Editing Glosses as Networks

Exploring the Explorative Edition

Peter Boot, Evina Steinová

Keywords

Glosses, editing of glosses, Isidore of Seville, network-based edition, network visualizations, explorative edition, distant reading.

Abstract

Creators of digital editions have historically been among the early adopters of computational tools in the humanities. It is surprising, therefore, that digital editions have been very slow in accepting the distant reading paradigm, which has been widely embraced as a fruitful approach for studying textual collections. In this article we make the case that digital scholarly editions should become explorative editions: while maintaining philological accuracy, they can and should include interactive tools that help the reader study and work through the edition's riches. We use the example of the early medieval glosses to the first book of Isidore's *Etymologiae* and argue that a network representation is better suited than the traditional stemma to represent the relations among manuscripts that carry these glosses. Then we proceed to show that an edition of these glosses can use interactive network graphs to provide high-level overviews of the information that the edition contains. The graphs provide access points into the edition as well as ways to explore the relations among the manuscripts, Isidore's text and the glosses. We also discuss several examples where interactive network graphs could help in the study of other types of texts.

1 Introduction

This article presents a proof-of-concept explorative edition of the glosses to the first book of Isidore of Seville's *Etymologiae*. The edition contains embedded interactive network graphs that help access its scholarly contents. In doing so, we are making two claims: first, that in editing organic corpora of glosses a traditional stemma may not be workable and a network model is a better alternative; and second, that by and large digital scholarly editions have up to now failed to embrace the distant reading paradigm and this is limiting their value and usability.

In section 2, we argue that since organic corpora of glosses do not behave as literary or narrative texts, their transmission cannot be reconstructed using a traditional approach. Nevertheless, glosses in one manuscript can be usually related to other manuscripts and it can

be shown that certain sets of glosses appearing in multiple manuscripts (clusters) represent relics of transmission. For this reason, rather than a stemma, we assume a network where the links (edges) between the manuscripts (nodes) are constructed based on the glosses these manuscripts share. In one of our network displays, we show how these gloss clusters define relations between manuscripts.

In section 3, we note that scholarly editors were among the first humanities scholars to embrace the digital medium. The scholarly edition was a hypertextual construct before hypertext existed and editors were quick in taking advantage of the new possibilities. Prominent creators of digital editions have also called for analytical tools to be included in the digital edition (Robinson 2004; Siemens et al. 2005). These calls, however, have mostly gone unheeded. Indeed, while 'distant reading' has been embraced as a fruitful paradigm and rallying cry for many practitioners of Digital Humanities, digital editions seem to be mostly unaffected. Editors have continued on their path of creating philologically sound editions that can be searched and navigated but that often expose the reader to the risk of 'drowning by versions' (Dahlström 2000). They have usually not been very creative in the creation of (visual) tooling that can help explore and analyse the underlying material in the edition.

In section 4, we describe our edition of the glosses to the first book of the *Etymologiae*.[1] We describe how we used TEI to store the relevant objects and relations and how we present the main text, the glosses, manuscripts and clusters. We also show three interactive network graphs that facilitate the study of the material contained in the edition. In section 5, we argue that a similar approach can be used for many different types of text, and that this will enhance the usability of the digital edition. We present our conclusions in section 6.

2 Editing Glosses

2.1 Glosses as a product of medieval textual culture

While most scholarly editions of Latin texts remain grounded in the genealogical method connected with the name of the German philologue Karl Lachmann, it has been long observed that certain types of historical documents cannot be edited using this paradigm (Palumbo 2020). This is particularly true for non-literary texts (e.g., letters, recipes, instructions), texts that don't have well-defined author(s), non-linear or discontinuous texts (e.g., glossaries), textual forms that were transmitted other than by copying (e.g., orally), and for those classes of written material that defy a classification as iterations of a single text (e.g., collections). A notable example examined in this article are glosses, a well-known product of (not only) medieval Western written culture, which have been rarely edited critically (Teeuwen 2016). This article is concerned more specifically with editing the glosses to the first book of the *Etymologies* of Isidore of Seville, the most important medieval Latin encyclopaedia.

[1] The edition is available at https://db.innovatingknowledge.nl/edition/. The github repository holding the source code and data is at https://github.com/HuygensING/isidore-glosses.

A medieval Western gloss can be defined as a minute textual entity that clarifies or comments on a different text. Glosses tend to be preserved both separated from the text they comment on (as glossaries) and together with this text as marginalia surrounding it in a manuscript of this text (in annotated manuscripts). Here, we are concerned with the latter case. A layer of glosses in a medieval annotated manuscript may consist of hundreds or thousands of glosses and therefore outwardly resemble a continuous and sequential text. Nevertheless, even a dense layer of glosses is but a chain of glosses, since they remain discrete and self-sufficient: they can be added or removed, combined together, and their position and relative order can also be altered (Teeuwen 2008b). A layer of glosses, thus, never entirely ceases to be a collection of micro-texts (the glosses) and it is perhaps appropriate to consider it a kind of a pseudo-text with limited cohesion, integrity and sequentiality.

The textual reality of glossing can be compared in some respects to Lego: a corpus of glosses to an author, text, or genre of literature reconstructed by an editor resembles a box of Lego blocks. Blocks from this box are found in manuscript witnesses assembled into larger structures, the layers of glossing. In these witnesses, it is possible to always encounter layers that are significantly similar or whose differences can be explained as alterations to a prescribed design. However, it can also happen that the same blocks are assembled in distinctly different configurations. In theory, one can build two entirely different or significantly dissimilar objects from the same selection of glosses, a situation attested in glossaries (Dionisotti 1996). A good editor must pay attention to both the similarities, because they will notice the same blocks and sets of blocks time and again, and dissimilarities, because these blocks will appear in different positions. At the same time, they may encounter certain fixed components consisting of several blocks recurring unaltered in dissimilar larger structures. Indeed, the textual reality of glossing usually does not restrict itself to the gloss as the minimum textual unit (akin to a Lego block) and the layers of glossing as preserved in a specific manuscript witness as the unit of the highest degree (akin to a structure built from the blocks). Glosses tend to coalesce into various mid-level textual elements: clusters, batches, and sets, which served as intermediary building blocks of medieval glossators and often inform us about how the material was transmitted in the Middle Ages. A scholarly editor should ideally strive to capture all levels of textuality present in a corpus of glosses: the total corpus, the layers present in individual manuscript witnesses, individual glosses, as well as any clusters.

