Twenty-One* Pseudo-Chrysostoms and More

Authorship Verification in the Patristic World

Thibault Clérice, Anthony Glaise

2023-11

ALMAnaCH, Inria Paris; CESR, Université de Tours

Table of contents

Patristics & Chrysostom

Patristics ?

Vocabulary

Late Antiquity classification of the litterature based on the period (incl. christ. litt)

Patristics study of the texts and influence of Church Fathers

Church Father vs. ecclesiastic author Doctrinal influence and recognition by the institution

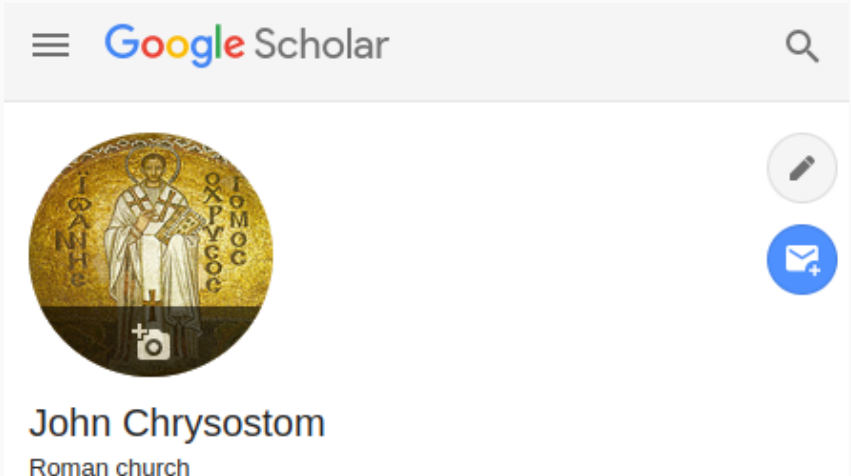
Christian litterature in the first millenium

- First in Ancient Greek, then in Latin (first texts are Paul's *Epistles* (51-55 CE))
- Chronological coverage 1st–8th century CE (more or less)
- "Golden age": 4th century

John Chrysostom

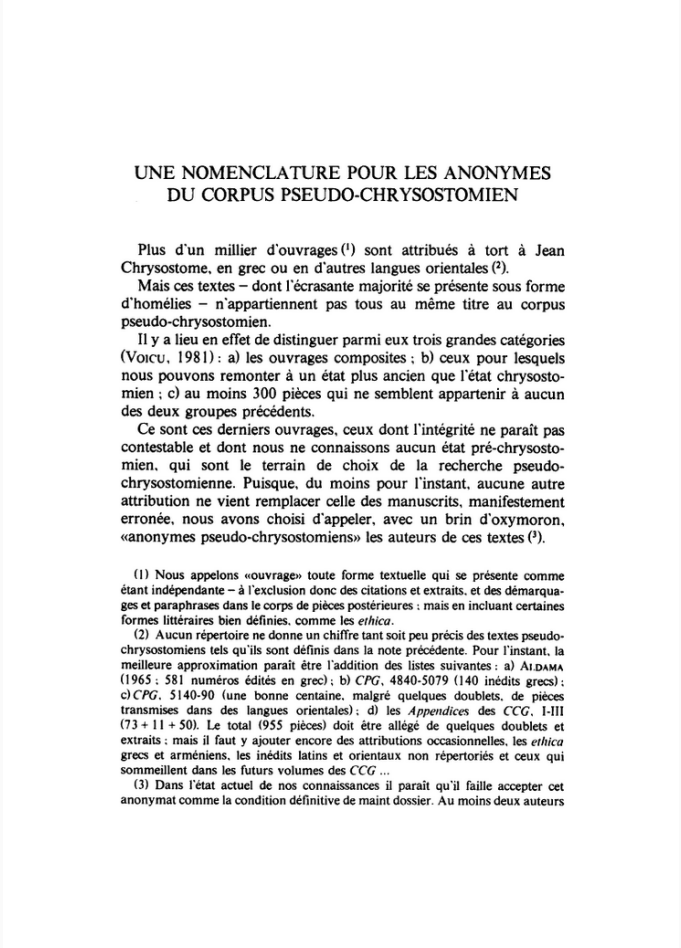
- Born in 349 CE
- Various high ranked positions in the Church (Antioch Diaconate and Priest, Archbishop of Constantinople)
- Exiled twice (fights between his supporters and others ended up costing the walls of a cathedral (burnt technically))
- Dies in an exile from exile (further banished because he had too much influence remaining in Constantinople)
- Wrote a lot, was very influential (very high h-index potential)

One way to present him, the first time it was introduced to me, is that Chrysostom is the **Greek-speaking equivalent of Augustine** (and I love this comparison).



The issue with writing a lot

- Around 300 works written by Chrysostom (which we are more or less sure of, 18 volumes of the *Patrologia Graeca*)
- S. .J. Voicu [4] estimates that a little more than 1000 works are wrongly attributed to Chrysostom (and as such are Pseudo-Chrysostomian), including Armenian and Latin translation.
- Some authors have been namely identified, such as Severian of Gabala (originally friendly with C, ended up judging him for exile) who wrote around 60 homelia ending up attributed to Chrysostom (irony ?)



UNE NOMENCLATURE POUR LES ANONYMES
DU CORPUS PSEUDO-CHRYSOSTOMIEN

Plus d'un millier d'ouvrages (1) sont attribués à tort à Jean Chrysostome, en grec ou en d'autres langues orientales (2).

