

Looking at the iceberg from below the waterline: Stylometric answers to unattributed texts in early Arabic periodicals

Till Grallert

Humboldt-Universität zu Berlin, Berlin, Germany

ORCID: 0000-0002-5739-8094

Abstract: From the mid-nineteenth century onward, Arabic periodicals were the first mass-medium of South West Asia and North Africa (SWANA) and formed an increasingly globalised public sphere. Consequently, they attract increasingly scholarly attention. However, most articles up until at least World War I were published without a proper byline, commonly causing scholars to conflate owners-cum-editors with their journals. This paper computationally addresses the question of authorship through stylometric authorship attribution applied to data sets derived from digital corpora of journals published in Baghdad (Lughat al-‘Arab), Cairo (al-Zuhūr, al-Muqtabas), and Damascus (al-Muqtabas, al-Ḥaqā’iq) between 1906 and 1918. It demonstrates, firstly, that titles were stylistically sufficiently different to identify a distinctive authorial voice for each periodical. Secondly, it shows that individual anonymous articles cannot be reduced to a single author for each journal and that owners-cum-editors should not be considered strong candidates for their authorship. Thirdly, it also rejects owners-cum-editors as strong candidates for the bulk of anonymous texts originally published in short chunks as part of news sections, announcements, miscellanea etc. Finally, this paper rejects all known authors in our corpus as potential editors thus necessitating further, non-computational research.

Keywords: Stylometry, Authorship attribution, Network analysis, Arabic Periodicals, 19th century

1. Background and research question

From the mid-nineteenth century onwards, periodicals became the first mass media of the Ottoman Eastern Mediterranean. Magazines, journals, and newspapers in the region’s many languages, from Arabic to Ottoman, Greek, Armenian, Coptic or Hebrew, to French, English or Italian were at the core of social developments and formative discourses of modern(ising) societies that carry continued prominence across the region and beyond: the Arabic (cultural) renaissance (*nahda*), Arab nationalism or the Islamic reform movement (*ṣalafiyya*), to name just a few. Many papers operated on large transnational cultural planes far exceeding their local contexts. This is particularly true for the large Arabic speaking communities in North Africa and the diaspora in the Americas and Europa. Yet, the vast majority of Arabic periodicals remain obscure and understudied beyond a few well-known titles from Beirut and Cairo.

The overwhelming majority of articles in Arabic periodicals from the late Ottoman Eastern Mediterranean (c.1850–1918) carried no explicit authorship information. This is particularly true for newspapers but also for magazines and journals.¹ Based on the very small corpus of fully modelled digital editions allowing for such queries, we can estimate that on average 75 per cent of all articles in

¹ Till Grallert, Catch Me If You Can! Approaching the Arabic Press of the Late Ottoman Eastern Mediterranean through Digital History, in: *Geschichte und Gesellschaft* 47 (2021), pp. 58–89; Nicole Khayat, What’s in a name? Perceptions of authorship and copyright during the Arabic nahda, in: *Nineteenth-Century Contexts* 41 (2019), pp. 423–440.

magazines were published anonymously (see the section on corpus below and tbl. 1). Yet, the question of authorship has not received much attention in existing scholarship and is strikingly absent from standard works in the field.² The common implicit hypothesis considers publishers listed in mastheads, imprints or encyclopaedic works as the sole authors of all anonymous texts within a periodical.³

In the absence of any documentary sources detailing the actual processes of editing and publication, we have to rely on a combination of circumstantial knowledge, which strongly implies the fallaciousness of the hypothesis as it generally ignores the known practical and infrastructural realities of periodical production as well as individual publishers's biographies and the specific contexts of individual titles.⁴ In this short paper, I will, therefore, turn to the material artefact of the periodicals themselves and apply stylometric authorship attribution to a corpus of digitised Arabic periodicals in order to empirically test the hypothesis that publishers authored most if not all anonymous texts in their periodicals.

Periodical	Place	Dates	Articles	with author	% of total	DOI
al-Ḥaḡā'iq	Damascus	1910–13	389	163	41.90	10.5281/zenodo.1232016
al-Muḡtabas	Cairo, Damascus	1906–18	2964	377	12.72	10.5281/zenodo.597319
al-Ustādḥ	Cairo	1892–93	435	24	5.52	10.5281/zenodo.3581028
al-Zuhūr	Cairo	1910–13	436	181	41.51	10.5281/zenodo.3580606
Lughat al-'Arab	Baghdad	1911–14	939	152	16.19	10.5281/zenodo.3514384
average			1032.6	179.4	23.57	

Table 1: authorship information based on the “Open Arabic Periodical Editions” corpus

2. Method

Computational stylistics or stylometry is a well-established approach in linguistics and literary studies for authorship attribution and genre detection for major languages of the Global North and has been successfully applied in English and German periodical studies.⁵ ‘Style’, in this context, refers to patterns in the distribution of most frequent linguistic features (most commonly, token or character n-grams). These patterns are captured as lists of most frequent features (MFF), which are then clustered by similarity according to a variety of distance measures (for example, delta, cosine, euclidean, manhattan, etc.).⁶ Controlled corpora of texts with known authorship have shown that this comparative approach can identify authorship of specific texts with high accuracy.

There are two challenges for stylometric analysis of periodical articles. First, the precision of the approach rapidly declines below a minimal length of analysed texts. Based on his empirical work Eder recommends random sampling and at least 5000 word tokens as a conservative threshold for meaningful attribution of prose in English, German, Hungarian, and Polish after which accuracy did not improve

² E.g. Ami Ayalon, *The press in the Arab Middle East: a history*, New York 1995.

³ E.g. Samir Seikaly, *Damascene Intellectual Life in the Opening Years of the 20th Century: Muhammad Kurd 'Ali and Al-Muḡtabas*, in: Marwan Rafat Buheiry (ed.), *Intellectual Life In The Arab East, 1890-1939*, Beirut 1981, pp. 130, 131; Christiane Czygan, *Zur Ordnung des Staates: jungosmanische Intellektuelle und ihre Konzepte in der Zeitung Hürriyet (1868-1870)*, Berlin 2012, p. 120; Kais Ezzerelli, *The publicist and his newspaper in Syria in the era of the Young Turk Revolution, between reformist commitment and political pressures: Muhammad Kurd Ali and al-Muḡtabas (1908-17)*, in: Anthony Gorman/Didier Monciaud (eds.), *The press in the Middle East and North Africa, 1850-1950: politics, social history and culture*, Edinburgh 2017, pp. 176–206.

⁴ C.f. Grallert, *Catch me if you can!*, pp. 84–85.

⁵ Francesca Benatti / David King, *Hidden Authors and Reading Machines: Investigating 19th-century authorship with 21st-century technologies*, in: SHARP 2017: *Technologies of the book*, Victoria 2017; Mike Kestemont / Gunther Martens / Thorsten Ries, *A Computational Approach to Authorship Verification of Johann Wolfgang Goethe's Contributions to the Frankfurter gelehrte Anzeigen (1772–73)*, in: *Journal of European Periodical Studies* 4 (2019), pp. 115–43.

