspapers in the Digital Age. “Search All About It!,” London 2018, pp. 149 – 157; Nanna Bonde Thylstrup, The Politics of Mass Digitization, Cambridge 2018, pp. 79 – 100.} This forces scholars working on texts in non-Western languages and/or written in non-Latin scripts to engage in substantial corpus-building efforts and severely limits the scope of our computational scrutiny. I, therefore, introduce my own corpus-building project “Open Arabic Periodical Editions” (OpenArabicPE) as a framework to address the outlined challenges to mass digitization of Arabic periodicals by combining transcriptions of a small number of early twentieth-century journals from shadow libraries with digital facsimiles from various vendors for the purpose of validating the former.\footnote{Open Arabic Periodical Editions, https://openarabicpe.github.io. I use “shadow libraries” as proposed by Thylstrup, The Politics of Mass Digitization, p. 81 to describe mass digitization efforts that “operate in the shadows of formal visibility and regulatory systems” and in order to avoid the term “pirate” with its colonial connotations.} Consequently, any corpus built with

these affordances and the dependence on the work of anonymous others will not be systematically tailored to our research questions. But it is currently the only corpus of Arabic periodicals that can be subjected to computational analysis. After substantial modelling efforts, this essay presents the first results of computational analyses of bibliographic datasets of four periodicals published in Baghdad, Beirut, Cairo, and Damascus between 1906 and 1918 and digital full-text editions of three of them with a total of circa 2.65 million words. It is the first systematic attempt to empirically answer core questions for the nascent field of Arab periodical studies, which are in turn indispensable for a proper source critique if one wanted to employ these periodicals for the historiography of the late Ottoman Eastern Mediterranean.

I. The Promised Land of Digitized Arabic Periodicals

The better known and at the time widely popular Arabic journals of the late Ottoman Empire do not face the ultimate danger of their last copy being destroyed in the current onslaught from iconoclasts, institutional neglect, and wars raging through Syria, Libya, Yemen, and Iraq. Yet, copies are scattered across libraries and private collections worldwide. Many collections remain unknown to scholarly communities. If catalogues exist, they are not necessarily available online and union catalogues have fallen out of fashion. An example shall illustrate this point. A search in WorldCat for the nine volumes of al-Muqtabas returns six different bibliographic entries, the first of which has 13 variants (called “editions” by WorldCat), pointing to 34 libraries. If one follows each entry to the holding library’s catalogue, one will find that the large majority of collections is incomplete and that collections commonly combine original volumes, reprints, microfilms, microfiches and even photo copies. This makes it almost impossible to trace discourses across journals and

---

with the demolition and closure of libraries in the Middle East, copies are increasingly accessible to the affluent Western researcher only. Digitization promises an “easy” solution to the problems of preservation and access. Access to tens if not hundreds of thousands of digitized periodical issues frequently invites the imaginaire a promised land of instantaneous one-click answers to any question one might have. The public and many scholars expect to be able to put a computer to such diverse tasks as a keyword search across the ideosphere of the early Arabic press between Morocco and Iraq since its beginnings to track semantic changes; or a social network analysis of the discursive field of authors and their texts and its changes over time. These are highly relevant questions. Unfortunately, the eager student of digitized Arabic periodicals will immediately find tools, data and skills lacking. The first question we encounter in our attempt to track the network of authors and texts is to which extent can we submit digitized periodicals to computational analysis, or rather, what is the meaning of digitized and access? The answer is very different from periodical corpora in Western languages: from large corporate and institution-backed platforms to shadow libraries digitized in the context of Arabic and Ottoman periodicals commonly means the provision of digital facsimiles and, sometimes, even “fakesimiles”: a digital text rendered with a layout meant to emulate the material artefact and served as an image file. While scanning and hosting millions of pages is a laudable endeavor, digital facsimiles provide access only to human readers and solely enable close-reading approaches similar to how we encounter the original artefact or microfilm copies.

15 A map based on the results of this and a similar query to the Arabic Union Catalogue is available at Till Grallert, Map of Geographic Distribution of Library Holdings of al-Muqtabas, https://doi.org/10.5281/zenodo.4154171.
18 The almost hegemonic interface to digitized collections focuses on a Google-like search bar and makes browsing titles – the classic way of accessing periodicals – nigh impossible. The de-contextualizing of strings of text from the page and the wider context of the periodical immanent to keyword search has been repeatedly criticized as inadequate for the study of periodicals; e.g. Laurel Brake, The Longevity of “Ephemerata,” Library Editions of Nineteenth-Century Periodicals and Newspapers, in: Media History 18. 2012, pp. 7 – 20, here p. 17, https://doi.org/10/b7x6ps; Adrian Bingham, The Digitization of Newspaper Archives. Opportunities and Challenges for Historians, in:
Optical character recognition (OCR), the technology to convert an image into machine-readable text, has come a long way and even hand-written text recognition (HTR) is fairly successful at least for Latin script.\(^{19}\) Automatic recognition of Arabic script, however, is severely lagging behind for a variety of reasons beyond the scope of this essay.\(^{20}\) Despite promising developments with the application of machine-learning technologies to pattern recognition,\(^{21}\) automatic conversion of images of early Arabic periodicals is hampered by three factors: first, all OCR technologies depend on training sets of “gold standard” transcriptions as ground truth; second, low-quality fonts, inks, and paper employed at the turn of the twentieth century will inevitably result in poor print quality and reduce the reliability of automatic transcription through variance; and third, text recognition depends on layout recognition and multi-column texts with various intersections of boilerplate, ads, et cetera pose serious challenges. Consequently, these texts can currently only be reliably digitized by human transcription.\(^{22}\) Funds for transcribing the tens to hundreds of thousands of pages of an average mundane periodical are simply not available, despite of their cultural significance and unlike what is being done with valuable manuscripts and high-brow literature.

Some platforms either ignore the problem, actively pretend it doesn’t exist, or claim to have solved it and therefore foreground search functions. Reasons range from Arabic being a marginal language in their corpus to the need to return a profit on investment through selling extremely expensive subscrip-
tions to institutions. The advertised search functions are severely limited and often deceptive. Hathitrust, a Google-powered conglomerate of mostly American universities, is obviously dysfunctional for Arabic text if one has a look at the text layer. The commercial “Early Arabic Printed Books” (EAPB) project was developed by Cengage Gale in collaboration with the British Library and makes repeated claims of employing “[n]ewly-developed optical character recognition software (OCR) for early Arabic printed script.” But since they share neither text layers nor error rates or software, their claims cannot be verified. East View’s digital Arabic periodical collections (some free to use, some subscription-based) cover the middle ground. While they rely on extremely messy OCR data, which they invite their users/customers to manually improve, the interface focuses on the search box. As long as platforms show search results as highlights superimposed upon the facsimile, one can catch (the many) false positives. We will, however, never know the extent of false negatives.

The digital text of a periodical is necessary but not sufficient for many analytical queries and distant reading and it is certainly insufficient for close reading. The full text of a periodical would be nothing but a string of words. But periodicals unite different texts of various genres from multiple authors. These texts are commonly grouped into issues and volumes. Longer ones are frequently serialized and scattered across issues. Some of these texts are reprints from other periodicals or first printed editions of much older manuscripts. Therefore the full text has to be modeled in order to make sense for both humans and machines. Will Hanley’s effort of modeling the OCR’ed issues of the newspaper Egyptian Gazette (Alexandria, 1905 – 1908) with the help of his students as part of a digital micro-history course since 2016 demonstrates how tedious this work is even with an English baseline. Even the provision of digitized facsimiles and raw OCR output does not mean access for all and to everything. Digital infrastructures, despite all promises towards the opposite, are rooted in the hegemony of late twentieth-century

---


capitalism and the Global North. Most digitized periodicals are kept in opaque data silos. In the absence of provisions for interchange or interoperability in the form of application programming interfaces (APIs) or the option to bulk download data in standardized, open file formats, access to these silos is restricted to human readers and provided through proprietary web interfaces that are commonly neither tailored to the display of Arabic material nor themselves available in Arabic.\textsuperscript{28} Such (close) reading access is further restricted by paywalls, licenses, and geo-fencing.\textsuperscript{29} Downloading content in order to circumvent ill-suited interfaces is limited to individually identifiable users. Bulk download frequently violates terms of use and most vendors try to prevent this on the technical level.