Mais ces textes – dont l'écrasante majorité se présente sous forme d'homélies – n'appartiennent pas tous au même titre au corpus pseudo-chrysostomien.

Il y a lieu en effet de distinguer parmi eux trois grandes catégories (Voicu, 1981) : a) les ouvrages composites ; b) ceux pour lesquels nous pouvons remonter à un état plus ancien que l'état chrysostomien ; c) au moins 300 pièces qui ne semblent appartenir à aucun des deux groupes précédents.

Ce sont ces derniers ouvrages, ceux dont l'intégrité ne paraît pas contestable et dont nous ne connaissons aucun état pré-chrysostomien, qui sont le terrain de choix de la recherche pseudo-chrysostomienne. Puisque, du moins pour l'instant, aucune autre attribution ne vient remplacer celle des manuscrits, manifestement erronée, nous avons choisi d'appeler, avec un brin d'oxymoron, «anonymes pseudo-chrysostomiens» les auteurs de ces textes (3).

(1) Nous appelons «ouvrage» toute forme textuelle qui se présente comme étant indépendante – à l'exclusion donc des citations et extraits, et des démarquages et paraphrases dans le corps de pièces postérieures ; mais en incluant certaines formes littéraires bien définies, comme les *ethica*.

(2) Aucun répertoire ne donne un chiffre tant soit peu précis des textes pseudo-chrysostomiens tels qu'ils sont définis dans la note précédente. Pour l'instant, la meilleure approximation paraît être l'addition des listes suivantes : a) Aldama (1965 ; 581 numéros édités en grec) ; b) *CPG*, 4840-5079 (140 inédits grecs) ; c) *CPG*, 5140-90 (une bonne centaine, malgré quelques doublets, de pièces transmises dans des langues orientales) ; d) les *Appendices* des *CCG*, I-III (73 + 11 + 50). Le total (955 pièces) doit être allégé de quelques doublets et extraits ; mais il faut y ajouter encore des attributions occasionnelles, les *ethica* grecs et arméniens, les inédits latins et orientaux non répertoriés et ceux qui sommeillent dans les futurs volumes des *CCG* ...

(3) Dans l'état actuel de nos connaissances il paraît qu'il faille accepter cet anonymat comme la condition définitive de maint dossier. Au moins deux auteurs

1000 pseudo-chrysostomian works ?

In 1981, S. J. Voicu provides a summary of the status of attribution of pseudonymous work into anonymous clusters (PCn). It covers around 21 PC, for which some hypothesis date back to the 17th century, and features argument based on various features: style, theological, extra-textual (references or cites something posterior to Chrysostom) and sequentiality (texts clearly continuing each others).

Cluster	Number of texts	Original Hypothesis	Confidence	Additional analysis	Confidence	Argument type
1	2	Montfaucon		Altendorf	Refuted	theological
2	2	Montfaucon		Voicu	Confirmed	continuity
3	3	Montfaucon		Voicu	Refuted	
4	7	Montfaucon		Voicu	Low	
5	3	Montfaucon		Voicu	Partially refuted	
6	3	Montfaucon		Voicu	Low	
7	7	Montfaucon		Voicu	Possible	
8	2	Montfaucon		Voicu	Refuted	stylistic
9	5	Marx		Voicu	Refuted	
10	2	Weyer				
11	3	Nautin				
12	2*	Liébart				theologic
13	3	Leroy		Voicu	Refuted	
14	4	Voicu	High			
15	5	Voicu	Mostly high			
16	5	Voicu	High	Wenger	Partially Possible	
17	2	Voicu	High			
18	2	Voicu	Possible			stylistic,theologic
19	2*	Rilliet	High			
20	5	Datema	High	Voicu		
20b	12	Datema	Possible	Voicu		
21	2	Voicu	High			extra-textual