2.2 Problems posed by editing of glosses

The genealogical method has been principally designed for the editing of long, continuous, sequentially ordered texts, such as literary and narrative texts. Moreover, it presupposes that a text is an entity with a single creative point of origin (such as a single author or intention of composition) rather than a collection of material originating in many authorless contexts. Some medieval corpora of glosses are relatively stable and display a level of fixity and uniformity of authorial intent that has allowed editors to treat them like literary texts. The genealogical method was, thus, successfully deployed on the high medieval systematic *glossae*, from the pen of known scholars, that came into being in the cathedral schools and universities (Kostoff-Kaard 2015,

Morard et al. 2016), and on those early medieval corpora of glosses that have been described as glossing traditions or commentaries (Grifoni 2003, O'Sullivan 2004, O'Sullivan 2010, Cinato 2015). However, the more fluid, unstable, and authorless a particular corpus of glosses is, the more the genealogical method proved ill-suited for the task. The early medieval corpora of glosses, in particular, are significantly 'unruly', to the extent that the genealogical method usually produces no satisfactory results and sometimes cannot be deployed at all. This is also true for the glosses to the first book of the *Etymologies* of Isidore of Seville.

The fluidity of this corpus, as of many other early medieval corpora of glosses, is rooted in its organic nature. Today, approximately 4, 200 glosses to the first book of the *Etymologies* survive across 50 manuscript witnesses. They were generated between the seventh and the tenth century in the environment of early medieval monastic centers. Rather than reflecting a systematic engagement with a text by a single author or circle, they emerged from the immediate interests and needs of many isolated anonymous individuals and groups who had no literary ambition nor an intent to disseminate: teachers and students responding to the challenges of Latin education, scholarly-minded readers engaged in private study, and monastic scribes who wished to 'improve' particular manuscripts for the sake of their brethren (Steinová 2020). Due to processes beyond the control of the original glossators, these glosses were shared, traveling from one center to another (Baldzuhn 2011). Once diffused, they became often attached to other glosses, coalescing into large deposits in certain manuscripts (or rather the coalescing may have been a significant factor of preservation and therefore uncoalesced glosses rarely survive, Teeuwen 2008b). As a result, glosses to the *Etymologiae* reach us today in a bewildering range of amalgamated or fused forms rather than in the shape in which they were conceived. The layers of glosses preserved in surviving witnesses are each unique in their particular configuration, while at the same time, they contain identical glosses, pointing to sharing and circulation. Crucially, it is not necessarily fruitful for editors to try to disentangle different creative layers in any manuscript witness or the corpus to reconstruct the 'archetypal state' of particular batches of glosses, for this task may prove ephemeral (what to do, if it turns out that a corpus of several thousand glosses coalesced from hundreds of textual kernels, as seems to be the case with glosses to the first book of the *Etymologiae*?). Rather, an editor should attempt to represent the corpus in its entirety, acknowledging and fleshing out its heterogeneous character and complex history.

Some scholars responded to the problems of early medieval organic corpora of glosses by editing the layer of glosses as appearing in a single important witness (Wieland 1983, Teeuwen 2008a, Eisenhut 2010, Bauer, Hofman, Moran 2017). However, such an approach can be adopted only if there is one prominent and representative witness of the entire corpus, which is not the case with the glosses to the *Etymologies* nor with many organic corpora of glosses. Another solution facilitated by the use of digital technologies is to transcribe and treat each of the witnesses separately and to publish the resulting collection of material as a digital archive (Aris, Wiener, Hellmann, Posselt, and Ullrich 2010, Moran 2022). However, such a publication project would not address the questions of relationship between the witnesses and transmission of glosses nor clarify the existence of possible clusters of material. Indeed, both approaches may rather obscure these aspects.

Our solution to the problem of organic corpora of glosses is to turn to network analysis as an editorial framework. The genealogical method shares many similarities with network analysis and can be seen as its lighter substitute. In fact, the stemma, the model for representing the relationship between witnesses utilized by the genealogical method, may be considered a type of a network graph, namely a rooted directed acyclic graph (Flight 1992, Hoenen 2020). Our objective is not to construct a stemma, as the glosses have many independent sources there is no 'archetypal state'. Rather, we wish to represent and explore the relationship between the witnesses of our corpus to understand its structure and layering. We therefore sketch the contours of our corpus as a network graph that is not subjected to the constraints assumed by the genealogical method: we allow it to be unrooted, undirected, and cyclic. Crucially, this method is not sensitive to the position, order and mutual closeness of shared glosses as is the genealogical method and it can therefore easily handle the fluidity of layers of glosses. Since we do not assume that glosses were transmitted between manuscripts in the same manner as narrative texts, for example because they could be transmitted orally and via memory, rather than by copying from an exemplar, we do not wish to trace their descent, but rather explore their similarities, represent the relationships between them, and identify clusters that may enlighten us about how they circulated. While our network approach does not directly address the question of transmission, the establishment of relationships and similarities can provide insight into this matter. The network approach is, however, not a panacea to all woes of editing corpora of glosses and comes with its own limitations (see the discussion in section 6 below).

2.3 The method

In our digital edition, we treat manuscript witnesses as nodes (just like a stemma would). We distinguish between shared glosses (those that appear in two or more manuscripts, 1639 glosses from our corpus) and isolated glosses (those that survive in a single witness and therefore cannot be used for tracing relationships between them, 2574 glosses). When deciding whether a gloss is shared by multiple manuscripts, we ignore spelling variation, abbreviations, morphological form (e.g., whether the gloss accepts the case, number, etc. of the lemma or has a different form), word order, presence of certain Latin phrases characteristic for glosses that have no implication for the meaning of the gloss (e.g., *id est, hoc est, scilicet, sicut, quasi, vel*), textual corruptions that likely occurred as a result of mechanical error in a single witness, and omission of or variation in minor textual elements that do not alter the meaning of a gloss (e.g., prepositions and prefixes). In the case of long glosses, we also consider them shared even if they are imperfectly reproduced in several witnesses, as long as most of the complete gloss is present and therefore identifiable. Each shared gloss creates an edge between the manuscripts. Since many glosses could be shared by the same pair of manuscripts, rather than multiplying the edges, we plot just a single edge and assign it a weight based on the number of glosses and their individual ranks (see below). Thus, the more glosses are shared between two or more manuscripts, the heavier the edge that connects them. In this manner, we can examine and visualize the degree of similarity between the individual manuscript witnesses without presuming that it informs us about their genetic relationship.

