



Classical and Digital Approaches to Greek Palaeography
PATMOS · GREECE 14–18 SEPTEMBER 2026

Al and stemma codicum

Elpida Perdiki, PhD – Democritus University of Thrace



Sarah Connor watching you
use AI for everything



Ways to apply AI to stemmatics

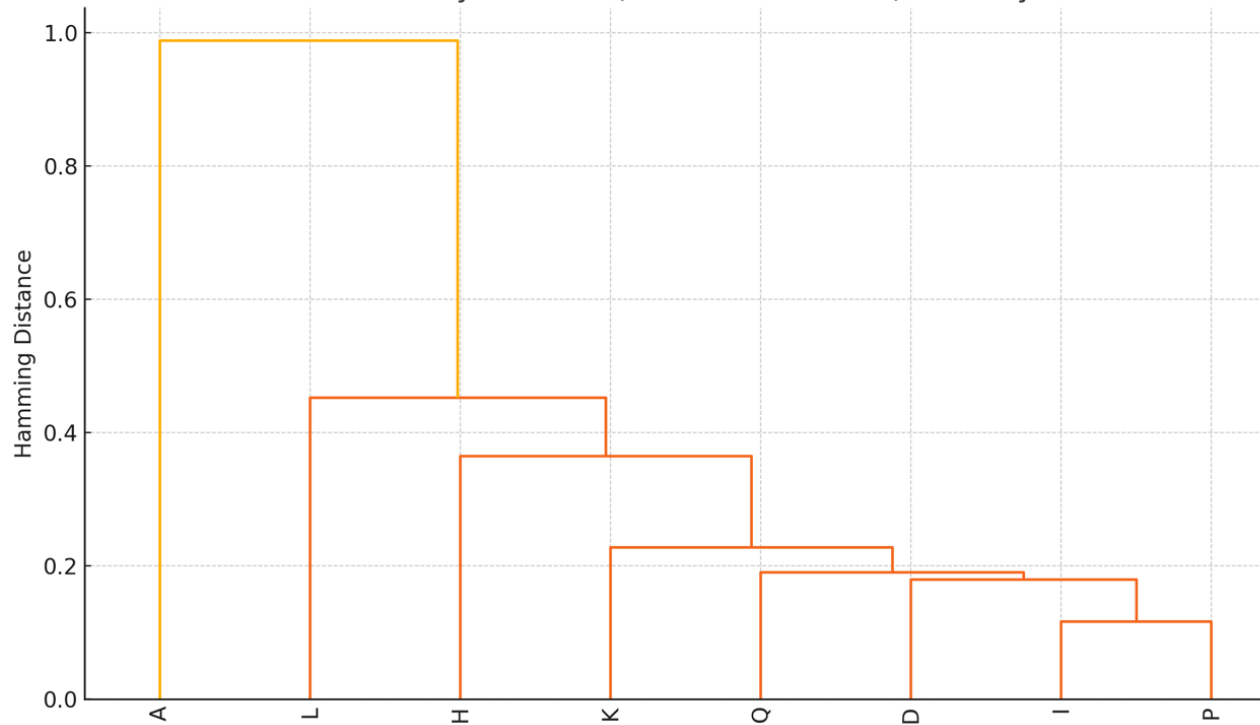
- **The obvious:** direct generation of stemma codicum
- **The medium:** text extraction and post-correction
- **The pro:** stemma codicum detection and (possibly) reconstruction

Bozzi, A., & Corradini, M. S. (2002). New trends in philology: a computational application for textual criticism. *Euphrosyne*, 30, 267-285.