⁶ See John Burrows, ‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship, in: *Literary and Linguistic Computing* 17 (2002), pp. 267–287; Maciej Eder, *Does size matter? Authorship attribution, small samples, big problem*, in: *Literary and Linguistic Computing* 30 (2015), pp. 167–182; Moshe Koppel / Jonathan Schler / Shlomo Argamon, *Computational methods in authorship attribution*, in: *Journal of the American Society for Information Science and Technology* 60 (2009), pp. 9–26.

any further.⁷ For nineteenth-century Arabic, we could show that accuracy plateaued much earlier, which allowed for almost 100% matching between authors at a sample length of 2500 tokens.⁸ Despite such a reduced minimal length, most anonymous periodical texts still fall below this threshold and cannot be directly attributed through stylometric approaches. Second, as a comparative method providing clusters of texts with high relative similarity within a given corpus, all results are extremely sensitive to the composition of that corpus. Adding a very unlikely candidate (and thus stylistically very different text) will cause all other texts to become more similar to each other. Stylometric authorship attribution consequently works best for corpora comprising a relatively large number of text for each author as well as a number of samples from the real author of an anonymous text. Both is not the case for periodicals.

Stylometric analysis was done in R with the `stylo()` package⁹ applying our experimentally established parameters for reliable authorship attribution of nineteenth-century Arabic texts.¹⁰ Accordingly, all texts were tokenized into words and sampled into bags of words of 2500 tokens. Authorship attribution was then based on computing Eder's simple delta as a distance measure for tables of MFF with a length of 200 to 500 tokens, which we incremented in steps of 100 tokens. No culling was applied. `stylo()` allows to use the resulting lists of distances between all texts for each MFF setting as a voting measure to weight distances over multiple iterations, computing a nearest neighbour as well as the first two runners-up. Results were written to an edges table for a *consensus network*. These networks were then analysed using the `tidygraph()` package to compute various centrality measures and detect communities. Results were plotted with `ggplot2()` and `ggraph()`¹¹.

3. Corpus and data set

Arabic periodicals from the late Ottoman Eastern Mediterranean resist computational analysis for a number of reasons, which can be summarised as the following layers of inaccessibility:¹²

1. A massive knowledge gap about the publication history of the vast majority of periodicals published from the late eighteenth century onwards. We simply do not know enough about any periodical to establish a definitive or comprehensive list of editors or even the identities of potential editors beyond the names provided by mastheads and title pages. In addition, periodical publishing had not yet evolved into fully differentiated roles and corresponding terminology. While mastheads used terms such as *owner* or *cessionnaire* (*ṣāḥib*, *ṣāḥib al-imiyyāz*), *responsible director* (*al-mudīr al-mas'ūl*) and *editor-in-chief* (*ra'is al-tahrīr*), they were frequently used interchangeably, while the functions themselves converged. Periodicals, thus, rarely provided more than a single name.
2. A collection bias with most known and traceable collections concentrated in the Global North, forming a patchwork of scattered print runs and issues (fig. 2). This collection bias is to a large

⁷ Eder, Does size matter.

⁸ Maxim Romanov / Till Grallert, Establishing parameters for stylometric authorship attribution of 19th-century Arabic books and periodicals, in: Book of Abstracts, Tokyo 2022, pp. 346–348.

⁹ R Core Team, R: A language and environment for statistical computing, manual, R Foundation for Statistical Computing, Vienna 2018; Maciej Eder / Jan Rybicki / Mike Kestemont, Stylometry with R: A Package for Computational Text Analysis, in: The R Journal 8 (2016), pp. 107–121.

¹⁰ Romanov; Grallert, parameters for stylometric authorship attribution.

¹¹ Thomas Lin Pedersen, tidygraph: A Tidy API for Graph Manipulation; Thomas Lin Pedersen, ggraph: An Implementation of Grammar of Graphics for Graphs and Networks; Hadley Wickham, ggplot2: Elegant graphics for data analysis, New York 2016.

¹² C.f. Till Grallert, Open Arabic Periodical Editions: A framework for bootstrapped scholarly editions outside the global north, in: Digital Humanities Quarterly 16 (2022), secs 4–17; Till Grallert, Digitalisierung des Kulturerbes des globalen Südens, in: Laura Busse et al. (eds.), Clio Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften, 3rd Ed., Berlin forthcoming; Till Grallert, Editing mundane texts across the digital divide: the case of Arabic periodicals from the late nineteenth-century Eastern Mediterranean, in: Constance Crompton et al. (eds.), Digital Humanities in Practice, London forthcoming.

part the result of the submission of the material's regions of origin to colonial policies. It consequently intersects with the knowledge gap.

3. A digitisation bias resulting from the former two and socio-technical infrastructures of the digital rooted in the epistemic hegemony of late twentieth-century, anglo-american neoliberal capitalism, which results in the lack of support for Arabic—the worlds second most important script after Latin and one of humanity's major languages with more than 420 million active speakers and one of only six official languages of the United Nations—across digital infrastructures. Such lack of support is, for instance, manifest in the absence of meaningful and accessible OCR until the recent advent of machine-learning based approaches.

Fig. 1 shows the distribution of all Arabic periodical titles published across South West Asia and North Africa (SWANA) between 1789 and 1929 based our Jarā'id data set.¹³ It also provides information on the ratio of titles with known holdings and digital remediations. Almost one third of the 775 or 21,83 per cent of all 3550 titles in the data set that can be located in collections—233 or 6.56 per cent of all titles—have digital remediations. While the digitisation quote of titles in collections is surprisingly high, it must be kept in mind that digitisation almost exclusively means digital facsimiles without a meaningful text layer and that we cannot currently resolve information on the extent of digitisation. Even if only a single issue of a given title was digitised, the periodical will be included in this count.