We, furthermore, assign two properties to each shared gloss. First, since similar glosses may come into being independently multiple times (Conti 2020), we rank each shared gloss based on its (non-)triviality, indicating the likelihood that it is polygenic (because it is trivial and could be easily coined independently many times) or monogenic (because it contains non-trivial elements that make it likely that it was coined precisely once and therefore indicates relationship). A shared gloss with the highest value 4 (45 glosses) is so non-trivial that its occurrence in multiple witnesses can be only explained by a genetic relationship. This is, for example, the case with the gloss explaining the name of the metrical foot *Pyrrichius* (the Pyrrhic) in *Etym*. I 17.2, in our edition having the identifier L17.2.2, *quasi certator ludicus, vel a Pyrrho filio Achillis nominatus* ('named as if after a competitor in a contest, or from Pyrrhus, the son of Achilles'), which appears in two manuscripts. Here, the number of elements shared by the two manuscripts (9 words) and the originality of information (i.e., the imaginative connection between the Pyrrhic and Achilles's son Pyrrhus) make it clear this cannot be a polygenic gloss. A shared gloss with the lowest value 1 (77 glosses) is so trivial that it could be derived many times independently and it cannot be accepted as evidence for a relationship. This is, for example, the case with the gloss *scilicet verba* ('that is words') attached to the word *grammaticorum* ('of grammarians') in *Etym.* I 9.2 (L9.2.1) in two manuscripts. Since it clarifies an obvious ellipsis in the text (*Verborum genera duo sunt: grammaticorum atque rhetorum*; 'There are two types of verbs: [those] of grammarians and [those] of rhetoricians'), we may assume it was coined multiple times. Most shared glosses in the corpus were assigned the triviality rank of 2 (unlikely but possibly derived multiple times independently, 404 glosses) or 3 (highly unlikely derived multiple times and therefore most likely monogenic, 170 glosses). The ranking of glosses in our edition by their (non-)triviality ultimately resembles the assessment of variants in genealogical method - it, too, relies on a deep philological knowledge and the exercise of the *iudicium*, which may in theory lead to the same gloss being ranked differently by different editors.[2] The triviality rank is used in our model to add additional weight to edges in the network graph, as an edge is a compound of the ranks of constituent glosses. Moreover, the triviality rank can be used to eliminate shared glosses whose value for the establishment of relationship is limited (e.g., those with a rank of 2 or less).

Each shared gloss is also assigned to a cluster, which groups shared glosses exclusive to a certain pair or group of witnesses.[3] For example, we label glosses that are shared by manuscripts Orleans 296, Harley 3941, and Reims 426 as F1 and glosses shared by Orleans 296 and VLO 41 as I1. We identified 211 unique connections (edges) among our 50 witnesses. On average, a connection consists of only 3.3 glosses and has a weight of 7.5, while the average rank of a

[2] Among the criteria that we used to assign shared gloss a high non-triviality rank (3 or 4) are: a) number of shared elements appearing in the same sequence (in our case, at least four words in a gloss); b) the presence of the shared gloss in a significant number of manuscript witnesses (in our case, at least five); c) if glosses are citations from authoritative sources; d) the presence of identical highly idiosyncratic, unusual or erroneous information (e.g., identifying Mount Parnasus as an island, L39.13.5); d) the presence of identical textual errors or corruptions; f) if glosses form logically coherent sets within the text (e.g., explaining the Greek etymology of grammatical terms appearing in a specific text segment); and g) in case of lemmata that attracted many different isolated glosses, if we encountered shared glosses (e.g., L22.1.11, which was glossed by five different glosses, four of them isolated, and one shared).

[3] An overview with definitions is given in the edition at https://db.innovatingknowledge.nl/edition/#left-clusters.

shared gloss in the corpus is 2.1. We selected 14 ‘clusters of high importance’ (A-B, D-I, M-Q, S) having weight larger than 20, 4 ‘clusters of medium importance’ (T-W, Z) with a weight larger than 10 but lower than 20, and one cluster of special interest consisting of only one or two glosses, but with the ranks of 3 and 4 (C). Because of their significant above-average weight or non-triviality, these clusters reflect the most important links between manuscripts and therefore are most likely to represent collections of glosses circulating in the early Middle Ages. We assigned the remaining shared glosses to one of the two generic clusters (X1 and X2), which group those shared glosses that cannot be considered to indicate a relationship between manuscripts because their weights or their ranks are too low.[4] For example, the clustering method we just described reveals that 148 of the 211 connections we identified are constituted by one or two glosses of lower ranks (1-2).

3 Digital editions, Distant Reading and the Explorative Edition.

Scholarly editors were among the first humanities scholars to use the computer in scholarly work (Dalbello 2011). Robinson wrote that a print scholarly edition was essentially saying 'I am a hypertext: invent a dynamic device to show me' (Robinson 2005). Editors with foresight understood that using the computer to model their text set them free of many of the limitations of print publishing (Unsworth 2002), long before the advent of XML made this a practical thing to do for most editions. According to Sahle, a scholarly edition is even fundamentally impossible in printed form: 'Die Grundstruktur einer Edition mit ihren vielstufigen Repräsentationsformen und deren Vernetzung mit erschliessenden und kontextualisierenden Informationen wiederspricht grundsätzlich der linearen Struktur typographischer Publikationsformen' (Sahle 2003:79).