Authorship verification

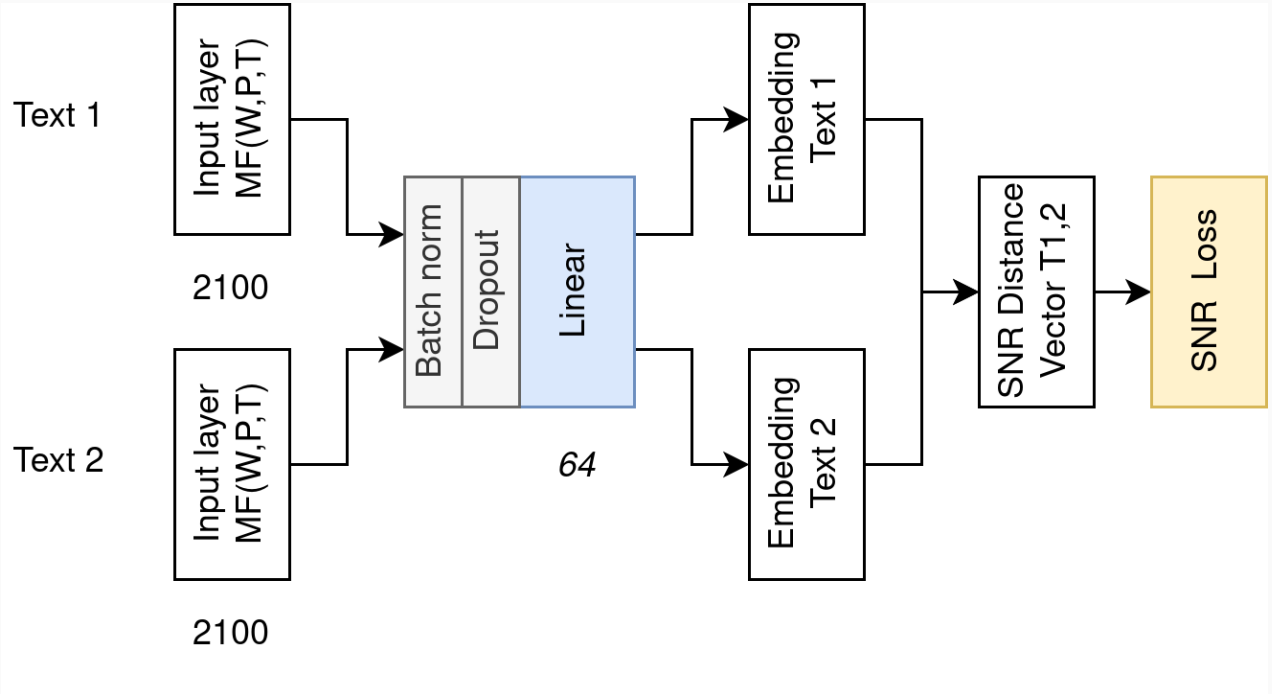
Authorship verification

We have very small samples of texts (between 2 and 7 texts per Pseudo-Chrysostom, one exception with 20b being 12 texts): so classification is nearly out of the question (hard to test accuracy on a two text sample).

We propose to approach this with authorship verification, and specifically, we want to merge approaches from NLP and Computer Vision (using Siamese Networks, specific distance) and features usually used in DH (avoid to use complete sentences, prefer style markers that are (nearly) atopical such as function words, POS 3-grams, affixes [2].

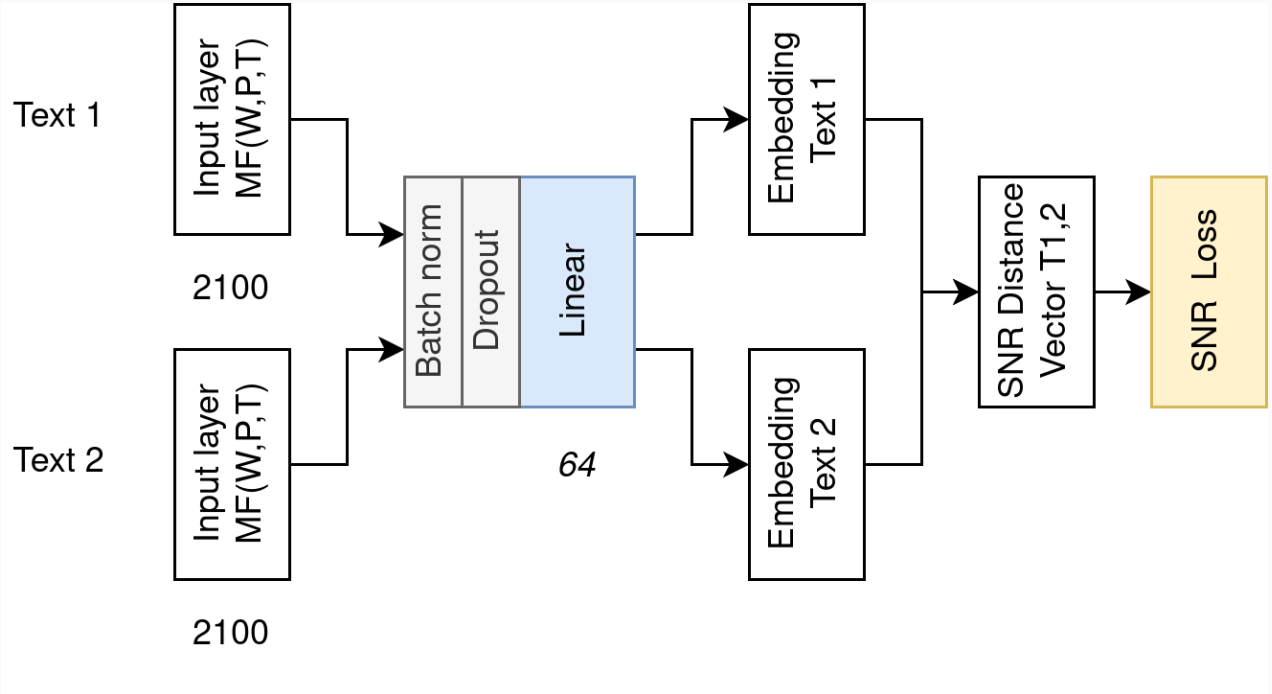
Siamese network ?

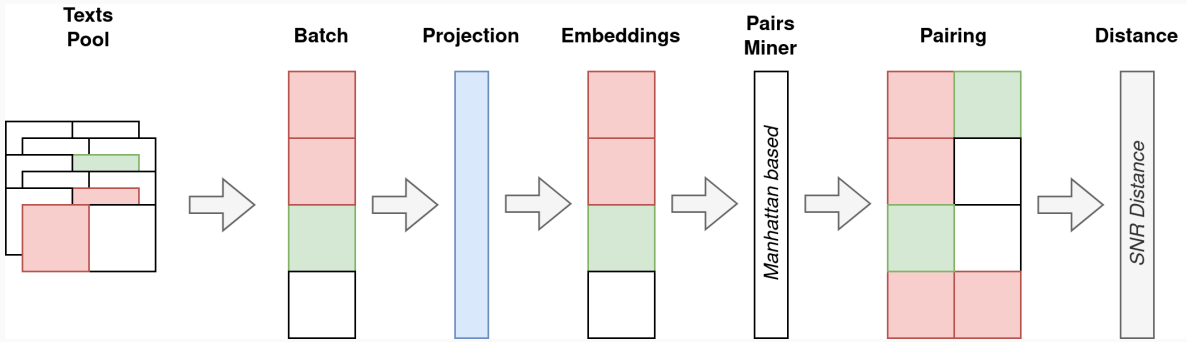
Architecture of our models: vectors are passed in parallel in the embedding projection layer and a distance vector is then computed.