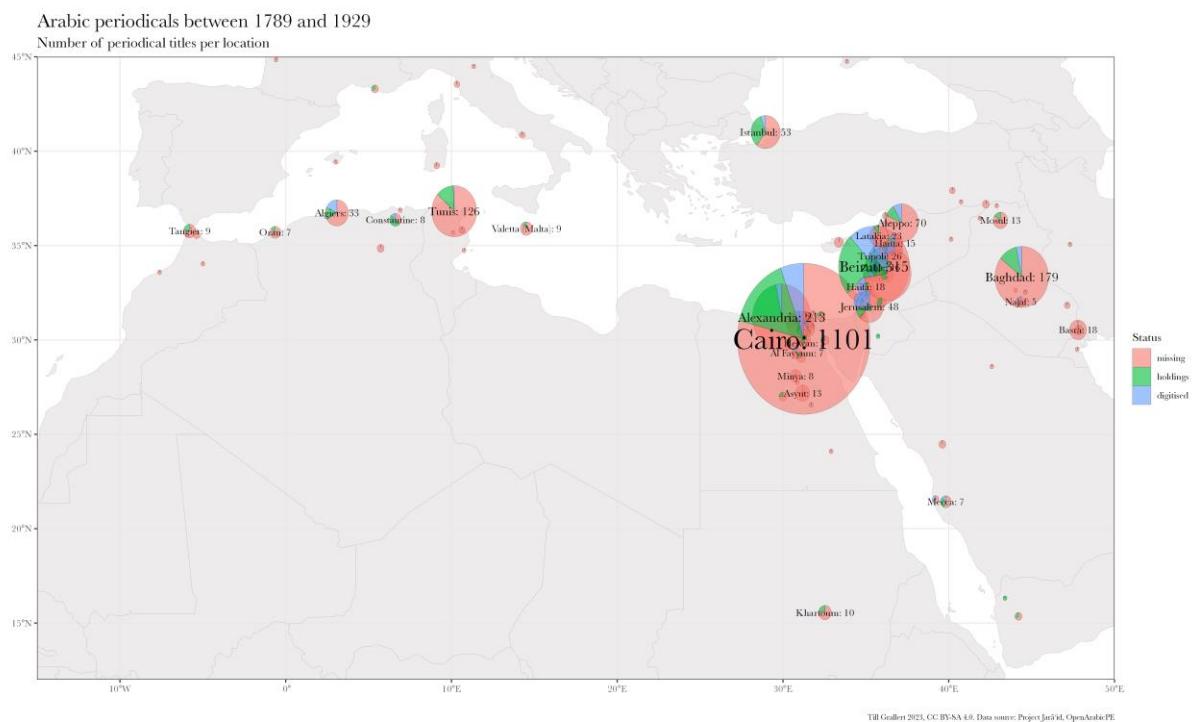


Fig. 1: Geographic distribution of Arabic periodical titles published across South West Asia and North Africa (SWANA) between 1789–1929. The size of the pie charts corresponds to the total number of titles published at a location. Slices show the percentage of known holdings and digitized collections.

¹³ Adam Mestyan / Till Grallert/et al., Jarā'id: A chronology of Arabic periodicals (1800-1929), Zenodo; Adam Mestyan / Till Grallert, A Chronology of Nineteenth-Century Periodicals in Arabic (1800-1900): A research tool, 2012–2015, <https://web.archive.org/web/20160422071133/https://www.zmo.de/jaraid/> (19. 11. 2020); Adam Mestyan / Till Grallert, Jara'id: A Chronology of Arabic Periodicals (1800-1929). 2020 Edition, 2020, <https://projectjaraid.github.io/> (19. 11. 2020).

Periodic al	Place	Dates ¹⁶	Vol. s	No. s	Words	Article s	with autho r (%)	2500 + word s	words / article	Authors	DOI
<i>al-Haqā'iq</i>	Damascu s	1910–13	3	35	298090	389	41.90	22	832.66	104	10.5281/zenodo.1232016
<i>al-Muqtabas</i>	Cairo, Damascu s	1906–18	9	96	1981081	2964	12.72	241	873.34	140	10.5281/zenodo.597319
<i>al-Zuhūr</i>	Cairo	1910–13	4	39	292333	436	41.51	6	695.09	112	10.5281/zenodo.3580606
<i>Lughat al-ʿArab</i>	Baghdad	1911–14	3	34	373832	939	16.19	21	485.21	53	10.5281/zenodo.3514384
total			19	204	2945336	4728		290	622.96		

Table 2: Our corpus of four periodicals from “Open Arabic Periodical Editions”

The corpus of TEI/XML files was converted into data sets of plain-text files with very basic normalisation of orthographic differences that can be attributed to printers’ and transcribers’ preferences. We removed hamza when it appeared on an *alif*-carrier (i.e. ٱ and ٱ to ٱ) and normalised final ٥ and ٥ as well as ٤ and ٤ to a single form each.¹⁷ To answer the two related research questions and accounting for the threshold of minimal required sample length, we have created a number of data sets from this corpus for stylometric analysis:

1. *Data set 1*: all 215 individual articles from our corpus that are longer than our experimentally established threshold of 2500 tokens. Serialised articles were compiled into single texts in order to increase the number of authors in this data set. Still, only a small number of texts is significantly longer than the minimal threshold. In addition, *data set 1* comprises a large number of authors (74) with only a small number of texts (112 in total) as well as a large number of unattributed texts (103). While *data set 1* includes texts from some of the periodicals’s known editors, it is unclear to which extent included authors could or should be considered potential candidates for authorship of the unattributed texts.
2. *Data set 2*: 88 compilations of anonymous articles below our threshold of 2500 tokens originally published in rubrics within a single issue. These mundane texts of news, notable events and inventions, book reviews etc. are assumed to be most likely to have been authored or heavily edited by the journals’ publishers.
3. *Data set 3*: all 612 issues from the four journals in our corpus. This data set will be used for basic sanity checks.

4. Results

In order to test the hypothesis that owners-cum-editors were indeed the authors of most, if not all, anonymously published articles in their respective periodicals, we commence with some sanity test. Namely, establishing whether individual periodicals are stylistically distinguishable at all. Fig. 3 demonstrates that this is clearly the case. It also shows that two issues of *al-Zuhūr* are stylistically different from the rest of the journal, which is confirmed by community-detection algorithms (fig. 4) and might indicate a shift in editorship. Similar shifts seemingly happened at other periodicals as well.

¹⁶ The current cut-off date is 1918

¹⁷ The code for the conversion is available at https://github.com/openarabicpe/convert_tei-to-markdown/.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