This dynamic character of the edition has always been seen as an essential characteristic. This is a requirement that goes beyond different views of the same text, say a diplomatic text and a reading text, or a side-by side view of different witnesses, as in the Versioning Machine.[5] Digital editions should not just be publications, they should become interactive tools for the study of the edited texts. Robinson, in a far-sighted article (2004) wrote that current electronic editions '[failed] to use new computer methodologies to explore the texts which they present'. He continues 'New systems of data analysis might offer ways into all this material, and so permit us to see patterns and relationships always there, but never before accessible. In turn, we could use the explicatory power of the computer to allow readers to discover these, just as we do for ourselves'. Editions that provide these facilities he describes as 'lean-forward' editions. These calls are echoed in Fraistat and Flanders' introduction to the *Cambridge Companion to Textual Scholarship* (2016), where they require digital editions to be dynamic and to encourage user interaction as well as to be 'scalable, allowing for microscopic and macroscopic inquiry'.

In practice, the dynamic possibilities of the digital edition have been realized only to a limited extent. Many editions allow the user to select what she will see (text, facsimile, notes) and to

[4] We distinguish X1 and X2 because some lemmata attracted more than one set of generic shared glosses and we wanted to distinguish these generic sets one from another for the purpose of their encoding. Overall, we assigned 639 shared glosses to these generic clusters: 615 to X1 and 24 to X2.
[5] http://v-machine.org/.

determine to some extent the layout of the screen (text only, text and image, etc.). Often, the user can select a critical or a diplomatic version of the text. Editions are often heavily hyperlinked: from (critical) text to annotation, from text to apparatus, from apparatus to witness, from named entity to an index of occurrences of that entity, etc. Of course the text can be searched, and the sometimes unwieldy 'advanced search' windows are being replaced or enhanced by powerful faceted searches. There is no doubt that modern digital editions are interactive and dynamic tools.

However, what is still lacking are tools that summarize or create an overview of the information that the editors so laboriously brought together. There are impressive digital editions that bring together many versions of a text but leave the user with the question: 'And now what? In which chapters do the interesting changes occur? Do I have to work through the entire edition to get a sense of the textual evolution of this work?' It is illustrative that the world of the digital edition (text encoding) and the world of developers of analytic tools have mostly been far apart. A panel abstract from DH 2012 talks of 'the way text encoding specialists and text analysis scholars do their work [...]: they hardly ever work together on the same projects. Because of this lack of connection, some of the long-promised benefits of markup for analysis remain unrealized' (Bauman et al. 2012).

This scarcity of tools that help create overview in the digital edition is strange, given the popularity of 'Distant Reading' within the Digital Humanities. Distant reading, a term coined by Franco Moretti (Moretti 2013) is now used as an umbrella term for those types of computational analysis of humanities objects that help explore and understand patterns and regularities within these collections. If distant reading is important in what we do, why don't we have editions that support the practice?

The integration of analytic tools into digital editions seems to be a feature of large (or largish) text collections rather than of scholarly editions. Larger collections have integrated topic modelling tools and network diagrams (Ravenek, Van der Heuvel and Gerritsen 2017), tools for intertextual analysis (ARTFL Project 2021), concordance tools (Mahlberg et al. 2016) and tools for text clustering (Unsworth and Muller 2009).

In editions, the use of maps and interactive timelines is no longer an exception[6] though in general these are visualizations only, not tools that are integral to handling the site and not linked back into the material that the site has to offer. An example where the timeline (not the map) does link back into the edition is given by The (De)collected War of the Worlds, an edition of Wells' *War of the Worlds*.[7]

Deeper integration of more advanced tools is still very rare. If we look specifically at networks, we can bring up several examples. In their edition of a collection of April fool letters to Mark Twain, Myrick and Ohge (2017) include a network visualisation that helps explore the social network that the letters originated from. The Letters 1916-1923 project[8] features an interactive network

[6] E.g. APW Digital, https://apw.digitale-sammlungen.de/
[7] https://decollected.net/.
[8] http://letters1916.maynoothuniversity.ie/

visualisation of correspondents (as well as an interactive map). Its purpose was described by Hadden (2016) as 'facilitat[ing] exploration at multiple levels'. In both these projects the network map is informative and beautiful, it increases the usefulness of the site. However, it is not linked back into the edition, it is shown as if it were an extra, a separate component not really part of the main interface. The only example we have found where a network diagram links directly back into the edited collection, and thus can be used for navigating the edition, is in the collection of Saint Louis Circuit Court Records.[9] This collection includes a network diagram of court cases and the persons involved in them. Clicking a case node brings up that case's record in the edition. Here, the diagram is still very much a *Fremdkörper* within the collection; it is not embedded within the site but on a page of its own that has no headers, background or other information that tells us where we are. It is, however, a clear step in the direction of what we argue for in this contribution: editions that go considerably beyond the inclusion of text analysis tools as required by Siemens (2005), editions that contain interactive visualisations integral to the navigation of the site, that can help a user explore and work through the edited material and provide an experience which truly leaves behind the print paradigm. These editions, which we propose to call explorative, allow a distant reading of the textual material that they contain (Fraistat and Flanders' 'macroscopic inquiry') integrated with the facility for zooming in on textual detail ('microscopic inquiry'). In the next section we describe what our attempt at such an explorative edition for the glosses to Isidore's *Etymologiae* looks like.

4 A proof-of-concept edition

In section 2, we showed that a network representation is the most adequate for representing the glossing tradition that we are interested in, and in section 3 that there is still a dearth of digital editions that use advanced tools to present and make accessible textual material. In response to these requirements, we designed our edition not only to allow but to encourage users to actively engage with the material through the constitution of network visualizations. We also want our edition to show not just the glosses by manuscript but also the intermediate clustering. In this section we discuss first the XML structure that we used (4.1), the traditional components of the edition (4.2), a number of general characteristics of the network graphs that we include (4.3) and the individual graphs in greater detail (4.4 through 4.6).

The edition that we present is a proof-of-concept edition. It should be usable, but its primary purpose is to illustrate the argument we are making here. It is not meant to be the only way that users can access the glosses. Apart from the edition data being available on GitHub, the information about the glosses will also be integrated in the Gloss Corpus project (Moran 2022). A very obvious limitation of our edition is that it does not include facsimiles, which could not be incorporated because of practical reasons.[10]

[9] http://digital.wustl.edu/c/ccr/.