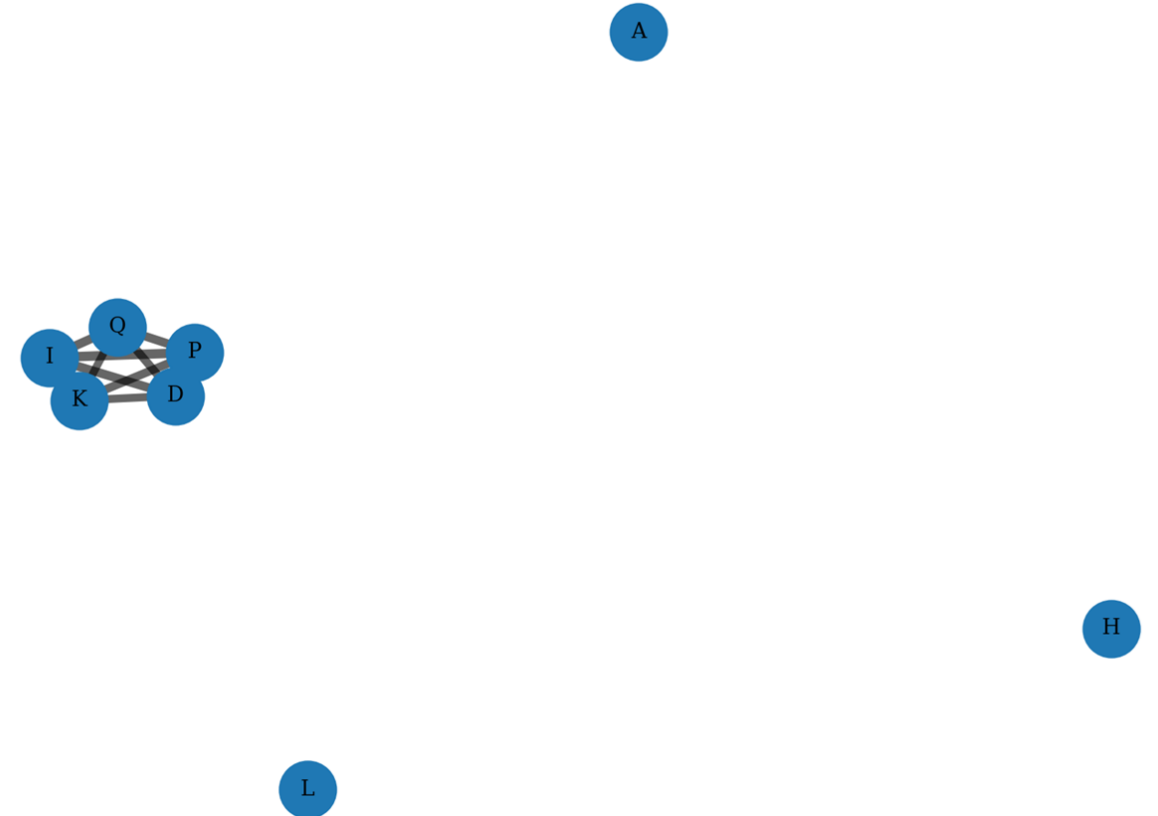
**please generate a stemma codicum from
these manuscripts**

Can I trust you, ChatGPT?

Preliminary Stemma (Lachmannian Tree) - Homily 1



Contamination-Aware Manuscript Network - Homily 1



Manuscript **collation** & **human-in-the-loop**

pros

- similarity metrics and metadata
- involve human evaluation
- stemma proposal
- Graph Theory & bioinformatics

cons

- general witness similarity
- non available training data

Mancini, L. (2021). An open collaborative system for Digital Stemmatology. *TAIS 2021 Proceedings*. 23.

Computer vision: Embeddings vs ATR

- ResNet-18
- Hungarian Algorithm
- clustering glyphs isolation with visual embeddings
- groups of identical letterforms
- frequency per witness counting
- high shared glyph profile = scribal proximity

Hoenen, A. (2026). The Devil is in the Details – From OCR for Old Church Slavonic to Purely Visual Stemma Reconstruction. arXiv preprint arXiv:2604.11724. <https://aisel.aisnet.org/itais2021/23>

please **extract text from these
manuscripts**

AI as a medium

- ATR (Automatic Text Recognition)
- text collation
- witness dating
- ATR errors post-correction

De Gregorio, G., Ferretti, L., Pena, R. C., Marthot-Santaniello, I., Konstantinidou, M., & Pavlopoulos, J. (2024, August). A New Framework for Error Analysis in Computational Paleographic Dating of Greek Papyri. In *International Conference on Document Analysis and Recognition* (pp. 102-118). Cham: Springer Nature Switzerland.