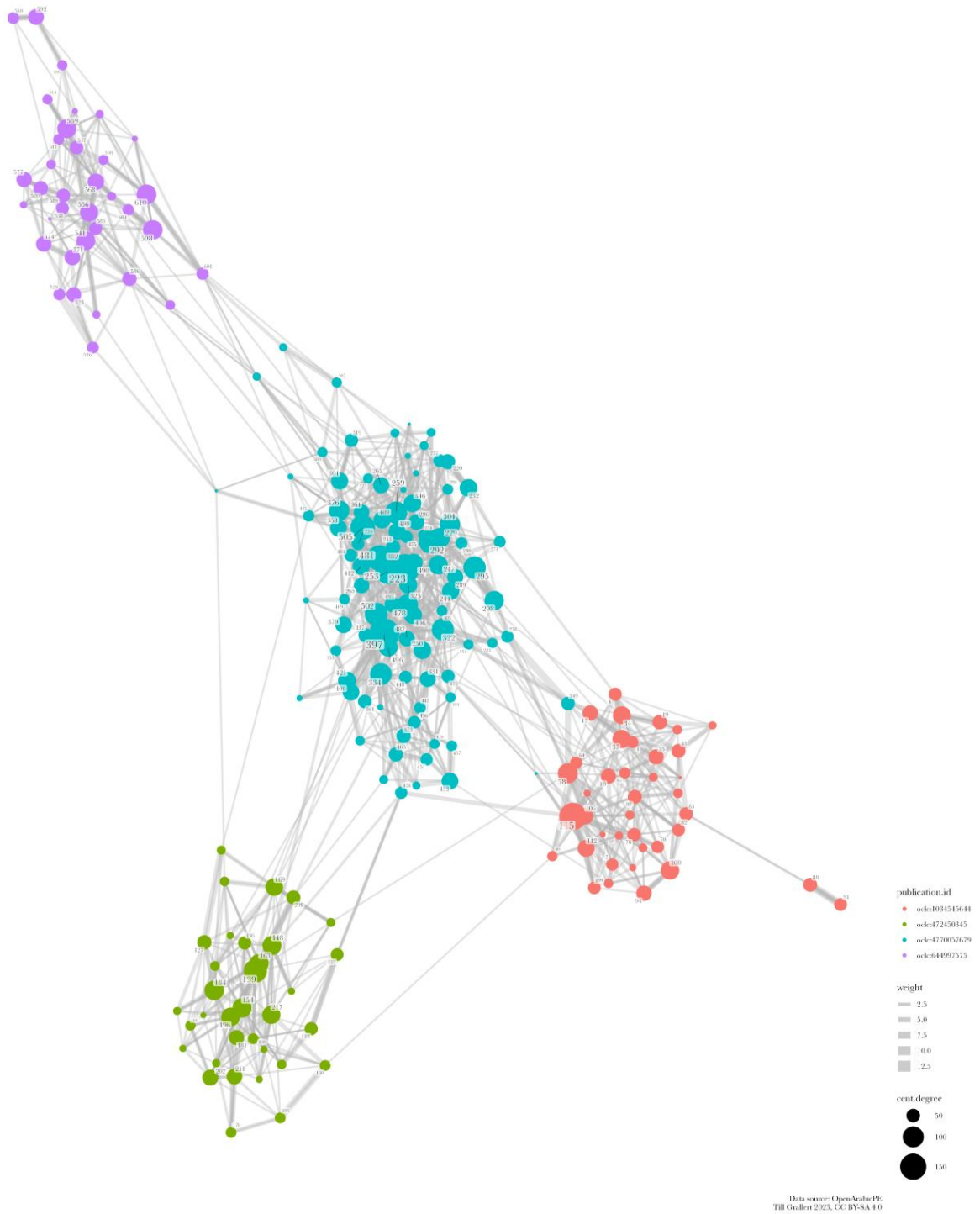


Fig. 3: Network plot of stylometric similarity in *data set 3*. Node colours indicate periodicals: purple = *al-Haqā'iq*, turquoise = *al-Muqtabas*, green = *Lughat al-'Arab*, red = *al-Zuhūr*.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

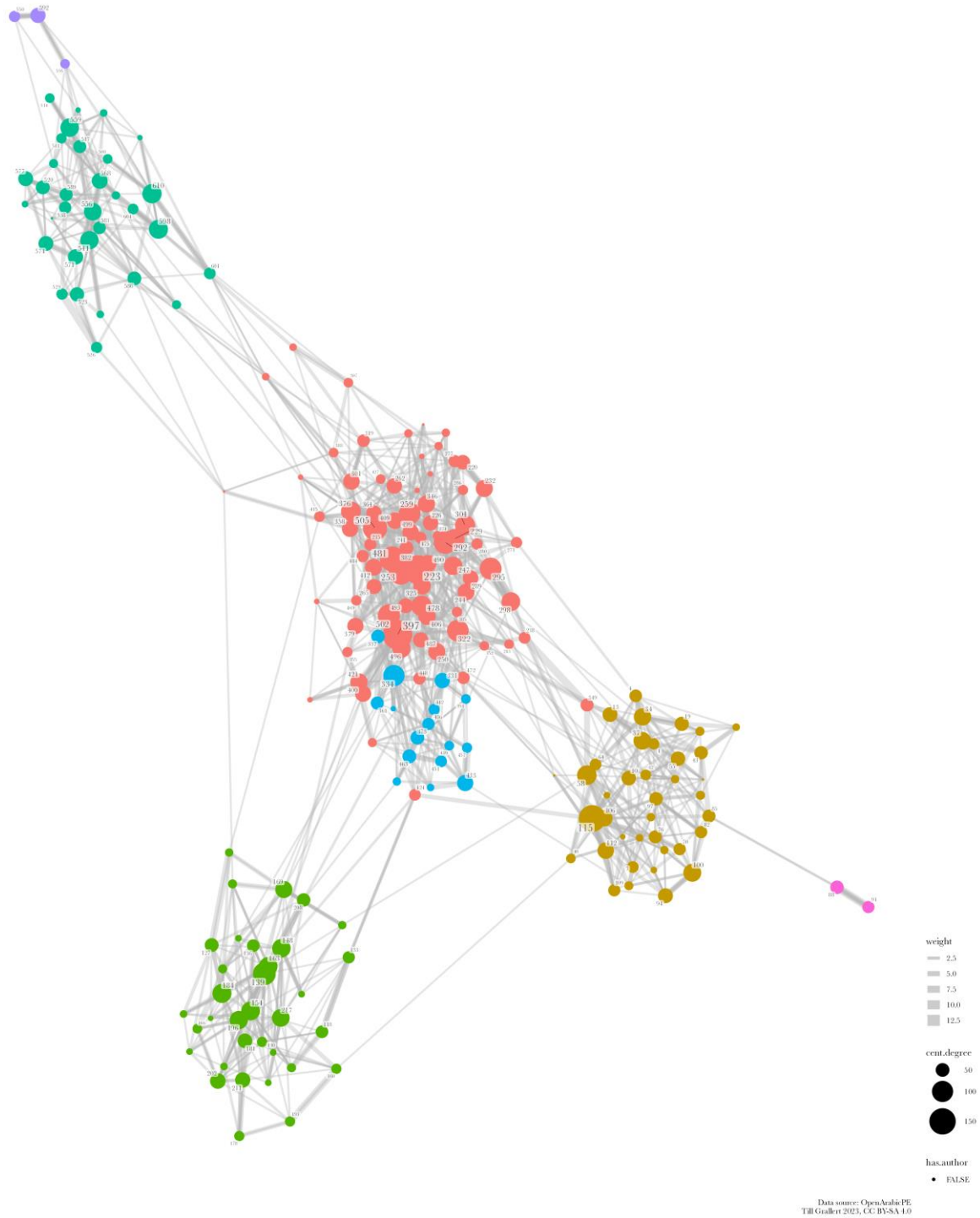


Fig. 4: Network plot of stylometric similarity in *data set 3*. Node colours indicate communities established by the walktrap algorithm.