[10] The main obstacles proved to be the limited time and rights issues (the 50 manuscripts used for the edition are held in different manuscript collections following distinct policies of image-reuse, many of them not open to sharing their facsimiles in a digital edition).

4.1 Data structure

For the edition we built an XML file following the *Guidelines* of the Text Encoding Initiative (TEI Consortium 2021), with a few local extensions. XML, though itself a text-based format, lends itself very well to visualization (Jung et al. 2003). More specifically, many aspects of the XML structure can be translated into network terminology and lend themselves to visualisation as networks (Beshero-Bondar 2020, Bingenheimer 2011). We discuss here first the encoding that we adopted, then the implied network structure. The edition XML, with the software used for creating the edition, is available in a github repository.

Since the glosses are dependent on a text that they comment on, we began with encoding the text of the first book of the *Etymologiae* in the <text> <body>. As our basis for this text, we use the 1911 critical edition of the *Etymologiae* by W.M. Lindsay. Following the structure of Lindsay's edition, we divide the text of the first book into 44 chapters using , and each chapter further into sections using <ab>. The key element of the mark-up of the main text concerns the identification of lemmata to which glosses were attached. As the <lem> (lemma) element in TEI is reserved for textual variation, we encode them using <seg> (segment) elements. We chose to encode several features of the glossed text: the presence of Greek words, verse quotations, citations, and notable variant readings printed in Lindsay's edition in square brackets[11] Summary information about the individual manuscript witnesses (identification, information about glossing hands, place and time of origin, provenance, and manuscript type) is kept in <msDesc> elements.

The glosses are encoded separately from the text in the recently introduced <standOff> element using the <gloss> element. We introduce two container elements: a <glossGrp> grouping <gloss>es belonging to the same lemma and a <listGlossGrp> to group the <glossGrp>s. The <glossGrp> points to the lemma using the @target attribute. The <gloss> points to the manuscript in which it appears using the @corresp attribute and may further point to one of the glossing hands within the manuscript using the @hand attribute. Optional attributes of a gloss include its language (e.g., Old High German, Old Irish, etc.; Latin is not encoded), source (if it is a citation of a known text), and a specifying type in case of glosses to glosses. Sections of glosses that were copied in shorthand or dry-point (scratched in with a pen) are encoded via <seg>s with an appropriate @type and @hand attributes.

Similarity among the glosses is encoded via a <glossCluster> (a specialization of the TEI linkGrp element) that contains pointers (<ptr>) which point to the similar glosses using the @target attribute. Thus, all <gloss> elements pointed to in a single <glossCluster> are considered manifestations of the same shared gloss. Importantly, not all manifestations of a shared gloss are necessarily attached to the same lemma, since, as was noted, shared glosses can appear in different positions.[12] The <glossCluster> is placed within the <glossGrp> that holds most of the

[11] Finally, <pb> elements provide the link to <surface>s in the various manuscripts included in the <facsimile> using the @facs attribute.In the current edition, these links are unused.

[12] To clarify, in our digital edition, we decided to account only for shared glosses occuring in mutual vicinity (e.g., within the same chapter section or chapter). However, we also noticed identical glosses

manifestations of the shared gloss. The two properties of the shared glosses we mentioned earlier, the cluster to which they belong and their (non-)triviality ranking, are encoded as @ana and @weight attributes of the <glossCluster> respectively. The 19 specific clusters and 2 generic clusters are defined as <interp> (interpretation) elements, to which the @ana attribute points.

In our digital scholarly edition, this XML encoding allows for the definition of nodes and edges of a network in three ways: each hierarchical relation implicitly defines an edge between two nodes, each explicit pointer creates an edge, and edges can also be created by co-occurrence of attribute values (our edition does not use the last option). From these initial edges, further edges can be derived by concatenation. For instance, edges between manuscripts (see section 4.6) can be defined by going from manuscript to gloss, from there to similarity group (glossCluster), then to cluster, and three more steps going to the other manuscript in the edge.

4.2 The edition interface

We chose the digital medium and the particular configuration for the edition to fulfill the editorial criterion for corpora of glosses we articulate above: that the users can traverse all levels of the edited corpus while also being able to make various selections from the material based on their particular choices (e.g., regional groupings of manuscripts, manuscript type, presence of notable features, etc.).

The edition uses two panels. The left panel is reserved for the text of the *Etymologiae* and for the introductory texts by the editor, which appear in this panel when the edition is opened. The right panel accommodates three different displays: gloss display by manuscript or chapter as well as network visualizations. Edition users can switch between these displays by clicking on one of the tabs at the top of this panel. When the edition is opened, one of our network graphs (see section 4.4) is displayed in the right panel. The user can use it to analyse the contexts of the edition, make a selection of glosses from the larger corpus (e.g., by chapter or manuscript), or as an entry point into the textual part of the edition.

We also prepared a more traditional roadmap for the edition by including a table of contents for the 44 chapters at the leftmost side of the left panel. Clicking on a chapter number in this table of contents opens the text of that chapter in the left panel and a corresponding gloss display in the right panel. The lemmata in the main text that received glosses are highlighted in red, marked with a superscript g (for gloss) and hyperlinked to glosses in the right panel. Glosses in the right panel are organized by lemmata and identified by their unique identifiers. Both shared and isolated glosses are equipped with a manuscript tag that indicates in which manuscript witnesses they appear. These tags are hyperlinked to separate displays of the glosses by manuscript (opening in the right panel) as well as to more detailed descriptions of manuscript witnesses in the project database associated with the edition (Steinová 2021). The gloss display contains two

appearing far apart, which could be considered manifestations of a single shared gloss. These far-removed shared glosses just demonstrate the extent of fluidity and discontinuous nature of corpora of glosses.

additional sets of information for shared glosses: their weight (as w=...) and a cluster label (A-Z). The cluster label is also displayed in the left panel of the edition as a superscript of a lemma, so that the edition users can quickly scan how shared glosses are distributed per chapter. Users can hide all isolated glosses by ticking off a box at the top of the right panel. The glosses they are going to see are also going to be affected by the choice of filter (see below).