To computationally answer the above questions, however, one would need unrestricted access to truly digital editions – that is, machine-readable editions of the full text with embedded structural and semantic information and in a standardized exchange format.\textsuperscript{30}

In the absence of digital editions, any meaningful computational analysis of the connections between authors, texts, and periodicals as a venue for publication and review requires access to reliable standardized bibliographic metadata as a bare minimum. Unfortunately, even this data is practically non-existent. This is due to a combination of factors: first, ambiguity and incorrect data found in the original artefact; second, lacking familiarity with the particularities of these artefacts among cataloguers, librarians and scholars; and third, a software stack ill-suited for anything but Western concepts of dates and names and Western scripts.

Periodicals seem to provide no dating challenges as publication dates were conveniently recorded in a masthead. However, periodicals across the Arabic-speaking late Ottoman Eastern Mediterranean made use of at least four calendars. Newspapers and journals provided dates in any combination of the


\textsuperscript{29} The US-based HathiTrust, for instance, does not provide public or open access to its collections, even to materials in the public domain under extremely strict US copyright laws, when users try to access them from outside the US. On the issue of unequal access to digitized collections see Gooding, Historic Newspapers in the Digital Age, pp. 145 – 170, and especially his figure 6.1 showing global access (or lack thereof) to the British Library Nineteenth-Century Newspapers.

Ottoman fiscal, or mālī calendar and the reformed Julian calendar as well as the better known Islamic hijrī and Gregorian calendars. In addition to at least three different year counts, these calendars and their users also differed in their conception of the calendric day. Most retained the old notion of a day commencing at sundown, while others adopted alla franca time with 24 equinoctial hours and a date change at midnight.31 Unfortunately supplied dates from mastheads frequently neither matched each other nor the day of the week the paper was supposedly printed on.32 How should one record this bibliographic nightmare? And which date-calendar combination should be considered the authoritative one? What if recorded publication dates were fictional to simulate a regular publication cycle and should therefore be conceived of as issue numbers that have only limited relation to an actual date?33

Any attempt to answer these questions relies on the affordances of available information systems, that is people and their skills, abstract concepts, and actual tools to record and retrieve these data points. But cataloguers, librarians and even specialists of the late Ottoman Eastern Mediterranean are frequently unfamiliar with calendric systems beyond the solar Gregorian and the lunar Islamic hijrī calendars. Mālī years are frequently misread as hijrī years, which introduces a margin of error of up to two years for the last decades before World War I.34 Moreover, most software is unable to work with anything but


33 Al-Muqtabas, for instance, was severely lagging behind its publication schedule by summer 1909. Volume 4, issue 7 was scheduled for Rajab 1327 A H (July / August 1909) according to its masthead, but only published in the first week of April the following year; see Jarıdat al-Muqtabas, 7. 4. 1910, p. 3.

34 Both Stefan Weber and Jens Hanssen, for instance, missed the fact that the birthday of Sultan ʿAbdūlhamīd II (1876 – 1909) was celebrated according to the Islamic hijrī calendar and thus rotated through the solar year. The annual celebrations of the anniversary of ʿAbdūlhamīd II’s accession to the throne were celebrated according to the empire’s mālī calendar. Yet, leading scholars read these dates as pertaining to the hijrī calendar. Due to a mix-up in 1872 and the resulting growing difference between the two calendars, ʿAbdūlhamīd II’s silver jubilee on the throne is wrongly dated to 1901 instead of 1900. Jens Hanssen, Fin de Siècle Beirut. The Making of an Ottoman Provincial Capital, Oxford 2005, p. 238 and p. 246; Stefan Weber, Damascus. Ottoman Modernity
Gregorian dates out of the box. Even if cataloguers were able to correctly establish the calendar used in a periodical’s masthead, the computing infrastructure would not allow them to enter this date into the digital record.\textsuperscript{35} Finally, bibliographic data is not commonly shared in a standard-compliant and machine-actionable format even when it is internally kept in structured form.\textsuperscript{36}

A good example for this state of affairs is the British Library’s otherwise excellent Endangered Archives Programme (EAP), which digitized periodical holdings of the al-Aqsa Mosque’s library in Jerusalem (EAP119).\textsuperscript{37} If we look at the fourth volume of the journal \textit{al-Muqtabas} available through EAP, we find that bibliographic information is solely provided in unstructured plain text.\textsuperscript{38} Publication dates are provided as Gregorian months even though the cover clearly states that \textit{al-Muqtabas} follows the “Arabic,” i.e. Islamic \textit{hijrī} calendar and despite each issue reporting the publication date as \textit{hijrī} month. Consequently, there is a dissonance between the facsimile and the bibliographic information. The first issue of \textit{al-Muqtabas}’ fourth volume recorded...
the month of Muḥarram 1327 A.H in its masthead. Depending on the local observation of the moon in Damascus, the journal’s place of publication, this month began around 27 January 1909. Should this issue therefore be considered the January or the February issue? The cataloguers at EAP clearly thought the latter or their cataloguing software did not allow for date ranges.

Even if we had perfectly reliable digital re-mediations of the bibliographic information found in the periodical issues themselves, the vast majority of articles would remain outside our analytical scopes because publishers did not provide (meaningful) bylines – most articles in journals and newspapers from Baghdad, Beirut, Cairo or Damascus did not credit their authors. One approach is to subject all articles to stylometric analysis for authorship attribution (more on this below) but this again presupposes truly digital editions.

II. Building a Corpus of Digital Arabic Periodicals

This state of digitized Arabic periodicals puts the onus of building a digital corpus on us – scholars of the late Ottoman ideosphere interested in leveraging computational approaches to answer pressing questions of the field. Like many others, I turned to the shadow libraries of Arabic literature, such as al-Maktaba al-Shāmila, Mishkāt, Ṣayyid al-Fawāʾid or al-Waraq. They provide access to (mostly classical) Arabic texts including transcriptions of unknown provenance, editorial principles, and quality for a small number of periodicals. These informal “editions” lack information linking the digital representation to the original artefact, namely bibliographic metadata and page breaks, which makes them almost impossible to validate and therefore employ for scholarly research.

Since we do not have the resources to proof and correct these texts, I conceived of Open Arabic Periodical Editions (OpenArabicPE) as a framework for open, collaborative, and fully-referencable scholarly digital editions of early Arabic periodicals. OpenArabicPE addresses the above-mentioned issues of existing collections of digitized Arabic periodicals with an emphasis on accessibility, sustainability, and credibility. It builds on the simple idea of combining the virtues of immensely popular, but non-academic shadow libraries with academic and commercial scanning efforts as well as editorial expertise. Starting with the mostly Damascene periodicals al-Muqtabas and al-Haqqāʾiq, we devised workflows and tools to transform digital texts (badly formatted HTML) from al-Maktaba al-Shāmila into an open, standardized file format

39 Al-Maktaba al-Shāmila; Mishkāt; Ṣayyid al-Fawāʾid; al-Waraq.
(XML) based on the Text Encoding Initiative (TEI)’s guidelines, to generate bibliographic metadata, and to render a parallel display of text and facsimile in a web browser. We model the periodicals through adding structural mark-up for articles, sections, authors, and bibliographic metadata. Our schema to do so also addresses the problems outlined above and devised ways how to encode and – as far as possible – computationally normalize non-Gregorian dates and Arabic-Ottoman entity names. Finally, we link each page to facsimiles from various sources, namely EAP, HathiTrust, and Arşif al-majallat al-adabiyya wa-l-thaqāfiyya al-‘arabiyya. The latter step, in the process of which we also make first corrections to the transcription, although trivial, is the most labor-intensive because page breaks were commonly ignored by al-Maktaba al-Shāmila’s anonymous transcribers. Each of the circa 8,500 page breaks in al-Muqtabas and al-Ḥaqāʾiq needed to be manually marked by volunteers in order to link facsimiles to the digital text and thus make the text verifiable for human readers. So far Dimitar Dragnev, Talha Güzel, Dilan Hatun, Hans Magne Jaatun, Jakob Koppermann, Xaver Kretzschmar, Daniel Lloyd, Klara Mayer, Tobias Sick, Manzi Tanna-Händel and Layla Youssef have contributed their time to this task.