Siamese network ?

Architecture of our models: vectors are passed in parallel in the embedding projection layer and a distance vector is then computed.





Signal-to-noise ratio, yet another distance ?

If we consider T_i as the features representing a text, $f(T_i)$ a function reducing it as a vector E_i, we can optimize f towards the enhancement of the authorial signal (or stylome) A_y and the reduction of noise N_i around it[1] such that $E_i = N_i + A_y$ where $N_i \approx 0$ if f converges.

Given two texts i and j, if they are from the same author y, $E_i - E_j = N_i + A_y - N_j - A_y = N_i - N_j$. If the function f converges, given $g(T_i, T_j) = var(f(T_i) - f(T_j))$, then $g(T_i, T_j) \approx 0$. As a result, considering the signal-to-noise ratio $SNR(E_i, E_j) = \frac{var(E_i)}{g(E_i, E_j)}$, then the distance $SNRD = \frac{1}{SNR} \approx 0$ for the same author.

If $E_k = N_k + A_z$, $g(E_i, E_k) = var(N_i + A_y - N_k - A_z)$, given $N \approx 0$, f should optimize such that $var(A_y - A_z) > var(A_y)$ and as such, $SNRD(E_i, E_k) > 1$[2]

[1] The noise could then be considered as containing remaining information about the genre, the period, the topic, etc.

[2] Many thanks to Pierre Mercuriali for our discussions which led to this slide.

Parameters finding

Features

Input concatenation of

1. 0, 100 and 200 most frequent POS-trigrams using a Greek BERT [3][3]
2. 0, 250, 500, 750, 1000 most frequent words[4]
3. 0, 250, 500, 750, 1000 most frequent affixes.

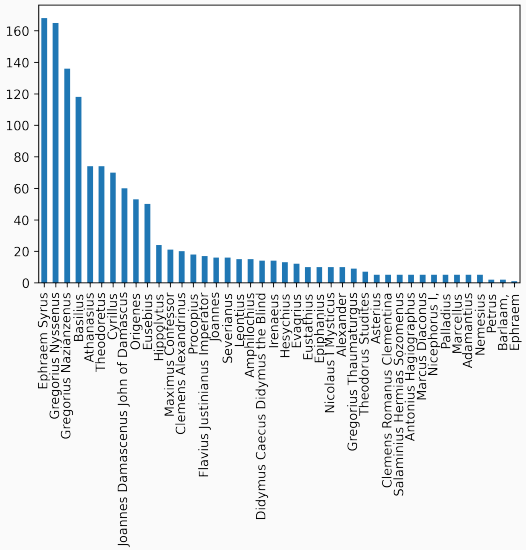
Distance SNR-D [5], Manhattan (L1), Euclidean (L2)

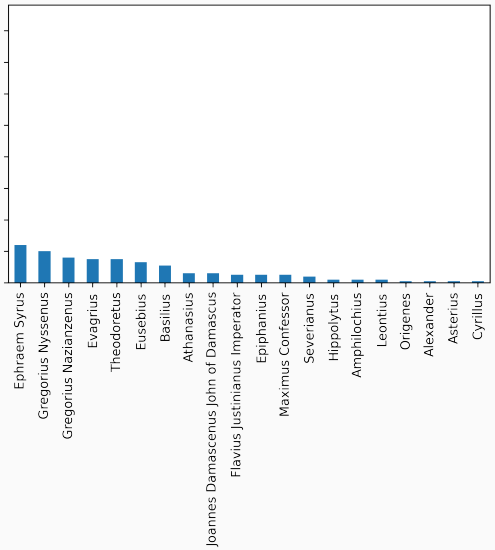
Hyper-params: Adam, $lr = 1e^4$, embedding size of 64, batch size of 64, .3 dropout, class sampling of 2, and a minimum of 100 epochs for training. Early stopping 20 epochs using dev loss

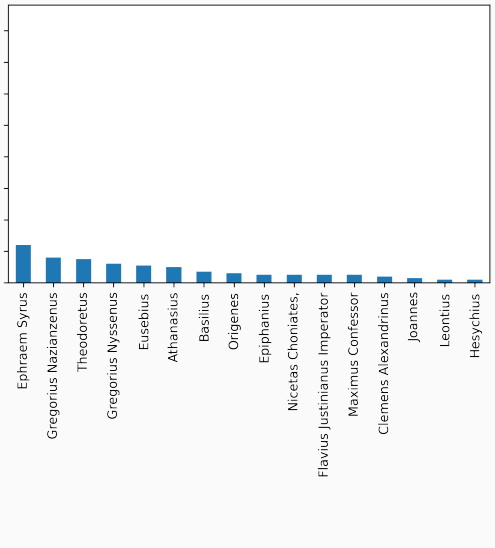
Trainable on CPU "quickly", on GPU in a shorter amount of time..