The various textual features of both the text of the *Etymologiae* and its glosses that are encoded in our XML are displayed in the left panel and gloss display respectively. For example, the text in the left panel includes and distinguishes passages found only in some manuscripts as far as they received glosses and therefore are relevant to our edition. The gloss display includes a standard set of editorial notes used in editing medieval Latin text (e.g., indication of omissions, additions, erasures, gaps, reconstructed and supplied readings, as well as of notable textual errors). Glosses copied in a special manner (i.e., shorthand and dry-point) are distinguished by the use of a different font. Additional information about glosses (such as notes on interesting features and language) are displayed in the rightmost column of the gloss display next to them .

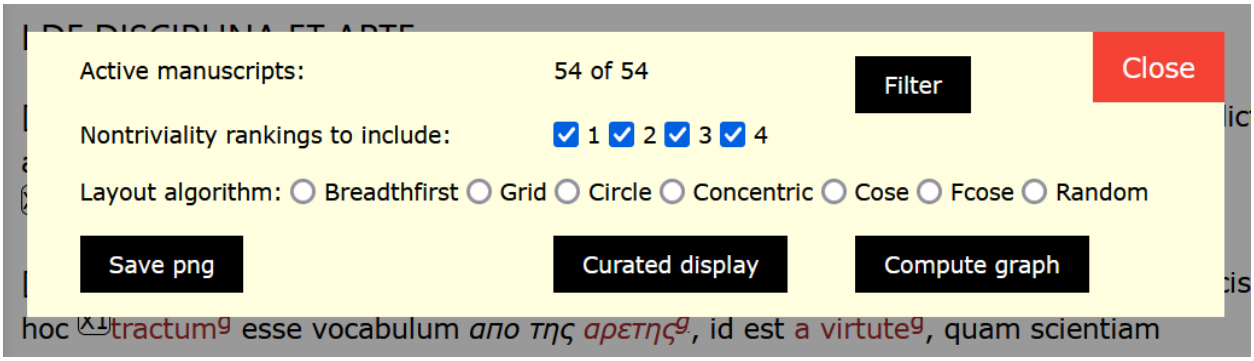
An editorial introduction with various displays of supporting information is also available in the left panel. It includes for instance an overview of sources and detailed description of the most important clusters. In the left panel, users can also open an overview of the available network visualizations. Here, they can select a network visualization, which will be opened in the 'Network' tab in the right panel.

4.3 Network visualisations

In the remainder of this section, we describe three network visualisations included in our proof-of-concept digital edition: (1) a bimodal network of manuscripts and chapters that shows which manuscripts contain glosses to which chapters of the first book of the *Etymologiae*; (2) a bimodal network of manuscripts and clusters, visualizing which manuscripts contain glosses from which clusters; and (3) a geolocated unimodal network of similarities between manuscripts (in which edges correspond to clusters displayed in network (2) as nodes). The three visualizations represent selected examples of networks that can be derived from our XML file. We chose them because they are most relevant to the context of this edition. Other examples of networks that could be derived from our XML file include: a network of groups of glosses copied by specific hands or a network of glosses sharing similar traits, such as language or manner of writing (e.g., dry-point or shorthand).

We designed our visualisations to be working instruments. That is, they are not just a set of illustrations embellishing the edition or communicating information conveyed elsewhere in text by graphic means. The complex set of relationships encapsulated by the edition is richer than what a text or a static image could convey and we feel that without these visualizations, the edition would be lacking an important aspect. Accordingly, we expect that researchers will want to manipulate the network in order to find their way through it and engage with the underlying information. The network visualizations are therefore designed to be interactive in a number of

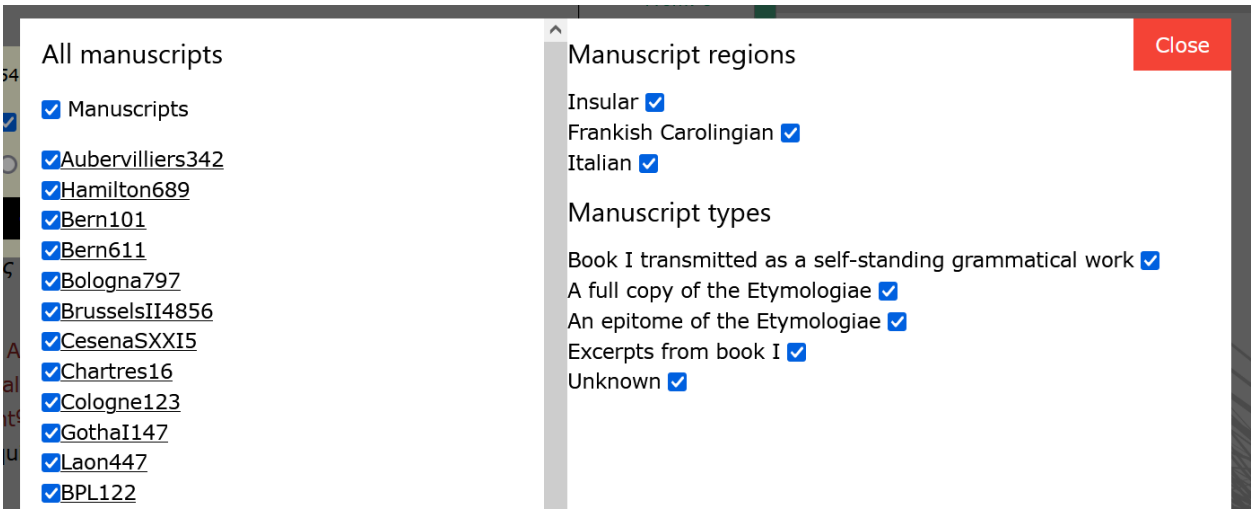
ways. A control panel is available to allow the user to control the content and layout of the graphs (figure 1). Depending on the network visualization, the control panel will display a filter option for the nodes, a filter option by gloss rank and options to experiment with different layout algorithms (see section 4.3).



[Figure 1. The control panel determining the layout of the networks.]