All tools and the editions are hosted on the code-sharing platform GitHub under the most permissive licenses for reading, contribution, and re-use. Retaining copyright of our own editorial contributions in the form of Creative Commons Attribution-ShareAlike 4.0 International is a reminder that the enormous amount of, often contingent, labor embodied in digital resources need to be transparently credited. We also provide structured bibliographic metadata for every article in machine-readable formats that can easily be integrated into larger bibliographic information systems. This bibliographic data is also accessible through a constantly updated public Zotero group, which can serve as a port of entry to the editions. With OpenArabicPE, I argue that by linking facsimiles to the digital text, every reader can validate the quality of the transcription against the original. We thus remove the greatest limitation of crowd-sourced or informal transcriptions

43 In other instances, such as the journals Lughat al-‘Arab, https://github.com/OpenArabicPE/journal_lughat-al-arab, and al-Usṭādh, https://github.com/OpenArabicPE/journal_al-ustadh, al-Maktaba al-Shāmila did provide page breaks that correspond to a printed edition.
45 Creative Commons, Attribution-ShareAlike 4.0 International, https://creativecommons.org/licenses/by-sa/4.0/.
and the main source of disciplinary contempt among historians and scholars of the Middle East. Anyone can improve the transcription as well as our modeling of a journal’s content with clear attribution of authorship and version control using .git and GitHub’s core functionality.47

1. OpenArabicPE’s Corpus

The resulting corpus comprises the full text of each issue of *Lughat al-‘Arab*, *al-Muqtabas* and *al-Haqa‘iq* until the end of World War I and a transcription of article titles and bylines for one volume of *al-Hasnā‘*, totalling 165 full-text journal issues with some 2.65 million words (table 1).48 Titles were selected based on the state of the digital editions and for their geographic distribution. This corpus is small if compared to the vast data sets available for the Global North through Chronicling America, Trove Australasia, the British Newspaper Archive et cetera, which gave rise to numerous distant reading projects.49 However, it is the only corpus of this material. Taken with a grain of salt, a systematic analysis of this corpus helps us test common hypotheses, challenge established narratives about the Arabic periodical press and direct the focus of further scrutiny, as I will show in the following sections after briefly introducing the constituent periodicals.

47 Such an approach was proposed by Christian Wittern, Beyond TEI. Returning the Text to the Reader, in: Journal of the Text Encoding Initiative 4. 2013, Selected Papers from the 2011 TEI Conference, http://jtei.revues.org/691. It has recently seen a number of concurrent practical implementations such as project GITenberg, https://gitenberg.git-hub.io, led by Seth Woodworth, or Jonathan Reeve’s Git-lit, https://github.com/Git-Lit/git-lit.


Muḥammad Kurd ‘Ali (1876 – 1953) established his journal *al-Muqtabas* (The Digest) in Cairo in 1906 and the first daily newspaper to be published in Damascus in 1908 (also confusingly called *al-Muqtabas*). He was the best known and, after the Young Turk Revolution of 1908, the most influential journalist and intellectual in Damascus. Before running his own periodicals, he had held minor government offices and worked at various public and private presses and periodicals in Damascus and Cairo. He was well-acquainted with leading figures of the Islamic reform movement in Egypt and Greater Syria. He was a member in Ṣaḥīḥ al-Jaza‘irī’s “senior circle” in the early 1890s in Damascus and later moved and worked in Rashīd Rıdā’s and Muḥammad ‘Abduh’s circles in Cairo. After the Young Turk Revolution, Kurd ‘Ali returned to his hometown and the publication of *al-Muqtabas* moved from Cairo to Damascus in the journal’s third year. In Damascus, *al-Muqtabas* soon became “the boldest, most coherent, consistent and committed proponent of reform and modernity […] prior to World War I.”

Due to conflicts with the authorities over the reprint of a poem, Kurd ‘Ali again fled Damascus for Cairo and Europe in 1912. Consequently, *al-Muqtabas* was published from Cairo for a couple of months before Kurd ‘Ali was allowed to return once again. During World War I and Cemal Pasha’s infamous term as commander-in-chief of the Fourth Army and governor general of Syria, Kurd ‘Alī was able to win his support. He thus escaped the fate of Shukrī al-‘Asālī, his co-editor at the newspaper *al-Muqtabas*, ‘Abd al-Ghānī al-‘Urayṣī and other journalists from Beirut and Damascus, who were publicly executed on charges of treason. Like their editor, the journal and the newspaper *al-Muqtabas* survived and continued publication until the final days of the war – albeit in shorter and less

Table 1: Summary of our periodical corpus.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Place</th>
<th>Dates</th>
<th>Vols.</th>
<th>Nos.</th>
<th>Arts.</th>
<th>With Authors in %</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Al-Hāqa‘iq</em></td>
<td>Damascus</td>
<td>1910 – 13</td>
<td>3</td>
<td>35</td>
<td>389</td>
<td>163</td>
<td>41.90</td>
</tr>
<tr>
<td><em>Al-Hasnā‘</em></td>
<td>Beirut</td>
<td>1909 – 10</td>
<td>1</td>
<td>11</td>
<td>173</td>
<td>63</td>
<td>36.42</td>
</tr>
<tr>
<td><em>Al-Muqtabas</em></td>
<td>Cairo, Damascus</td>
<td>1906 – 17/18</td>
<td>9</td>
<td>96</td>
<td>2,964</td>
<td>377</td>
<td>12.72</td>
</tr>
<tr>
<td><em>Lughatal-‘Arab</em></td>
<td>Baghdad</td>
<td>1911 – 14</td>
<td>3</td>
<td>34</td>
<td>939</td>
<td>152</td>
<td>16.18</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>16</td>
<td>176</td>
<td>4,465</td>
<td>755</td>
<td></td>
</tr>
</tbody>
</table>

frequent editions due to material shortages. After the end of the war and the
disintegration of the Ottoman Empire, Kurd ‘Ali abandoned the monthly and
left the editorship of the revived daily newspaper al-Muqtamas to his brother
Aḥmad. He founded the Arab Scientific Academy whose president he became
in 1919 and served twice as Minister of Education (1920 – 1922 and
1928 – 1932) during the French Mandate over Syria.51

Much less is known about the second Damascene journal in our corpus and the
people behind it. Al-Ḥaqāʾiq (The Facts) was a periodical of the conservative
Muslim establishment, who called themselves mutadayyīnūn (the very pious).
A total of three volumes with 35 issues were published between 1910 and 1913
by the ʿālim (religious scholar) ‘Abd al-Qādir al-Iskandarānī (1875 – 1943). The
journal had a strong focus on Islamic topics and published constant critiques
of Salafism and of what they saw as westernized journals, such as al-
Muqtamas.52 Its page layout conveys an apparent lack of familiarity with the by
then established conventions of Arabic journals, which underlines its nature as
an outlier in the ideosphere of Arabic periodicals. Al-Iskandarānī came from
the notable al-Kaylānī family, who were closely associated with the Qadriyya
Sufi order. Unlike Kurd ‘Ali, he did not claim a single byline in “his” journal or
any other periodical in our corpus. Pages contain much less text and issues are
much shorter than those of al-Muqtamas.