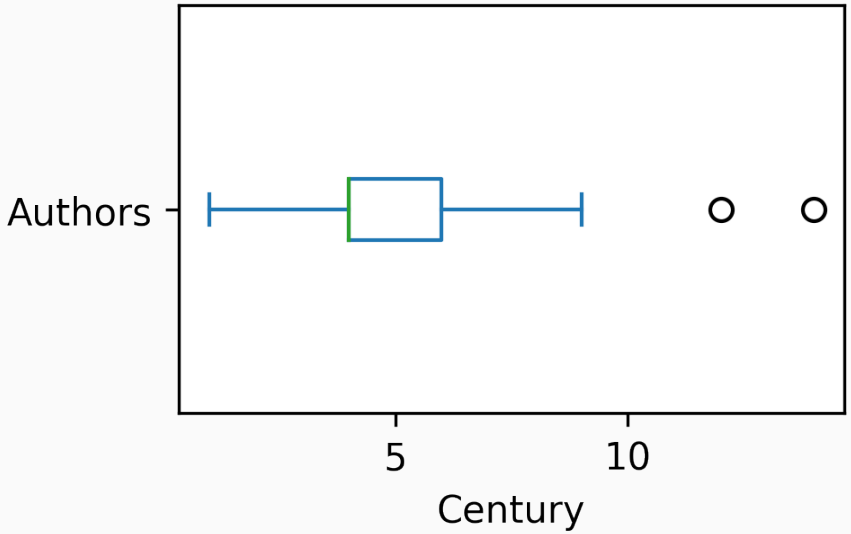
[3] All most frequent classes on a large corpus of Christian litterature
[4] We maxed our value at the size of our sample, 1000 words.

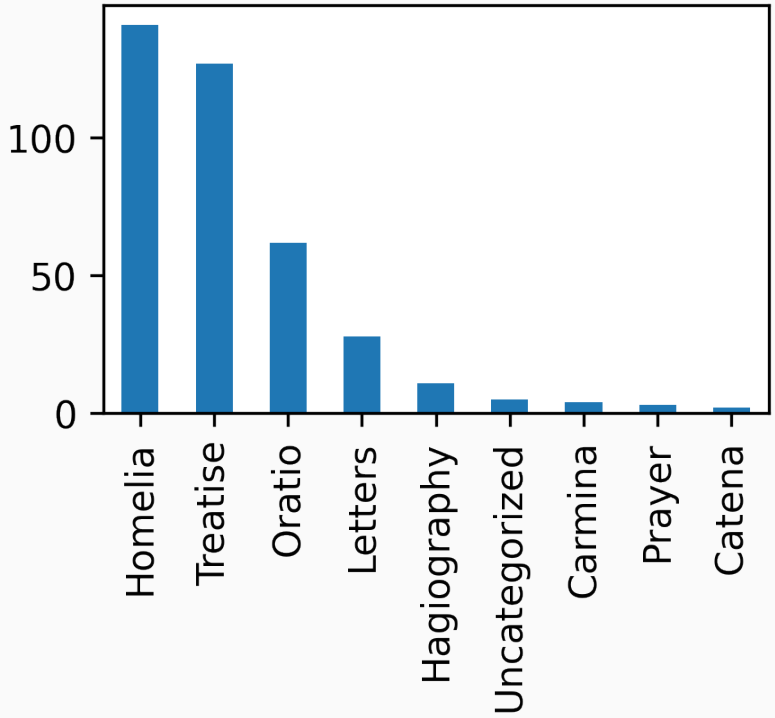
Corpus


160
140
120
100
80
60
40
20
0
Ephraem Syrus
Gregorius Nyssenus
Gregorius Nazianzenus
Basilius
Athanasius
Theodoretus
Cyrillus
Joannes Damascenus John of Damascus
Origenes
Eusebius
Hippolytus
Maximus Confessor
Clemens Alexandrinus
Procopius
Flavius Justinianus Imperator
Joannes
Severianus
Leontius
Amphilochius
Didymus Caecus Didymus the Blind
Hesychius
Evagrius
Eustathius
Epiphanius
Nicolaus Alexandrides
Gregorius Thaumaturgus
Theodorus Studites
Asterius
Clemens Romanus Clementina
Salamius Hermias Sozomenus
Antonius Hagiographus
Marcus Diaconus
Nicephorus I.
Palladius
Marcellus
Adamantius
Nemesius
Petrus
Barlaam,
Ephraem


Ephraem Syrus
Gregorius Nyssenus
Gregorius Nazianzenus
Evagrius
Theodoretus
Eusebius
Basilius
Athanasius
Joannes Damascenus John of Damascus
Flavius Justinianus Imperator
Epiphanius
Maximus Confessor
Severianus
Hippolytus
Amphilochius
Leontius
Origenes
Alexander
Asterius
Cyrillus


Ephraem Syrus
Gregorius Nazianzenus
Theodoretus
Gregorius Nyssenus
Eusebius
Athanasius
Basilius
Origenes
Epiphanius
Nicetas Choniates,
Flavius Justinianus Imperator
Maximus Confessor
Clemens Alexandrinus
Joannes
Leontius
Hesychius


Authors
5
10
Century


100
50
0
Homelia
Treatise
Oratio
Letters
Hagiography
Uncategorized
Carmina
Prayer
Catena

Results

Mean of standard-deviation across the parameters sweep depending on distances.

Distance	Dev Mean	±	Test Mean	±
L2	1.04	0.55	1.82	1.02
Manhattan	1.09	0.59	1.84	1.29
STN	1.18	0.68	1.14	0.59

Fleiss Kappa using 1000 models as annotators on the test set.