The next step is to run stylometric analysis on *data set 1* and look whether (a) anonymous articles cluster by periodical and (b) if the owners-cum-editors present in this data set can be considered strong candidates for their authorship. Fig. 5 shows that both hypotheses can be rejected for *al-Muqtabas* and *Lughat al-‘Arab*, whose editors, Muḥammad Kurd ‘Alī, Anastās Mārī al-Karmalī, and Kāzīm al-Duyalī, are present in the data set (labelled with their respective identifiers as oape:878, oape:227, and oape:396). The anonymous articles from *al-Muqtabas* form two distinct clusters with substantial internal differences and thus force us to reject any notion of single authorship (fig. 6). In addition, neither

Muhammad Kurd 'Alī nor Anastās Mārī al-Karmalīand Kāzīm al-Duyalī seemingly authored a significant number of articles in their respective journals.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

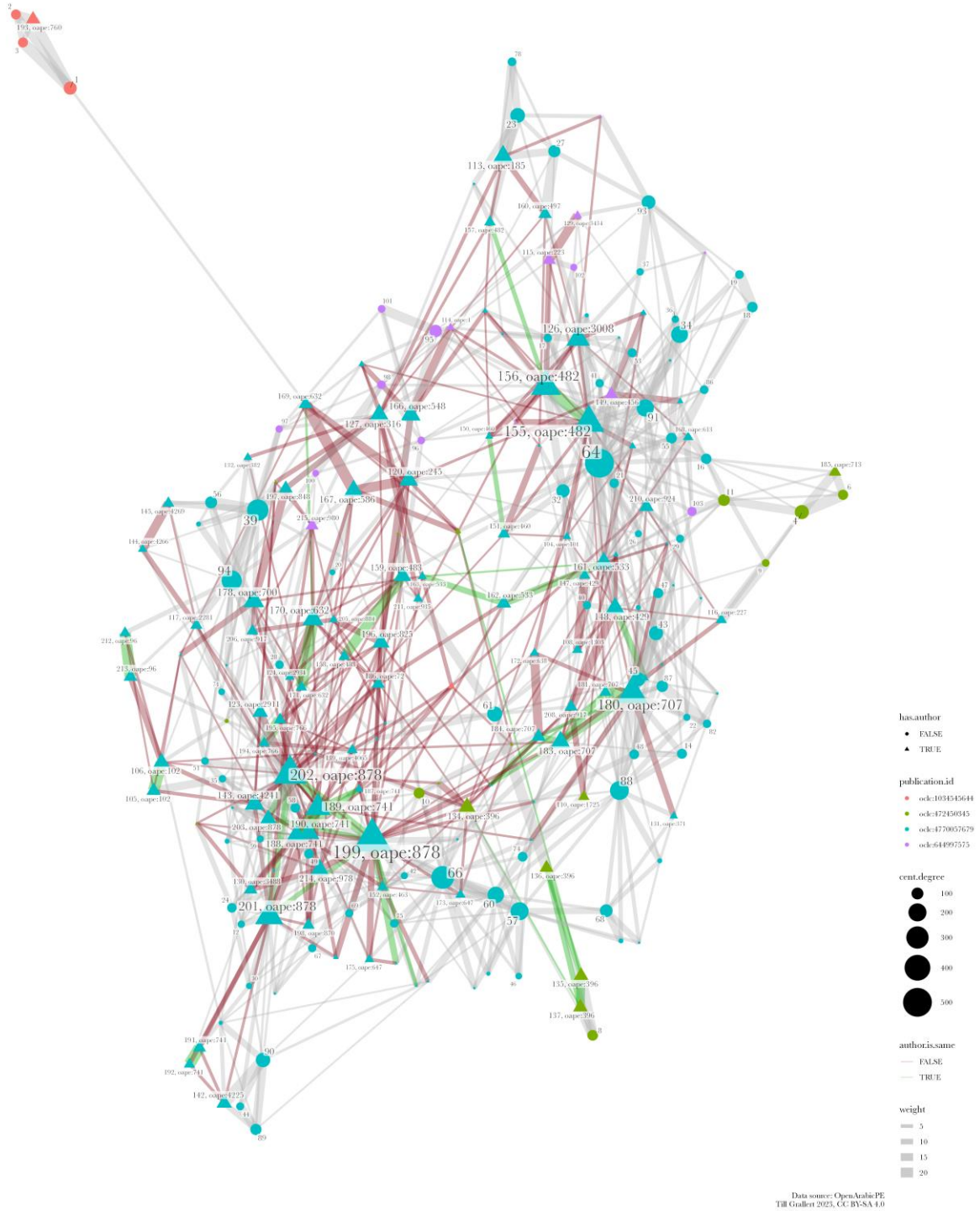


Fig. 5: Network plot of stylometric similarity for all articles from *data set 1*. Node colours indicate periodicals: purple = *al-Ḥaḡā`iq*, turquoise = *al-Muqtabas*, green = *Lughat al-'Arab*, red = *al-Zuhūr*.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