The Carmelite Father Anasta¯s Ma¯rı¯ al-Karmalı¯ (born Buṭrus ‘Awwād,
1866 – 1947) from Mount Lebanon established the monthly journal Lughat
al-˘Arab (The Language of the Arabs) in Baghdad in 1911. The journal
published 34 issues until its final number in June 1914. Al-Karmalī was exiled
to Kayseri in Anatolia upon the beginning of World War I and returned in
summer 1916. Lughat al-˘Arab recommenced publication only in 1926.53 It is
not clear to which extent al-Karmalī was involved as editor. Only a (small)
handful of articles in Lughat al-˘Arab carried his byline. Kāzım al-Dujaylı¯
(1884 – 1970), a self-taught journalist and a poet and student of the Baghdadi
Salafist Shukrı¯ al-Alu¯sı¯ and al-Karmalī, joined Lughat al-˘Arab as editor(-in-
chief?) soon after. Issues were similar in length to Ḥaqāʾiq.

51 For an autobiographic sketch see Muhammad Kurd ‘Ali, Khitāt al-Shām, vol. 6,
Damascus 1928, pp. 411 – 425, http://www.archive.org/details/kutat_cham_06. For in-
tellectual biographies see Seikaly, Damascene Intellectual Life; Rainer Hermann,
Kulturkrise und konservative Erneuerung. Muhammad Kurd ‘Ali (1876 – 1953) und das

52 For controversies between al-Muqtamas and al-Ḥaqāʾiq see James L. Gelvin, “Moder-
nity,” “Tradition,” and the Battleground of Gender in Early 20th-Century Damascus, in:
Die Welt des Islams 52. 2012, pp. 1 – 22, https://doi.org/10/ggwwhd; David Commins,
Islamic Reform. Politics and Social Change in Late Ottoman Syria, Oxford 1990,
pp. 118 – 122.

github.io/journal_lughat-al-arab/tei/oclc_472450345-i_35.TEIP5.xml#div_2.d2e316;
The monthly journal *al-Hasnāʾ* (The Fair Lady), published by Jirjī Niqūlā Bāz in Beirut between 1909 and 1911, was the first women’s magazine in Greater Syria. Together with Marī ‘Ajami’s *al-‘Arūs* (The Bride, Damascus 1910–unknown) and Salīma Abū Rashīd’s *Fatāṭ Lūbān* (Girls of Lebanon, Beirut 1914–unknown) and even though it was published by a man, *al-Hasnāʾ* was part of a second wave of “politically engaged female writing.”

III. Network of Authors and Texts

The quality and significance of the analysis of bibliographic data is directly dependent on the quality of the information provided by the periodicals themselves and of our mark-up in the digital editions. All relevant personal and place names in bylines and other source information must be marked up for retrieval. A core step is the necessary disambiguation of named entities through local and external authority files: “Anastās al-Karmalī,” “Buṭrus ‘Awwad,” “Sāṭīnsāʾ” and “The publisher of *Lughat al-‘Arab,*** for example, refer to the same person, “Ḥalab” and “al-Shabhā” both designated the city of Aleppo. By linking references to external authority files and the semantic web, we can harvest additional information on authors and locations, namely the geolocation for toponyms, transcriptions into Latin script necessary for working with most visualization tools, and life dates for persons. It must be noted that, as in any other step along the process of this research project, scholars of non-Western societies are at a significant disadvantage. Automated named-entity recognition (NER) is well established for western languages but there is currently no readily available software that supports Arabic. Openly available authority files, such as the Virtual International Authority File (VIAF) that aggregates tens of national authority files, which, in turn, depend on expensive infrastructures, are subject to the same digital affordances alluded to above (English interface, names as combinations of fore- and surnames, Gregorian calendar). As a result, they are heavily biased towards the Global North.


56 Virtual International Authority File (VIAF), https://viaf.org. There is a small number of efforts to build historical gazetteers for the Eastern Mediterranean. The most noteworthy is Masoumeh Seydi and Maxim Romanov, Al-Ṭurayyā Project, https://althurayya.github.io/. The Digital Ottoman Platform at Princeton University, in which I participated, aimed at building a digital gazetteer of the Ottoman Empire but all faltered after summer 2016.
1. Evaluating the Corpus. Network of Referenced Periodicals

Knowing that we work with a corpus whose composition is the result of external and unknown decisions by the contributors to al-Maktaba al-Shāmila as to which periodical to transcribe, we can evaluate the performance of this corpus in representing the larger ideosphere of the periodical press in the late Ottoman Eastern Mediterranean by looking at the network of referenced periodicals. Explicit references to periodicals indicated by “jarīda XYZ” or “majalla ABC” were automatically marked-up using XSLT and regular expressions and linked to local and external authority files for disambiguation and additional bibliographic information. I then counted the references to each mentioned periodical and plotted the result as a network graph. The plots feature the number of references by issue to account for the varying length of articles in each journal. Each node in the network plot (figure 1) signifies a periodical. Edges (connections) are drawn between nodes (periodical titles) when one references the other. The thickness of the edges indicates the number of issues that reference a periodical (weight). The size and shade of nodes reflect the number of journals in our corpus that mention this periodical (in-degree).

The first observation, common to all social networks, is that only a very small number of nodes are of relative importance, as measured by in-degree (number of edges connecting to a node) and weight of the edges connecting nodes. Out of a total of 465 different periodical titles, 421 or circa 90% were referred to by only a single journal. 344 periodicals are only mentioned in a single issue and 335 in a single article. The core of the network in figure 1 comprises only 44 periodicals mentioned by more than one journal. Only nine of those (or 2.13% of all periodicals) were referenced by three journals in our corpus. They are: al-Manār, al-Muqṭatāf, al-Hilāl and al-Diyyā from Cairo, al-Muqṭabas itself, al-Mufīd, al-Watān and al-Ḥaqīqa from Beirut and al-Ḥuqqūq from Mount Lebanon. The centrality of the three Cairene periodicals, al-Manār, al-Muqṭatāf, al-Hilāl, which were all published by Syrian immigrants, tentatively confirms standard narratives of the Arabic press.57 The remaining six, however, do not figure prominently in scholarly literature. If we had the means to construct our own corpus without the severe limitations alluded to above, these would be the journals to digitize. Second, this network is highly centralized in terms of geographic distribution. The thirty core nodes with identified publication places were published in only a handful of locations: Beirut (9), Cairo (7), Baghdad, Damascus, Paris (3), Alexandria, London, Mount Lebanon, Saida and Zahle (1).

A third observation of the larger network is that al-Muqṭabas accounts for the vast majority of references to other periodicals by some orders of magnitude even after we account for al-Muqṭabas having almost thrice as many issues as either al-Ḥaqīqa or Lughat al-ʿArab (table 1). If we assume that we haven’t

missed a significant number of references, then *al-Muqtabas* was more outward-looking and more involved in larger discourses of the day. Fourth, a closer look at the core nodes in the network reveals that all periodicals were primarily self-referential – indicated by the thickest edges connecting a journal to itself (for the purpose of this visualization and to prevent circular edges, source and target nodes were separated). Fifth, the core nodes include number of surprises: *al-Jina¯n* was published by Butrus al-Busta¯nı¯ and later his son Salı ¯m al-Bуста̀nı¯ in Beirut between 1876 and 1886. This means that either *al-Jina¯n* was still relevant for certain discourses long after it ceased publication or that the corpus, spanning the years 1906 until 1918, contains a number of historiographic texts mentioning important journals of the past.\(^{58}\) Articles in Ibrāhīm al-Yazījī’s *al-

---

\(^{58}\) It is highly unlikely that these are references to another *al-Jina¯n* as such a publication is not mentioned in *Filîb dı ¯ Taarrāzı*, Tārīkh al-ṣihāfa al-‘Arabiyya, vol. 4, Beirut 1933, http://hdl.handle.net/2333.1/crjdfsp1, which has been digitized as part of Grallert, Jarā’id.
Diya‘, published in Cairo between 1898 and 1906, were also referenced after the end of this journal. Finally, the group of periodicals mentioned in more than one journal in our corpus comprises a number of foreign titles such as Le Temps, Revue des Revues and Revue du Monde Musulman from Paris and The Times from London.