Evangelatos, A., Palaiologos, K., Gatos, B., Kaddas, P., Christopoulou, A., & Katsouros, V. (2025, August). Old greek ocr result correction using llms. In *Proceedings of the 2025 ACM Symposium on Document Engineering* (pp. 1-4).

Kahle, P., Colutto, S., Hackl, G., & Mühlberger, G., Transkribus - A Service Platform for Transcription, Recognition and Retrieval of Historical Documents. 2017 *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*, Kyoto: IEEE, 2017:19–24. Available at: <https://doi.org/10.1109/ICDAR.2017.307>

Kiessling, B., Tissot, R., Stokes, P., & Ezra, D. S. B., eScriptorium: An Open Source Platform for Historical Document Analysis. 2019 *International Conference on Document Analysis and Recognition Workshops (ICDARW)*, 2019:19–19. Available at: <https://doi.org/10.1109/ICDARW.2019.10032>.

Pavlopoulos, J., Konstantinidou, M., Perdiki, E., Marthot-Santaniello, I., Essler, H., Vardakas, G., & Likas, A. (2024). Explainable dating of greek papyri images. *Machine Learning*, 113(9), 6765-6786.

Perdiki, E. (2023). Preparing big manuscript data for hierarchical clustering with minimal HTR training. *Journal of Data Mining & Digital Humanities*, (Sciences of Antiquity and digital humanities).

Ströbel, Phillip Benjamin, Martin Volk, Simon Clematide, Raphael Schwitter, Tobias Hodel, και David Schoch. 'Evaluation of HTR models without Ground Truth Material'. *Proceedings of the Thirteenth Language Resources and Evaluation Conference, European Language Resources Association*, 03/2022, 4395–404. <https://aclanthology.org/2022.lrec-1.467>

Tsochatzidis, L., Symeonidis, S., Papazoglou, A., & Pratikakis, I., HTR for Greek Historical Handwritten Documents. *Journal of Imaging*, 2021;7(12):260. <https://doi.org/10.3390/jimaging7120260>.

Περδίκη, Ε. (2026). Διαχείριση μεγάλου όγκου χειρογράφων σε εκδόσεις μεσαιωνικών κειμένων: Ψηφιακά εργαλεία και μεθοδολογία χρήσης τους [Handling big manuscript data; digital tools and methodology for editing Byzantine texts] [Διδακτορική Διατριβή]. Δημοκρίτειο Πανεπιστήμιο Θράκης.

**please detect stemma codicum in this
edition**

Computer vision and stemma codicum

- VLMs (Vision-Language Models)
- detection of *stemma codicum* in critical editions
- Machine-readable graphs
- metadata detection

Bizais-Lillig, M., Camps, J. B., Ferraris, O., Gaglione, E., Garnier, M., & Godreau, U. (2026, September). Automated detection and transcription of stemmata codicum from digitised publications. In *European Association for Digital Humanities Conference 2026*.

please don't ruin my text

Limiting limitations



LLMs (Large Language Models) are simply probabilistic sequence predictors



NLP, LLMs, or agentic pipelines for ATR post-correction might eliminate the manuscript tradition (Leitfehler)



stemma reconstruction is pattern learning in manuscript tradition, not a new critical edition



glyph clustering has also an error rate and the risk of losing three important aspects:

Historical vs visual distance: same text and different script style witnesses might be related

Scribal identity vs textual tradition: different texts and the same scribe is not a textual relation

Shapes vs Errors: shared distinctive errors (Leitfehler), not visually shared characters



Print design: Josef Mounser

‘Let Ggge go, let her go,’
Eggg chuckles, ‘let her get
carried away by the
spurts of flying electrons.’

Calvino Italo. 1971. *The distance of the
moon.*