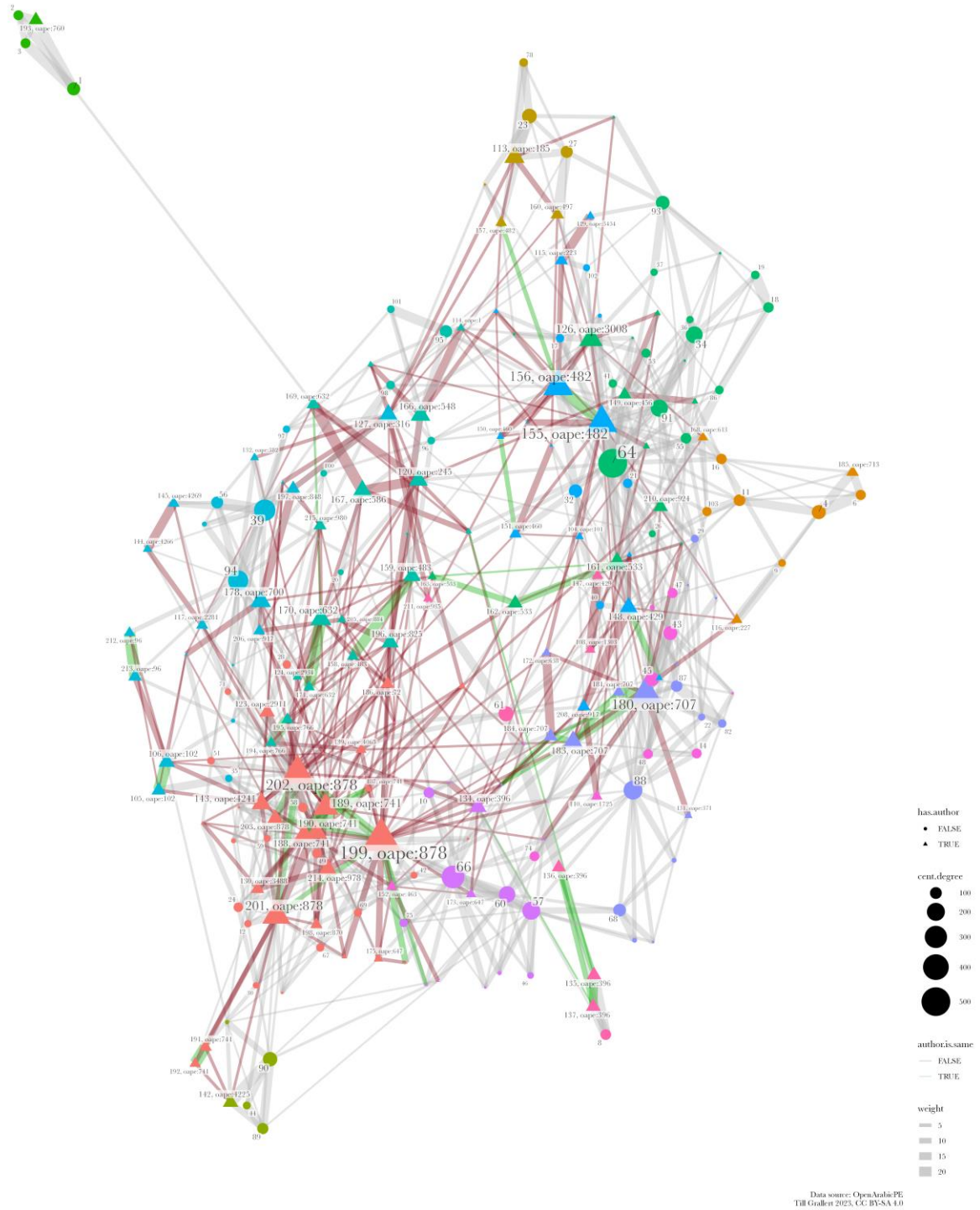


Fig. 6: Network plot of stylometric similarity for all articles from *data set 1*. Node colours indicate communities established by the Louvain algorithm.

The strongest candidates for anonymous texts likely to have been authored by the editors of a periodical are compiled into *data set 2*. To test the hypothesis, we divided *data set 2* by source publication and added texts from *data set 1* authored by the owners-cum-editors. For *al-Muqtabas* fig. 7 shows that without any additional contenders, Muḥammad Kurd ‘Alī could be considered the author of some issues while community detection algorithms hint at additional contributors for most issues.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

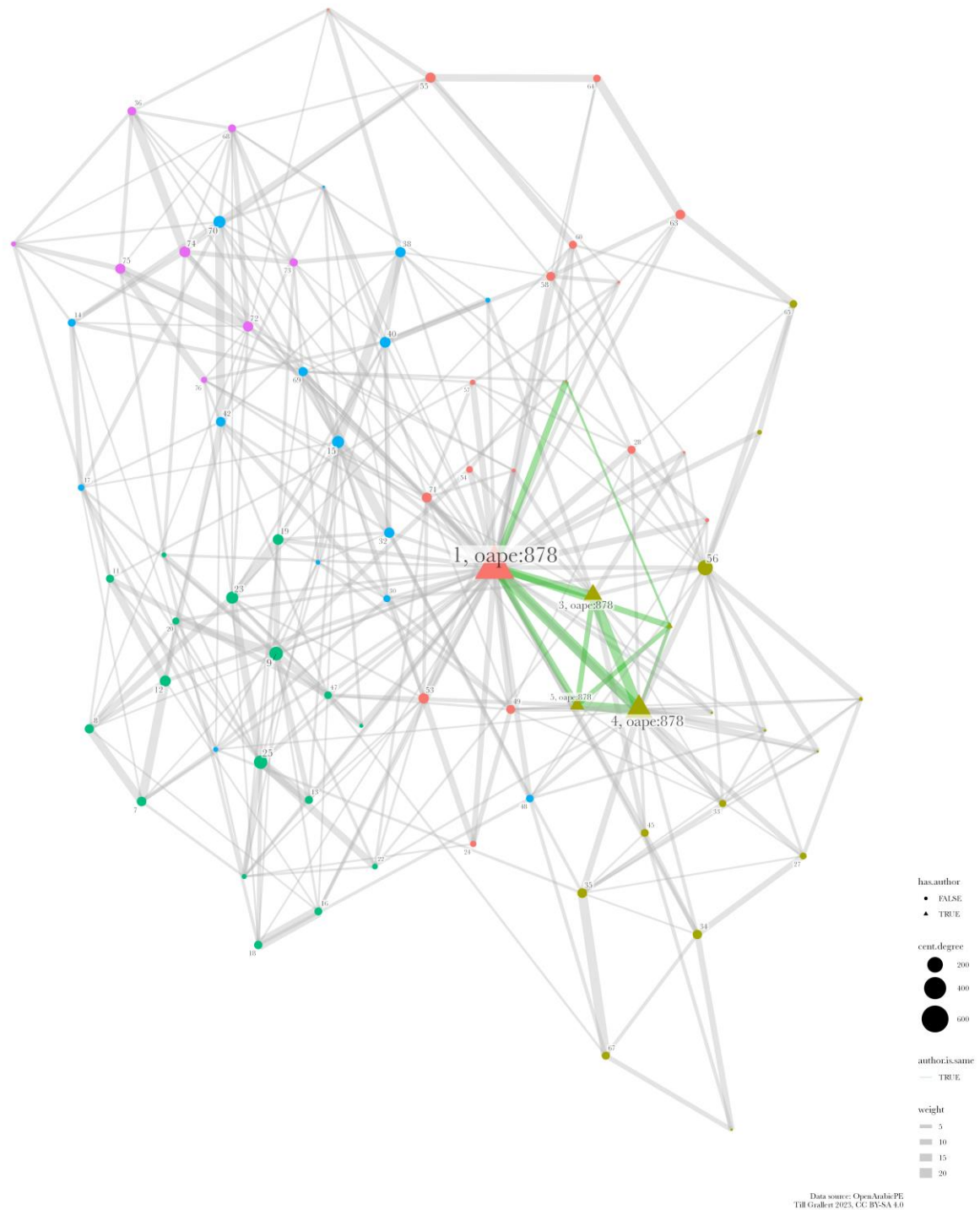


Fig. 7: Network plot of stylometric similarity for anonymous articles in *al-Muqtabas* from *data set 2* and articles by its publisher Muḥammad Kurd ‘Alī from *data set 1*.