2. Analysis of Metadata. Network of Authors

Sketching a network of periodicals and the references between them is only one part in the endeavor to layout the ideosphere of the late Ottoman Eastern Mediterranean. Another is the network of authors who published in these periodicals and the geographic distribution of places they wrote from. Knowing the importance of certain authors for an individual periodical is the basis for mapping the network of authors across the late Ottoman ideosphere.

The aim would be to map a network for the hundreds of journals and newspapers published between Alexandria and Aleppo, Jaffa and Basra. Such a network analysis, nevertheless, already provides valuable insights with our small corpus of bibliographic metadata. We can currently identify a total of 319 named authors within this sample: 139 for al-Muqtasab, 103 for al-Haqā‘iq, 52 for Lughat al-‘Arab, and 42 for al-Ḥasnā‘. Quite a significant number appear only with their initials, particularly in al-Ḥaqā‘iq, and – with the exception of al-Ḥasnā‘ – all of them were men.

Figure 2: Network of authors with bylines in al-Haqā‘iq (left), al-Ḥasnā‘ (right), Lughat al-‘Arab (top), and al-Muqtasab (bottom). Size and shade of nodes indicate the number of journals in which an author had bylines.

The first observation, again, is that only a very small number of nodes (14 of 319) are of relative importance as measured in-degree (number of edges connecting to a node) and weight of the edges. In the network plot (figure 2), edges were drawn between all authors who published in the same periodical. Shades and size of nodes signify the out-degree or the number of journals in our corpus in which an author had bylines. The thickness of the edges is a function of the number of articles carrying the byline of a given author. Nodes of authors who published only in a single journal form dense clusters. These are: al-Ḥāqaʿīq to the left, al-Muqtabas bottom centre, Lughat al-ʿArab top centre, and al-Ḥasnāʾ to the right.

A closer look at the central nodes of the network reveals that only one author published in all four journals: Maʿrūf al-Ruṣāfī was a famous poet from Baghdad who mostly authored qaṣīdas (poems) on current political affairs. He moved to Istanbul after the Young Turk Revolution, where he worked as an Arabic teacher at the Royal College and at the newspaper Sabil al-Rashad. He was elected to the Ottoman parliament as representative for al-Muthanna (Iraq) in 1912 and 1914. After World War I, he became a member of the Arab Scientific Academy, established by Muḥammad Kurd ‘Ali in Damascus. 59 Al-Ruṣāfī’s close ties to al-Muqtabas and Kurd ‘Ali are further evident in the announcement for the publication of a first collection (dīwān) of his poems in 1910. There, al-Muqtabas claimed that al-Ruṣāfī was known as “the poet of al-Muqtabas” and – wrongly – that “more than three quarters [of the qaṣīdas therein] had been published in this journal.” 60 The publication of al-Ruṣāfī’s qaṣīdas in so many different periodicals raises important questions regarding the production of periodicals: Did al-Ruṣāfī send his qaṣīdas to the editors of sometimes far-away periodicals unsolicited? Was he invited to contribute? Did editors take his texts from other sources such as the dīwān? 61

---


The other 13 central nodes had bylines in only two out of four journals. Only eight of the 14 authors can be found in international authority files, which at least means that they have not authored works catalogued in any of the contributing libraries (see table 2). Those for whom we have biographic information (employing more traditional close reading of Arabic prosopographic literature) were on average in their mid-thirties during the years under investigation.\(^{62}\) There is a surprising number of Iraqis and a notable absence of Syrians from this network of two Damascene journals and one periodical from Beirut and Baghdad each. Among the eleven identifiable authors, there are six Iraqis: Maʿrūf al-Ruṣāfī, Kāzīm al-Dujaylī, Iḥbīm Ḥīlmī al-ʿAmr, Anastās Mārī al-Karmalī (often writing under the pen name Sātīnā), and the two brothers Muḥammad Riḍā al-Shabībī and Muḥammad Bāqīr al-Shabībī; three Egyptians: Muṣṭafā Ṣādiq al-Raʿīfī, Aḥmad Muḥarram and Wālī al-Dīn Yakan; and only two Syrians: ʿĪsā Iskandar al-Maʿlūf and Muḥammad Rāghib Ṭabbākh. One would expect Syrians to figure much more prominently since the vast majority of articles was actually published in Damascus. In addition, Iraqis are much less prominent in the scholarly literature on the Arab renaissance (nahdā) as one would expect from looking at this network. The religious composition of this core group raises the same issue: The ratio of two Christians among a group of 14 is a mirror of the larger populations but surprising if one considers the emphasis on the importance of Christians for the nahdā in scholarly literature. It is also important to note that Ayalon, in his quasi-standard account of the Arabic press, mentions only one of the 14, ʿĪsā Iskandar al-Maʿlūf, and only in passing.\(^{63}\)

In terms of education and occupations the core nodes are exemplary for the bourgeois middle-class intelligentsia of their time: many attended Ottoman state schools in addition to more traditional, religious venues of education; many knew foreign languages in addition to Arabic and Ottoman; some were trained or even taught abroad in the colonial centres of Paris and London; some served in the Ottoman bureaucracy; some were educators. There is also a significant number of poets (7) among the central nodes\(^{64}\) and a small number of representatives of the Ottoman parliament. The more prolific of them were themselves journalists who at one time or another operated their own periodical(s): Maʿrūf al-Ruṣāfī, ʿĪsā Iskandar al-Maʿlūf, Iḥbīm Ḥīlmī al-ʿAmr, Muḥammad Bāqīr al-Shabībī, Kāzīm al-Dujaylī and Anastās Mārī al-Karmalī, the publishers of Lughat al-ʿArab. Looking at the latter, the importance of al-


\(^{63}\) Ayalon, The Press in the Arab Middle East, p. 53 and p. 219; Albert Hourani, Arabic Thought in the Liberal Age 1798 – 1939, Cambridge 1983 similarly does not mention any of them.

\(^{64}\) Al-Ruṣāfī, al-Dujaylī, the two al-Shabībī brothers, al-Rāfī, Muḥarram, and Yakan.
Muqtabas in this small network (and beyond) cannot be overstated: al-Karmalî had more bylines in al-Muqtabas than in his own journal Lughat al-Arab. The connection to Damascus and Muḥammad Kurd ‘Alî is further evident in four of the central nodes becoming members of the Arab Scientific Academy in Damascus after World War I, whose first president was Kurd ‘Alî himself.\footnote{Al-Ruşāfī, al-Dujaylī, al-Maʃūf, and Muḥammad Riḍā al-Shābībī.}

Another striking observation can be found in the proximity and overlap of clusters. Two of the journals in our corpus, al-Muqtabas and al-Ḥaqqāʾiq, were predominantly published in the same city but there is only very limited overlap. Their clusters are only loosely connected by a handful of people who have only one or two bylines in each journal. The ties between Cairo, Damascus and Baghdad as signified by al-Muqtabas and Lughat al-ʿArab, on the other hand, being much closer. The number of shared authors is only marginally bigger but they had many more bylines to their names. This contradicts the assumption that due to the very small size of local journalistic circles – in 1912, five monthly journals were published in Damascus and none had a print run of