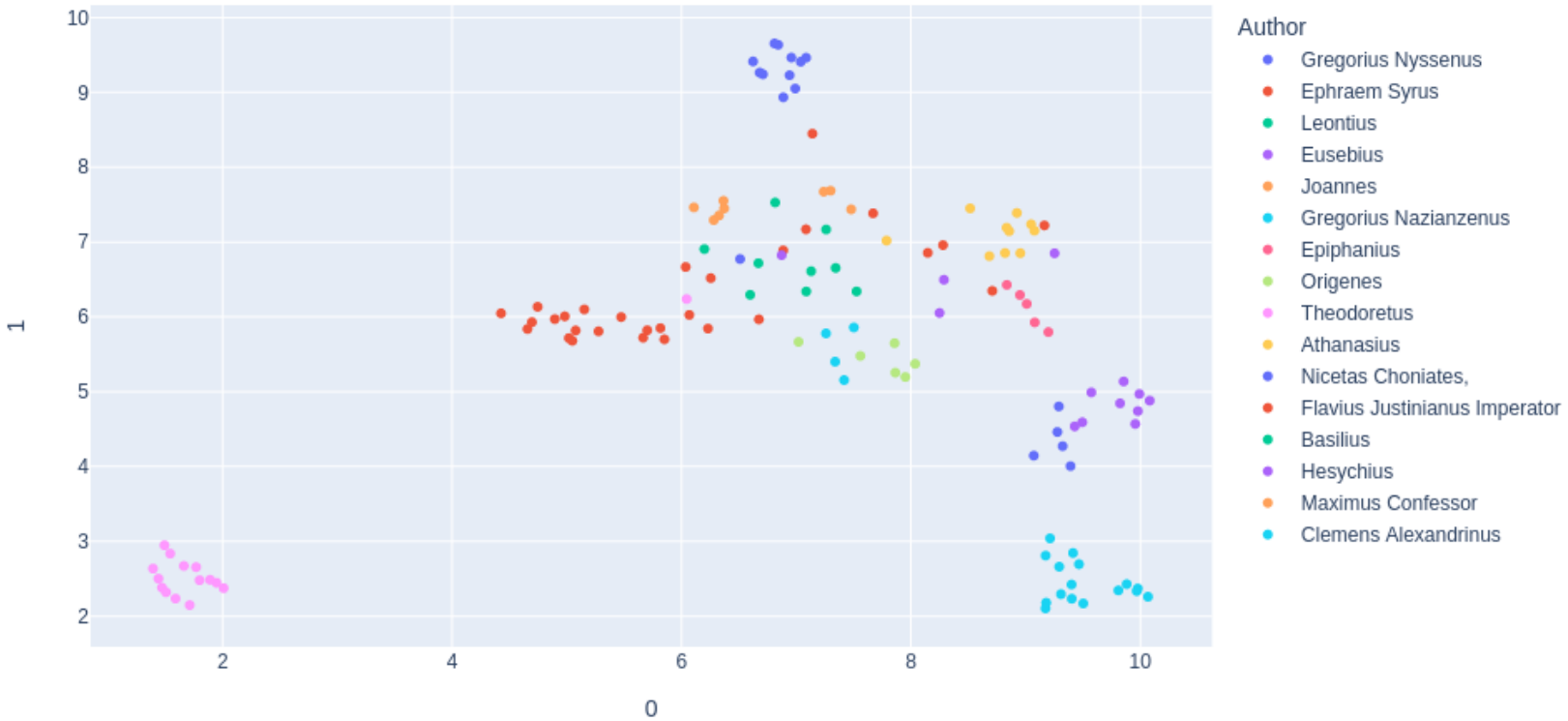
Development Precision	100%	99%	95%	90%	85%	80%
			Fleiss Kappa			
Manhattan	0.18	0.18	0.20	0.29	0.40	0.50
STN	0.50	0.60	0.69	0.72	0.72	0.72

Top 10 architectures

Parameters				Dev		Test	
POS	FW	Affixes	Distance	mean	std	mean	std
100	1000	1000	Manhattan	88.04	0.58	86.01	0.53
200	1000	750	Manhattan	87.63	0.83	85.88	2.03
100	750	1000	Manhattan	86.75	0.80	85.12	1.67
100	1000	750	STN	85.77	1.91	84.51	0.93
100	1000	1000	STN	87.03	0.65	84.42	1.31
100	1000	750	Manhattan	87.19	1.25	84.33	4.10
100	750	1000	STN	85.38	0.33	84.23	0.63
200	1000	750	STN	85.81	0.15	84.15	0.81
100	1000	500	Manhattan	86.85	1.44	83.68	4.48
200	750	750	Manhattan	86.64	0.51	83.63	1.08

- L2 is bad (what a surprise)
- Manhattan is more unstable in this setting
- Discrepancy between dev and test are lower for STN
- 100, 1000, 1000 seems to hit the spot

UMap


Author
Gregorius Nyssenus
Ephraem Syrus
Leontius
Eusebius
Joannes
Gregorius Nazianzenus
Epiphanius
Origenes
Theodoretus
Athanasius
Nicetas Choniates,
Flavius Justinianus Imperator
Basilius
Hesychius
Maximus Confessor
Clemens Alexandrinus
1
0
10
9
8
7
6
5
4
3
2
2
4
6
8
10