Finally, we added all attributed articles from *data set 1* as a control group. We argue that if they were less likely to be editors (and thus more stylistically different), they would push the assumed owner-cum-editors closer to the anonymous articles. Fig. 8 and fig. 9 show that this is neither the case for *al-Muqtabas* nor *Lughat al-‘Arab*. This indicates that Muḥammad Kurd ‘Alī and Anastās Mārī al-Karmālī should not be considered strong contenders for the authorship of anonymous reporting in their journals. We can also safely exclude all other 73 authors in *data set 1* as potential candidates for the editorship.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

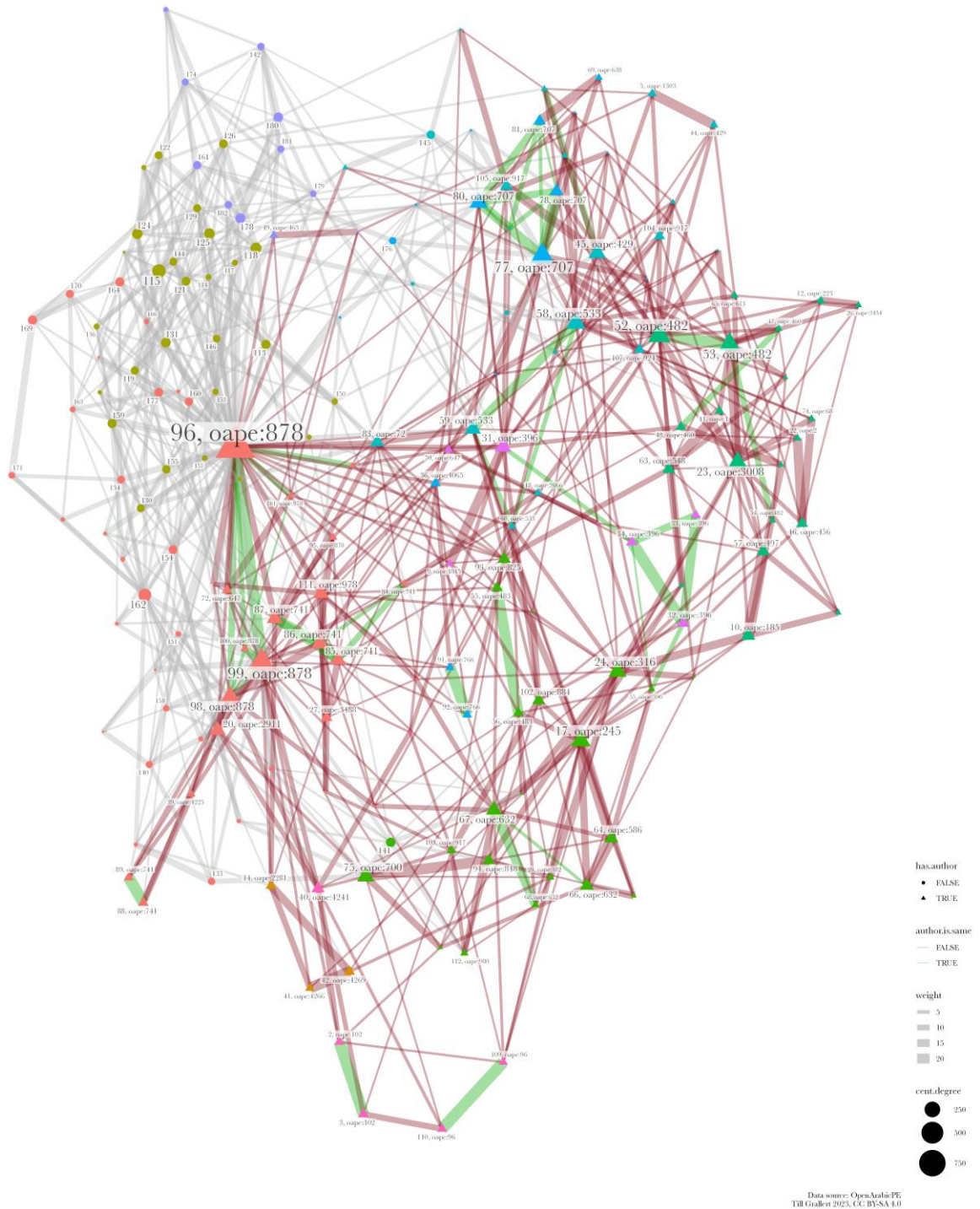


Fig. 8: Network plot of stylometric similarity for anonymous articles in *al-Muqtabas* from *data set 2* and all attributed articles from *data set 1*. Note how the anonymous articles cluster on the left.

Stylometric similarity
 MFW: 200-500. Increment: 100.
 Distance measure: Eder's simple delta.
 Sample length: 2500 W.