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>VIAF ID</th>
<th>Birth</th>
<th>Death</th>
<th>Journals</th>
<th>Articles</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maʿrūf al-Ruşāfī</td>
<td>14924300</td>
<td>1875</td>
<td>1945</td>
<td>4</td>
<td>31</td>
<td>15,038</td>
</tr>
<tr>
<td>2</td>
<td>Kāzîm al-Dujalî</td>
<td>1884</td>
<td>1970</td>
<td>2</td>
<td>28</td>
<td>38,050</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Ibrâhîm Hîlimî al-ʿAmr</td>
<td>1890</td>
<td>1942</td>
<td>2</td>
<td>22</td>
<td>40,747</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ʿĪsâ Iskandar al-Maʃūf</td>
<td>40250618</td>
<td>1869</td>
<td>1956</td>
<td>2</td>
<td>22</td>
<td>23,383</td>
</tr>
<tr>
<td>5</td>
<td>Sāṭîsnâ</td>
<td>39370998</td>
<td>1866</td>
<td>1947</td>
<td>2</td>
<td>15</td>
<td>20,106</td>
</tr>
<tr>
<td>6</td>
<td>Muḥammad Riḍâ al-Shābībî</td>
<td>22006374</td>
<td>1889</td>
<td>1965</td>
<td>2</td>
<td>12</td>
<td>24,156</td>
</tr>
<tr>
<td>7</td>
<td>Muḥammad al-Ḥâshîmî</td>
<td>2</td>
<td>6</td>
<td>2,717</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Muṣṭafâ Sâdiq al-Râfîî</td>
<td>236524859</td>
<td>1880</td>
<td>1937</td>
<td>2</td>
<td>6</td>
<td>3,711</td>
</tr>
<tr>
<td>9</td>
<td>Muḥammad Bâqîr al-Shābībî</td>
<td>1889</td>
<td>1960</td>
<td>2</td>
<td>5</td>
<td>3,331</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Abû al-Dîyâ</td>
<td>2</td>
<td>4</td>
<td>5,836</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Aḥmad Muḥarram</td>
<td>60500457</td>
<td>1877</td>
<td>1945</td>
<td>2</td>
<td>4</td>
<td>1,543</td>
</tr>
<tr>
<td>12</td>
<td>ʿA. J.</td>
<td>2</td>
<td>3</td>
<td>639</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Muḥammad Râghib Tâbbâkî</td>
<td>63117968</td>
<td>1877</td>
<td>1951</td>
<td>2</td>
<td>3</td>
<td>2,633</td>
</tr>
<tr>
<td>14</td>
<td>Wâlî al-Dîn Yâkân</td>
<td>36771043</td>
<td>1873</td>
<td>1921</td>
<td>2</td>
<td>2</td>
<td>795</td>
</tr>
</tbody>
</table>

Table 2: Authors who published in more than one journal in our corpus comprising al-Ḥaqqāʾiq, al-Ḥasnāʾ, Lughat al-Arab, and al-Muqtabas.
more than a few hundred copies \(^{66}\) there would be a substantial overlap in authorship between periodicals from the same provincial city.

**Individual periodicals**

The work on compiling the biographies of all 319 currently identifiable contributors is far from being done, but after looking at the most productive authors for each journal, we can identify certain trends in the author populations and their geographic distributions. For the purpose of this essay, I will contrast *al-Muqtabas* and *al-Ḥaqāʾiq*, the two Damascene periodicals in our corpus.

Only fifty authors published more than one article in *al-Muqtabas*. Two of the five most prolific authors with more than ten bylines to their names wrote from Baghdad (table 3): Maʿrūf al-Ruṣāfī and Anastās Mārī al-Karmālī. Īsā Iskandar al-Maʾlūf wrote mostly from Zahle and Yūsuf Jirjis Zakham from Omaha and Lincoln, Nebraska, USA. Only the fifth most prolific author was a native resident of Damascus: Muḥammad Kurd ‘Alī himself.\(^{67}\)

Table 3: The 15 most prolific authors in *al-Muqtabas* by number of articles (7 or more).

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>VIAF ID</th>
<th>Birth</th>
<th>Death</th>
<th>Articles</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N. A.</td>
<td>2,592</td>
<td></td>
<td></td>
<td>1243773</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Maʿrūf al-Ruṣāfī</td>
<td>14924300</td>
<td>1875</td>
<td>1945</td>
<td>27</td>
<td>58230</td>
</tr>
<tr>
<td>3</td>
<td>Īsā Iskandar al-Maʾlūf</td>
<td>40250618</td>
<td>1869</td>
<td>1956</td>
<td>20</td>
<td>105688</td>
</tr>
<tr>
<td>4</td>
<td>Sāṭisnā</td>
<td>39370998</td>
<td>1866</td>
<td>1947</td>
<td>14</td>
<td>87464</td>
</tr>
<tr>
<td>5</td>
<td>Muḥammad Kurd ‘Alī</td>
<td>32272677</td>
<td>1876</td>
<td>1953</td>
<td>13</td>
<td>239829</td>
</tr>
<tr>
<td>5</td>
<td>Yūsuf Jirjis Zakham</td>
<td>1880</td>
<td>13</td>
<td></td>
<td>97264</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ʿAbd al-Qādir al-Maghrīb</td>
<td>118432135</td>
<td>1867</td>
<td>1956</td>
<td>7</td>
<td>62440</td>
</tr>
</tbody>
</table>


\(^{67}\) The article count for Muḥammad Kurd ‘Alī depends on the definition of an “article” and whether we include texts without byline but whose authorship can be established without doubt. If we count each section in his series *gharaʾib al-gharb* and *fi diyar al-gharb* as independent articles (they were originally published as letters in the newspaper *al-Muqtabas*), his article count would significantly spike.
The four men out of the five, for whom we can find biographical records, are in many aspects exemplary of the modernizing late Ottoman Empire and the Middle East: Coming from a plurality of religious and social backgrounds – Greek Orthodox, Catholic and Sunni Muslim, priest and leading Salafi thinker of the second generation, part-time officials, of simple means and members of the old elites – they belonged to the same generation (born between the mid-1860s and mid-1870s) and worked as journalists, teachers, and occasionally politicians. All of them were highly mobile and well-travelled and had good command of local as well as foreign languages – to the extent that some of them published literary translations. The fifth man is not less exemplary, even though his story seems to be rather uncommon among journalists: Yusuf Jirjis Zakham was one of the many emigrants from Greater Syria to America. He arrived in the US in 1902 and was naturalized in 1904, settled in Lincoln, Nebraska, where he married Myra from Iowa and had at least five children. Both spouses were literate and Joseph George Zakem, as his name was recorded, provided his profession as newspaper correspondent in the 1910 US Federal Census.68

The geographic distribution and relative frequencies of locations mentioned in bylines conveys the same image as the network of referenced periodicals and the brief comments on the most prolific authors’ biographies: al-Muqtabas was a publication of at least regional importance. It reached well beyond Greater Syria to Egypt, Iraq and even America, turning the famous proverb “Cairo writes, Beirut publishes and Baghdad reads” upside down with Baghdad well ahead of even Damascus.69

The picture is different for al-Ḥaqāʾiq (table 4), which was repeatedly in conflict with al-Muqtabas over the latter’s supposed moral laxity. Its most prolific contributors were Damascene Sunni religious scholars from notable families, many of whom were at least one generation older than its opponents (the average year of birth for al-Ḥaqāʾiq is 1837 and 1869 for al-Muqtabas). Among them are Ibrāhīm Mardam Bek, Muḥammad ʿĀrif al-Munir al-Ḥusaynī (born 1847/48), Mukhtar al-Muʿayyad (born 1822) and Muḥammad al-Qāsimī (born 1843), whose son Jamāl al-Dīn al-Qāsimī was among al-Muqtabas’ contributors. The initially surprising finding of very limited overlap between the two networks of authors published in journals from the same city, becomes


less so against this backdrop. Looking at the top 14 contributors to both journals, we can also note that whereas only two authors from *al-Muqtabas* are missing from VIAF and thus international library catalogues, the same is true for eight of *Ḥaqāʾiq*’s most frequent authors (tables 3 and 4).