Evaluation on the 21* pseudo-Chrysostoms

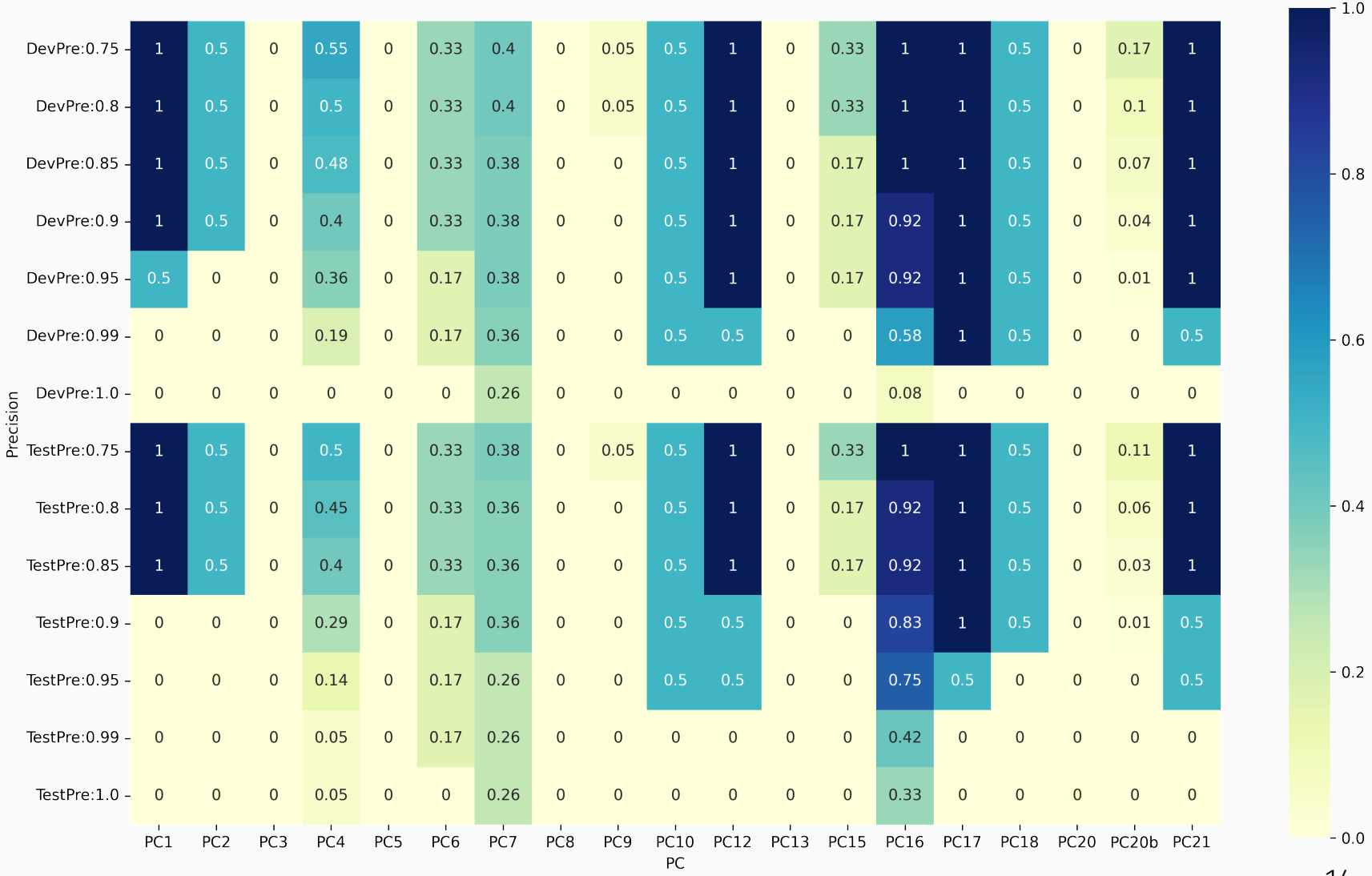
Making a decision

The model does not output a probability (future work ?) but does provite a distance.

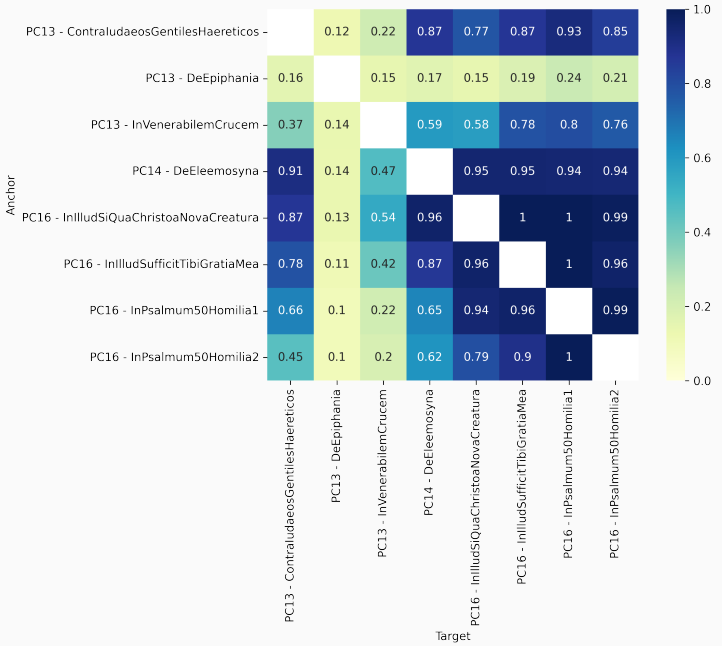
What we do:

- Distance is deemed positive based on the given precision on the dev or test set
- We look for each PC what's the percentage of pairs
- We plot.