Forms of interactivity include, first, that the user should be able to select the nodes to display in a visualization. From the control panel, another panel can be opened that allows the users to display only certain manuscripts or groups of manuscripts (e.g., those that originated in northern France) (figure 2). We also include a slider in the network display that allows for the filtering by edge weight or node weight (i.e. the summed weight of the adjacent edges). Second, the user should be able to experiment with network layout algorithms. The proof-of-concept edition is not intended to include a full-fledged network analysis environment, but depending on the interests and preferences of the researcher, a different layout algorithm might be appropriate and the edition should provide some flexibility in this respect. We therefore include several layout algorithms to choose from on the network control panel. Third, once the layout has been computed, it should be possible for the user to further manipulate the result to better understand the relations displayed in the graphs. Our prototype therefore provides the option to drag individual nodes with a mouse and to remove them by right-clicking and selecting this option from a context menu. Fourth, the user should be able to save the result of her filtering, manipulation, and application of layout algorithms as a png picture for her lab notes or for a publication.[13] Finally, and perhaps most importantly, the users can click on various elements of the network visualizations in order to navigate to specific parts of the editon, such as a chapter, or glosses present in a manuscript. Precomputed network representations of the XML are available from the project's github repository.

[13] For the future: export as a GraphML file for further analysis in a more powerful network visualisation tool such as Gephi.



[Figure 2. Part of the panel that facilitates filtering the manuscripts to be included in the visualisation.]

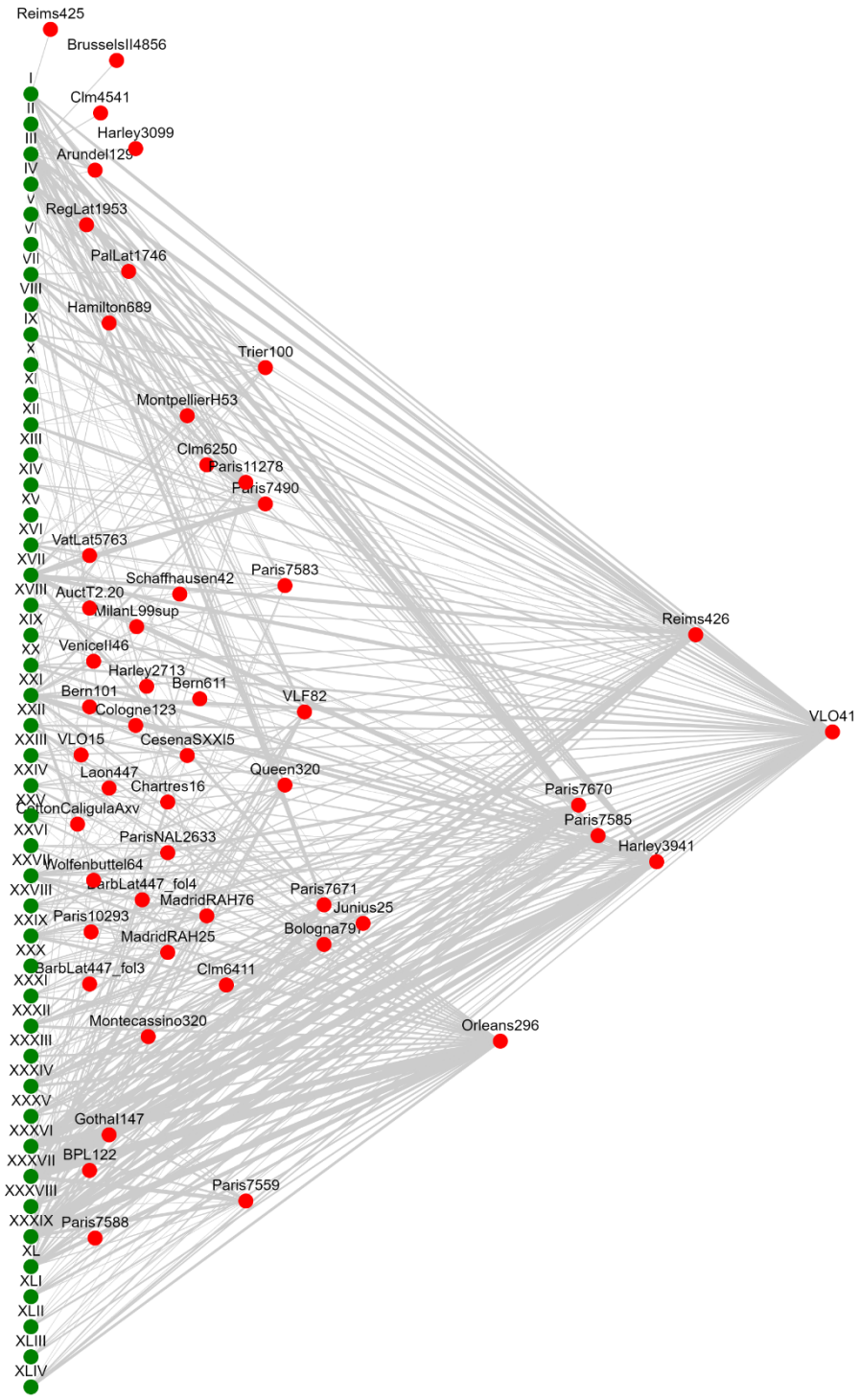
While we believe this interactivity is essential for a cognitively fruitful engagement with the network visualisations, we should also be aware that presumably any edition's editor, having deeply studied the edition's domain, is well qualified to provide an initial visualisation. We call this initial display the curated view: in our case a display that shows the most relevant manuscripts and possibly clusters, filters out the less relevant edges, and uses a layout algorithm that we consider the most enlightening for this network.

Technically, the edition's network visualizations are produced by the Cytoscape.js graph library (Franz et al. 2016).[14] We use a Cytoscape extension for the context menus (Dogrosuz et al. 2018). The popups are realized using Popper[15] and Tippy.[16]

4.4 Manuscripts and chapters

For a researcher new to this material an initial exploratory question might be: are the glosses in the various manuscripts more or less homogeneously distributed over the chapters of book I? Or do some manuscripts mostly gloss specific chapters? Our first network visualisation (Figure 3) is designed to help study this question. We display a bimodal network containing manuscripts and chapters. The edges between them have a weight determined by the number and rank of the glosses in that manuscript to a specific chapter.