Table 4: The 14 most prolific authors in *Ḥaqāʾiq* by number of bylines.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>VIAF ID</th>
<th>Birth</th>
<th>Death</th>
<th>Articles</th>
<th>Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N A</td>
<td></td>
<td></td>
<td></td>
<td>228</td>
<td>201,669</td>
</tr>
<tr>
<td>2</td>
<td>‘Abd al-Raḥmān al-Qaṣṣār</td>
<td>1863</td>
<td>1931</td>
<td></td>
<td>10</td>
<td>2,271</td>
</tr>
<tr>
<td>3</td>
<td>Ibrāhīm Khalil Mardam Bek</td>
<td></td>
<td></td>
<td></td>
<td>7</td>
<td>4,799</td>
</tr>
<tr>
<td>4</td>
<td>Muḥammad Farīd Wajdi</td>
<td>1875</td>
<td>1954</td>
<td></td>
<td>6</td>
<td>9,152</td>
</tr>
<tr>
<td>5</td>
<td>Muḥammad ‘Arif al-Munayyir</td>
<td>299025643</td>
<td></td>
<td></td>
<td>5</td>
<td>3,151</td>
</tr>
<tr>
<td>6</td>
<td>Ṣāliḥ al-Sharif</td>
<td>58892856</td>
<td>1869</td>
<td>1920</td>
<td>4</td>
<td>5,003</td>
</tr>
<tr>
<td>6</td>
<td>‘</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>2,867</td>
</tr>
<tr>
<td>6</td>
<td>Muḥammad al-Qāsimī al-Hallāq</td>
<td>51567828</td>
<td>1843</td>
<td>1900</td>
<td>4</td>
<td>10,156</td>
</tr>
<tr>
<td>6</td>
<td>Muḥammad Salīm al-Ḥifnī</td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>563</td>
</tr>
<tr>
<td>7</td>
<td>Mukhtar al-Muʿayyad</td>
<td>267054449</td>
<td>1822</td>
<td>1921</td>
<td>4</td>
<td>953</td>
</tr>
<tr>
<td>7</td>
<td>Abū al-Diya</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>5,098</td>
</tr>
<tr>
<td>7</td>
<td>ʿAbd al-Jawbārī</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>1,156</td>
</tr>
<tr>
<td>7</td>
<td>Abū al-Bāṣḥā</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>4,274</td>
</tr>
<tr>
<td>7</td>
<td>Ṣalāḥ al-Dīn al-Zaʿīm</td>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>2,448</td>
</tr>
</tbody>
</table>

A map of the relative frequency of locations mentioned in bylines confirms the brief overview of the authors’ biographies – *Ḥaqāʾiq* was a parochial paper with a focus on local issues. Its geographic network was mainly restricted to Damascus itself and the cities of the Syrian hinterland. Similarly distinctly regional distributions of authorship can be established for one of the two remaining periodicals in our corpus: *Ḥāsnā*: *Lughat al-ʿArab*, on the other hand, only rarely provided locations in bylines (26 of 939 articles), which doesn’t allow for meaningful observations.70

**IV. Authorship Attribution**

It is worth going back to the bibliographic data, its shortcomings and the resulting consequences for our analysis. We are particularly concerned with the number of articles that carried bylines or otherwise easily identifiable

---

authorship information. All journals in our corpus, like any other periodical at the time I have seen, published only limited authorship information. About 42 % of all articles in al-Haqa’iq carried authorship information (table 1). Second is al-Hasnā ‘with 36 %, followed by Lughat al-Arab with 16 % and al-Muqtabas with not even 13 %. In consequence and due to the heavy weight of al-Muqtabas in our corpus, we can only map 16.91 % of the entire network of articles by looking at explicit bibliographic information alone. More than four fifths are hidden from our view. Surprisingly the question of authorship has not received much attention. The, often implicit and accepted, hypothesis is that periodical editors authored all articles for which they did not provide a meaningful byline themselves. This raises a number of problems and considerations. Most importantly, the hypothesis remains untested. Second, we simply do not know enough about any given periodical to even name all editors. Cover pages of journals and mastheads of newspapers had a limited vocabulary to state responsibilities for an issue, not all of which were always provided: owner or concessionaire (şâhib, şâhib al-imtiyâz), responsible director (al-mudîr al-masûl) and editor-in-chief (ra’îs al-tahrîr). Commonly these functions converged and periodicals provided only a single name. Third, it is highly unlikely that a single person authored and edited almost the complete content of a periodical in addition to operating the whole business of publishing. Some owners-cum-editors ran more than one periodical. Muhammad Kurd ‘Ali, for instance, published a daily newspaper in addition to his monthly journal. We must therefore conceptualize potential authorship by editors as a collaborative endeavor. Fourth, owner-cum-editors were repeatedly absent from the place of publication and it is hard to believe they would have run their periodicals via the telegraph. Either a periodical suspended publication during

71 In addition to intrinsic bibliographic information, such as bylines and footnotes, I also explicitly encoded extrinsic information on authorship. Muḥammad Kurd ‘Ali, for instance, is well known to be the author of a series of letters titled Gharā‘ib al-Gharb (Oddities of the West). The gathering and encoding of extrinsic information cannot be considered systematic or comprehensive at this point.

72 The problem is not even mentioned in Ayalon, The Press in the Arab Middle East.

73 For examples of this implicit hypothesis see Seikaly, Damascene Intellectual Life, pp. 130 f., where he assigns a number of anonymous articles in al-Muqtabas to Muhammad Kurd ‘Ali without even discussing his assumption. Christiane Czygan, Zur Ordnung des Staates. Jungosmanische Intellektuelle und ihre Konzepte in der Zeitung Hûrriyet (1868 – 1870), Berlin 2012, p. 120, makes the same assumption by stating that it wasn’t entirely clear if Diyā Bey authored the final 36 issues of the journal Hûrriyet all by himself.

74 Ottoman press codes mandated identifiable information of publishers as well as the legal responsibility of the person signing off on the bottom of the (last) page for any paper’s content as if he was the author. Maṭbū‘ât nizâmnamesi, in: Düstür, vol. 2, Der-i Sa‘âdet 1872 (Tertip I), pp. 220 – 226. The function and responsibilities of the mudîr masûl were stipulated by Maṭbū‘ât kânûnû, in: Düstür, vol. 1, Der-i Sa‘âdet 1913 (Tertip II), pp. 395 – 403.
their absence or it must have been produced by someone else in their stead. In the case of al-Muqtabas, we know that Kurd ‘Ali fled Damascus and the Ottoman Empire twice for extended periods of time due to conflicts with the censors. During his newspaper’s suspensions between September 1909 and March 1910 and between April and August 1912, Kurd ‘Ali fled to Egypt and then traveled through Europe. His travelogues were serialized in his periodicals as Gharāʾib al-Gharb (Oddities of the West) and Fi diyār al-Gharb (In the Countries of the West) and later published in book form. The consequence of his absences on his potential authorship for articles in al-Muqtabas is not entirely clear. This is partially due to the difficulty of establishing actual publication dates as alluded to above. In 1909/10 the publication frequency of al-Muqtabas corresponded with Kurd ‘Ali’s absence from Damascus, while during the latter period, the publication of al-Muqtabas moved to Cairo – yet, Kurd ‘Ali was traveling through Europe and can therefore not be assumed as the main editor. This is supported by repeated calls from multiple periodicals on authors of anonymously submitted contributions to come forward and provide their identities to the publishers. Meanwhile, the newspaper al-Muqtabas explicitly rejected allegations that articles by pseudonymous authors were indeed authored by the editors.

(Computational) stylistics or stylometry is a common and established approach in linguistics and literary studies for authorship attribution and genre detection. It is based on the observation “that authors tend to write in relatively consistent, recognizable and unique ways,” which is particularly true for an author’s choice of words. In the context of stylometry, “style” commonly means a frequency count of words used in a given text. Stylometry

75 Ayalon, The Press in the Arab Middle East, p. 224, provides examples of two journals informing their readers that they will have to publish double issues due to absences of their publishers. Min wa-ilā ‘l-qirāʾ, al-Zuhūr, 1, no. 8, 1.10.1910, pp. 368 f., https://openarabicpe.github.io/journal_al-zuhur/tei/oclc_1034545644-i_7.TEIP5.xml#div_1.d2e2020, apologizes to its readers for the delay in publishing this issue because the editor was traveling outside Egypt/Cairo to Beirut.