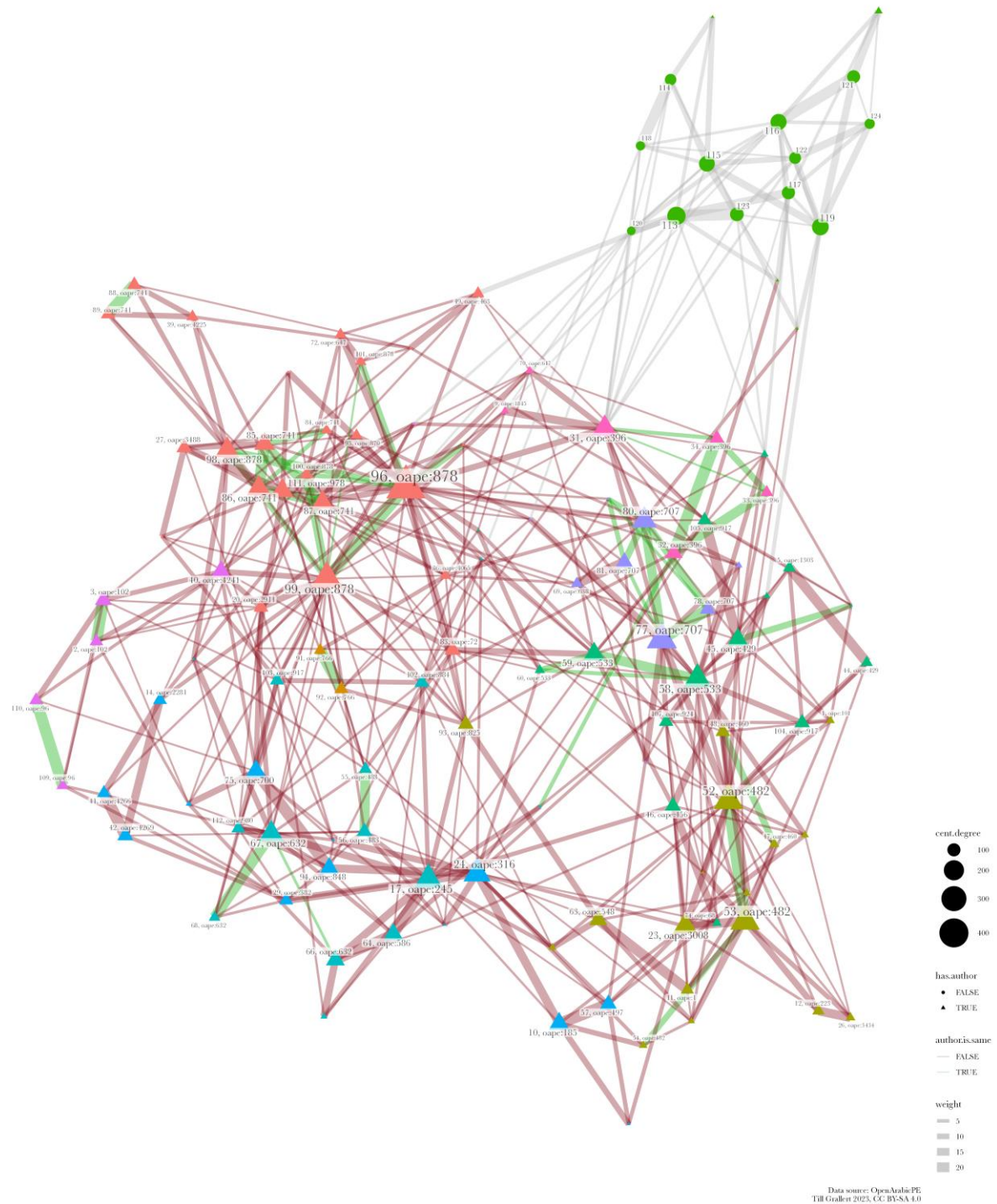


Fig. 9: Network plot of stylometric similarity for anonymous articles in *Lughat al-‘Arab* from *data set 2* and all attributed articles from *data set 1*. Note how the anonymous articles cluster in the top right.

5. Conclusion and future work

In this short paper and based on systematic stylometric analyses of the periodical texts themselves, I have empirically tested the hypothesis that the publishers of Arabic journals from late Ottoman Greater Syria and Iraq authored all or at least the majority of the anonymous articles in their periodicals. I was

able to falsify the hypothesis for the two journals *al-Muqtabas* from Cairo and, later, Damascus, and *Lughat al-‘Arab* from Baghdad for which we have sufficient textual data. This empirical evidence from computational stylistics compounds circumstantial evidence for *al-Muqtabas*, such as its publisher Muḥammad Kurd ‘Alī’s prolonged absences from the place of publication, which I have briefly discussed elsewhere.¹⁸

Future work will have to focus on identifying potential candidates for authorship as well as building the necessary data sets for such candidates and for stylometric analyses. We will also have to significantly expand the scope of analyses by systematically digitising a representative sample of Arabic periodicals (compare fig. 1).

6. Bibliography

- Ayalon, Ami, *The press in the Arab Middle East: a history*, New York 1995.
- Benatti, Francesca / King, David, *Hidden Authors and Reading Machines: Investigating 19th-century authorship with 21st-century technologies*, in: SHARP 2017: Technologies of the book, Canada 2017.
- Burrows, John, ‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship, in: *Literary and Linguistic Computing* 17 (2002), pp. 267–287.
- Czygan, Christiane, *Zur Ordnung des Staates: jungosmanische Intellektuelle und ihre Konzepte in der Zeitung Hürriyet (1868-1870)*, Berlin 2012.
- Eder, Maciej, Does size matter? Authorship attribution, small samples, big problem, in: *Literary and Linguistic Computing* 30 (2015), pp. 167–182.
- Eder, Maciej / Rybicki, Jan / Kestemont, Mike, *Stylometry with R: A Package for Computational Text Analysis*, in: *The R Journal* 8 (2016), pp. 107–121.
- Ezzerelli, Kais, *The publicist and his newspaper in Syria in the era of the Young Turk Revolution, between reformist commitment and political pressures: Muhammad Kurd Ali and al-Muqtabas (1908-17)*, in: Anthony Gorman / Didier Monciaud (eds.), *The press in the Middle East and North Africa, 1850-1950: politics, social history and culture*, Edinburgh 2017, pp. 176–206.
- Grallert, Till, *Open Arabic Periodical Editions: A framework for bootstrapped scholarly editions outside the global north*, in: *Digital Humanities Quarterly* 16 (2022), "Minimal Computing", <https://web.archive.org/web/20230302132915/http://digitalhumanities.org/dhq/vol/16/2/000593/000593.html>.
- Grallert, Till, *Catch Me If You Can! Approaching the Arabic Press of the Late Ottoman Eastern Mediterranean through Digital History*, in: *Geschichte und Gesellschaft* 47 (2021), pp. 58–89.
- Grallert, Till, *Digitalisierung des Kulturerbes des globalen Südens*, in: Laura Busse et al. (eds.), *Clio Guide. Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften*, 3rd Ed., Berlin forthcoming.
- Grallert, Till, *Editing mundane texts across the digital divide: the case of Arabic periodicals from the late nineteenth-century Eastern Mediterranean*, in: Constance Crompton et al. (eds.), *Digital Humanities in Practice*, London forthcoming.
- Kestemont, Mike / Martens, Gunther / Ries, Thorsten, *A Computational Approach to Authorship Verification of Johann Wolfgang Goethe’s Contributions to the Frankfurter gelehrte Anzeigen (1772–73)*, in: *Journal of European Periodical Studies* 4 (2019), pp. 115–43.

¹⁸ Grallert, *Catch me if you can!*, pp. 84–85.

- Khayat, Nicole, What's in a name? Perceptions of authorship and copyright during the Arabic nahda, in: *Nineteenth-Century Contexts* 41 (2019), pp. 423–440.
- Koppel, Moshe / Schler, Jonathan / Argamon, Shlomo, Computational methods in authorship attribution, in: *Journal of the American Society for Information Science and Technology* 60 (2009), pp. 9–26.
- Mestyan, Adam / Grallert, Till, *Jarā'id: A Chronology of Arabic Periodicals (1800-1929)*. 2020 Edition, 2020, <https://projectjaraid.github.io/> (19. 11. 2020).
- Mestyan, Adam / Grallert, Till, *A Chronology of Nineteenth-Century Periodicals in Arabic (1800-1900): A research tool*, 2012–2015, <https://web.archive.org/web/201604222071133/https://www.zmo.de/jaraid/> (19. 11. 2020).
- Mestyan, Adam / Grallert, Till/ et al., *Jarā'id: A chronology of Arabic periodicals (1800-1929)*, Zenodo, <https://web.archive.org/web/20230512084133/https://zenodo.org/record/4399240>.
- Pedersen, Thomas Lin, *tidygraph: A Tidy API for Graph Manipulation*.
- Pedersen, Thomas Lin, *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*.
- R Core Team, *R: A language and environment for statistical computing*, manual, R Foundation for Statistical Computing, Vienna (2018).
- Romanov, Maxim / Grallert, Till, Establishing parameters for stylometric authorship attribution of 19th-century Arabic books and periodicals, in: *Book of Abstracts, Tokyo 2022*, pp. 346–348.
- Seikaly, Samir, *Damascene Intellectual Life in the Opening Years of the 20th Century: Muhammad Kurd 'Ali and Al-Muqtabas*, in: Marwan Rafat Buheiry (ed.), *Intellectual Life In The Arab East, 1890-1939*, Beirut 1981, pp. 125–153.
- TEI Consortium, *TEI P5: Guidelines for Electronic Text Encoding and Interchange*, Zenodo.
- Wickham, Hadley, *ggplot2: Elegant graphics for data analysis*, New York 2016.