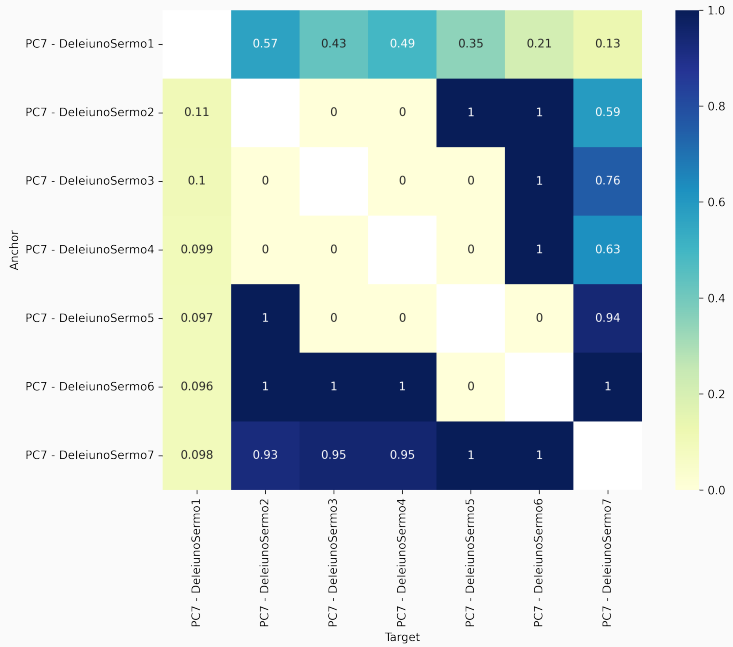
Overall


Precision
DevPre:0.75 1 0.5 0 0.55 0 0.33 0.4 0 0.05 0.5 1 0 0.33 1 1 0.5 0 0.17 1
DevPre:0.8 1 0.5 0 0.5 0 0.33 0.4 0 0.05 0.5 1 0 0.33 1 1 0.5 0 0.1 1
DevPre:0.85 1 0.5 0 0.48 0 0.33 0.38 0 0 0.5 1 0 0.17 1 1 0.5 0 0.07 1
DevPre:0.9 1 0.5 0 0.4 0 0.33 0.38 0 0 0.5 1 0 0.17 0.92 1 0.5 0 0.04 1
DevPre:0.95 0.5 0 0 0.36 0 0.17 0.38 0 0 0.5 1 0 0.17 0.92 1 0.5 0 0.01 1
DevPre:0.99 0 0 0 0.19 0 0.17 0.36 0 0 0.5 0.5 0 0 0.58 1 0.5 0 0 0.5
DevPre:1.0 0 0 0 0 0 0 0.26 0 0 0 0 0 0 0.08 0 0 0 0 0
TestPre:0.75 1 0.5 0 0.5 0 0.33 0.38 0 0.05 0.5 1 0 0.33 1 1 0.5 0 0.11 1
TestPre:0.8 1 0.5 0 0.45 0 0.33 0.36 0 0 0.5 1 0 0.17 0.92 1 0.5 0 0.06 1
TestPre:0.85 1 0.5 0 0.4 0 0.33 0.36 0 0 0.5 1 0 0.17 0.92 1 0.5 0 0.03 1
TestPre:0.9 0 0 0 0.29 0 0.17 0.36 0 0 0.5 0.5 0 0 0.83 1 0.5 0 0.01 0.5
TestPre:0.95 0 0 0 0.14 0 0.17 0.26 0 0 0.5 0.5 0 0 0.75 0.5 0 0 0 0.5
TestPre:0.99 0 0 0 0.05 0 0.17 0.26 0 0 0 0 0 0 0.42 0 0 0 0 0
TestPre:1.0 0 0 0 0.05 0 0 0.26 0 0 0 0 0 0 0.33 0 0 0 0 0
PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9 PC10 PC12 PC13 PC15 PC16 PC17 PC18 PC20 PC20b PC21
PC
1.0 0.8 0.6 0.4 0.2 0.0

Zooming in



(a) Precision threshold matrix for Pseudo-Chrysostoms 13, 14, and 16.



(b) Pseudo-Chrysostom 7

Conclusion

Conclusion

- It works, maybe more tuning and work around the question of prediction as a probability would be interesting.
- It can serves as a confirmation tool, but should not be used as THE single tool in an analysis (too much false negatives).
- We mostly align with Voicu and its peers.
- We have two very *bizarre* results:
  - We have more or less confirmation of a single author behind PC1, which was refuted by Altendorf [1].
  - We are completely unable to confirm PC20, which was a strong hypothesis of Voicu.

Questions?

https://github.com/PonteIneptique/Chryso-Voicu

References i

H. Altendorf.
Untersuchungen zu Severian von Gabala.
PhD thesis, Tübingen, 1957.

J.-B. Camps, T. Clérice, and A. Pinche.
Noisy medieval data, from digitized manuscript to stylometric analysis: Evaluating Paul Meyer's hagiographic hypothesis.
Digital Scholarship in the Humanities, 36(Supplement_2):ii49–ii71, 2021.

P. Singh, G. Rutten, and E. Lefever.
A pilot study for BERT language modelling and morphological analysis for ancient and medieval Greek.
In *5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature,*

References ii

co-located with EMNLP 2021, pages 128–137. Association for Computational Linguistics, 2021.

S. J. Voicu.
Une nomenclature pour les anonymes du corpus pseudo-chrysostomien.
Byzantion, 51(1):297–305, 1981.

T. Yuan, W. Deng, J. Tang, Y. Tang, and B. Chen.
Signal-To-Noise Ratio: A Robust Distance Metric for Deep Metric Learning.
In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4810–4819, 2019.