76 The back cover of al-Muqtabas volume 4, issue 4 articulates the hope that issue 5 and 6 would be published as a double issue in mid-August 1909 (mid-Sha’bañ 1327 A H). Yet, issue 7 was published only in the first week of April 1910. Jaridat al-Muqtabas, 7.4.1910, p. 3.


79 Instead, they claimed, not all ‘ulamāʾ and udābāʾ in support of progress can speak out. Ḥawādith dākhliyya, Jaridat al-Muqtabas, 17.2.1909, p. 3.


81 For an introduction to statistical methods for authorship attribution see Moshe Koppel et al., Computational Methods in Authorship Attribution, in: Journal of the American
then computes degrees of similarity between texts, called distance measure, through comparing multivariant frequency lists of textual features. The important catch is that stylometry is a comparative method. In order to establish similarities one has to have access to a significant corpus of digital texts by authors likely to be found among the unattributed texts. If we only compare every article in our periodical corpus to every other article in the same corpus, we cannot possibly identify any author not yet named in a byline. Instead, the best we could hope for would be to establish groups of texts that have a certain likelihood of having been authored by the same person.

The present essay is the first foray into stylometric authorship attribution for Arabic periodicals. Christiane Czygan’s work is the only attempt at stylometric authorship attribution for Ottoman periodicals I have come across, but, after developing a set of style markers for individual editors-cum-authors, the author did not apply them for actual authorship attribution.82

There is some debate as to which style-markers and distance measure should be considered for authorship attribution, but I settled on Most Frequent Words (MFW) and John Burrows’ Delta.83 Texts were not pre-processed by morphologizing or lemmatizing, as this would reduce the authorship signal to the vocabulary used. Similarly, function words, which, by definition, are the most frequent words, were not removed. Their frequency is independent of a text’s topic and it is unlikely that an author can consciously control this frequency.84

Because the number of MFW has a significant impact on the results, Maciej Eder suggested to use consensus networks in order to separate signal and noise. To this end, one computes the nearest neighbor as well as the first two runners-up for a sequence of MFW (for example from 100 to 1,000 MFW in increments of 100) and then combines the results in a single output, which serves as a form of self-validation for the more robust signals. The results can then be visualized using network analysis.85

Finally, there is an important caveat in applying stylometry to periodicals: Eder experimentally established a threshold length of 5,000 words as the

82 Czygan, Zur Ordnung des Staates. This work falls short on many counts beyond introducing a promising approach. Basic statistic measures about her corpus and computational approaches remain unknown and most importantly, she tested only one [sic!] anonymous article with a set of parameters she took eighty pages to develop.
84 Koppel, Computational Methods in Authorship Attribution, p. 11.
minimal required length of a text for meaningful attribution. Below 5,000 words, the signal was “immensely affected by random noise.” These findings have severe implications because most texts in our corpus are much shorter than 5,000 words and even the longer ones are too short for random sampling. Nevertheless, limiting our experiments with stylometric analysis to the some fifty articles of more than 5,000 words yielded promising results and shows at least three distinct signals: genre, author and translator/editor.

Our initial analysis of all articles of 5,000 words and more using bootstrap networks for 100 to 1,000 MFWs confirms the general applicability of stylometry to our corpus (figure 3). Articles form clusters based on edge weight and modularity around authors, editors and translators. Thus, we find clusters of articles authored by Jamāl al-Dīn al-Qāsimī, Ahmad Zakī, Charles Seignobos, and Muḥammad Kurd ʿAlī. The latter was the translator of Seignobos’s works from French, which is clearly visible in the close proximity between clusters of works by both authors. A similar relation can be observed

---

86 Eder, Does Size Matter?, p. 170. He also established that the accuracy does not improve any further beyond a length of 15,000 words.
in the proximity between clusters of works authored by Jamāl al-Dīn al-Qāsimī and classic texts he edited and commented on. In general, the network analysis shows only limited stylistic overlap between authors and we can assume with a high degree of confidence that the cluster of anonymous articles on the bottom right was authored by Kurd ʿAlī. Since the sample contains almost exclusively articles from al-Muqtabas, this would tentatively confirm the authorship hypothesis of editors. However, there is also a cluster of non-attributed articles in the centre between al-Qāsimī’s and Zakī’s that are stylistically distinct and far removed from Kurd ʿAlī’s texts. This contradicts the hypothesis and points to another, unknown author. Furthermore, the plot also shows a strong signal of genre: the cluster of unattributed texts most likely written by Kurd ʿAlī are all geographic works.

V. Conclusion

In this essay, I questioned hyperbolic promises of ubiquitous digitized knowledge from the marginal position of Middle Eastern intellectual history and by outlining the techno-infrastructural challenges faced by a digital history of societies outside the Global North. I showed how a digital episteme deeply rooted in twentieth-century, English-speaking capitalism requires mitigation strategies on every level of the digital workflow. These are placed on the individual scholar and involve significant investments in the making of corpora, resources and tools if we want to reap the promised fruits of the digital humanities. I also posed that one of the consequences of this episteme is a neo-colonial silencing of the material heritage of the societies in the Eastern Mediterranean.

Nevertheless, digital corpora and computational approaches are indispensable for scrutinizing the periodical press as an ideosphere. I argued that one has to transcend the individual periodical and engage in a systematic study of the periodical press at scale in order to better understand both the intellectual history of the Eastern Mediterranean and periodical production itself. The case study of four late Ottoman Arabic-speaking periodicals from the Eastern Mediterranean introduced and evaluated some of the mitigation strategies. After introducing my own efforts of building an open and scholarly digital corpus, I engaged in computational exploration through network analysis, mapping, and stylometry along the guiding question of what were the core nodes (authors and periodicals) in the ideoscape of the late Ottoman Eastern Mediterranean.

Modeling the network of references to periodical titles, I could confirm established knowledge about the importance of certain journals over others. The Cairene journals of al-Manār, al-Muqtatāf and al-Hilāl were indeed central to the late Ottoman Arabic ideosphere, even though they were published outside the Ottoman Empire. A future systematic exploration of
periodicals will have to digitize these and compare them to *al-Muqtabas*, which shows many traits of a periodical of transregional importance very different from the other journals in our corpus. The exploration of the network of article authors, on the other hand, provided a number of surprising results that will need to be addressed in future scholarship: The noted importance of Iraqi writers over Syrians among the core nodes of the network contradicts the common narrative about the Arabic renaissance (*nahda*). A similar importance of Baghdad over the more commonly assumed centres of Arab intellectual production is found in the geographic distribution of toponyms in bylines for *al-Muqtabas*. One would, again, need to test other transregional periodicals to see whether they provide a similar Iraqi connection or if *al-Muqtabas* is an outlier in this regard. The network analysis of authors also demonstrated lacunae in the historiography of the Arabic press and the intellectual history of the late Ottoman Arab ideosphere, namely the absence of the core nodes in our network from major works of scholarship.

Any analysis of authorship and socio-intellectual networks is limited by the fact that less than one fifth of all articles carry identifiable authorship information. I, therefore, presented a first empirical analysis of the common but untested hypothesis that editors authored the four fifths of anonymous articles themselves by submitting our corpus to stylometric authorship attribution. This provided significant hints towards authors of articles longer than 5,000 words. Although we cannot (yet) assign a specific name, the analysis of articles from *al-Muqtabas* returned one cluster of texts by an anonymous author very different from those articles that can indeed be attributed to the editor Muḥammad Kurd ʿAlī.

Dr. Till Grallert, Orient-Institut Beirut, Rue Hussein Beyhum 44, Beirut, Libanon
E-Mail: till.grallert@fu-berlin.de