[14] https://js.cytoscape.org/
[15] https://popper.js.org/
[16] https://atomiks.github.io/tippyjs/.



[Figure 3. Curated display of the chapter vs. manuscript network]

Our curated visualisation in this case displays the chapters on a vertical bar, while the manuscripts' horizontal distance from the bar is determined by their degree (i.e., the number of chapters that they gloss). Their vertical position is determined by the *fcose* algorithm (Dogrusuz et al. 2009), attempting to place them close to the chapters that they gloss most (and not too close

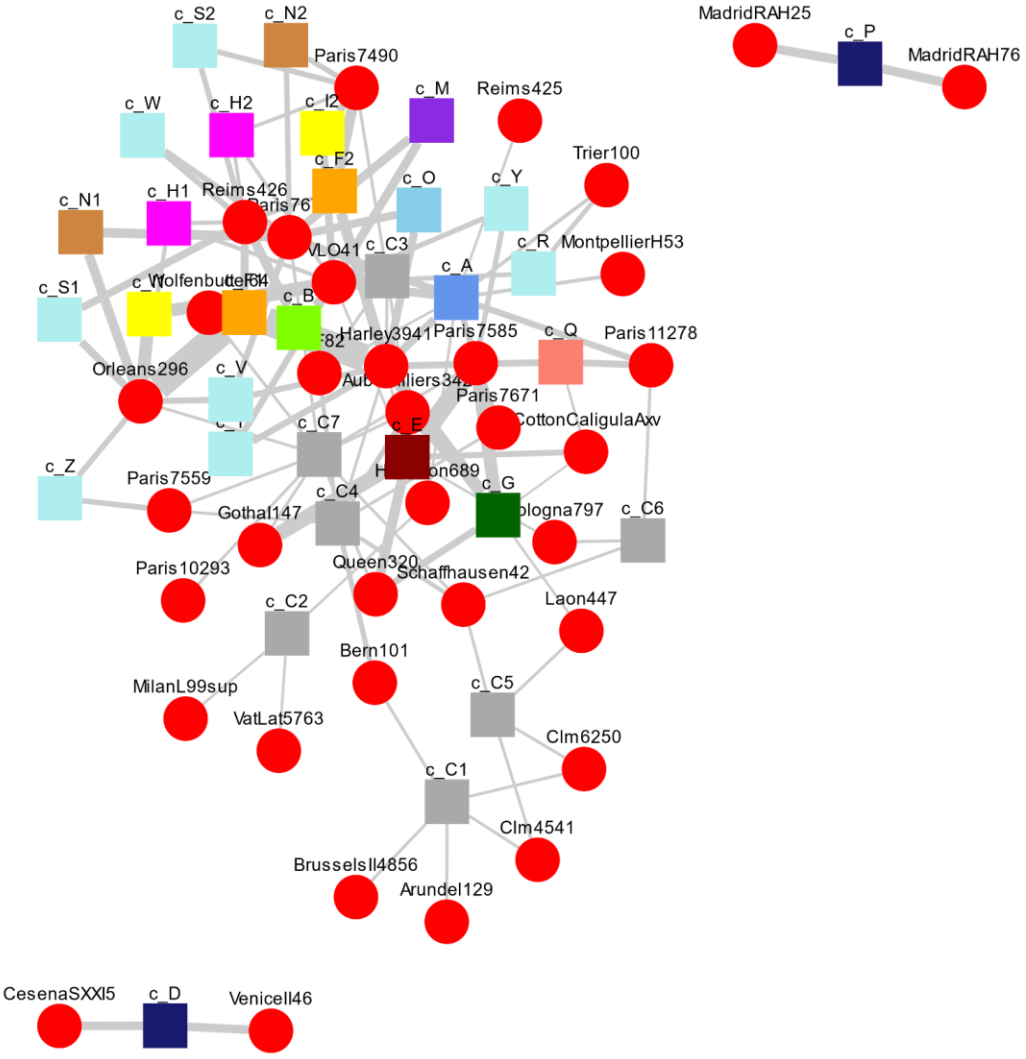
to each other). We chose this display because it makes immediately clear which manuscripts are the most important in terms of the number of chapters that they gloss and which chapters of book I of the *Etymologiae* received most attention. The width of the edges shows the importance in terms of number of glosses.[17] Clicking a node hides everything but the edges and nodes that this node is connected to and gives a better view of the importance of an individual manuscript or chapter. Applying a filtering by weight allows us to remove manuscripts that contain few glosses, and thus to focus on the more important ones.

What we can learn from the visualisation is, for example, that no manuscript glosses more chapters than VLO41 (41 out of 44 chapters, no manuscript is further away from the central axis), even though other manuscripts, Orleans 296 and Harley 3941, contain more glosses with higher ranks. More manuscripts with a higher degree appear in the lower part of the graph, indicating that medieval glossators were more interested in the second half of the text than the first. Indeed, many important annotated manuscripts that transmit the complete first book contain no or very few glosses to chapters 6-14 (e.g., Harley 3941, Paris 7585, Paris 7670, and Paris 7671). By contrast, Paris 7490 contains glosses specifically to these chapters, being an anomaly among the annotated manuscripts of the *Etymologiae*.

4.5 Manuscripts and clusters

Our second visualization (Figure 4) shows the relation between manuscripts (red discs) and clusters (squares in other colours). The width of the edges between the two types of nodes corresponds to the combined weight of glosses in each manuscript that can be assigned to the cluster. The layout is computed using the *cose* layout algorithm, which translates the relative similarity/dissimilarity of clusters and manuscripts into their mutual distance. Thus, this visualization allows the users to easily explore the mutual proximity and distance of manuscripts based on the clusters of glosses they contain. In the curated display, we leave out the two generic clusters X1 and X2, since they contain connections of low importance that probably do not indicate meaningful relations between the manuscripts (indeed, including the generic clusters would create edges between most manuscripts and clutter our visualization without adding meaningful information).

[17] To be precise, the edge's width is proportional to the square root of the weight. A linear relationship would make some edges invisible and others too wide. We use the same transformation in the other graphs.



[Figure 4. Curated display of the cluster vs. manuscript network]

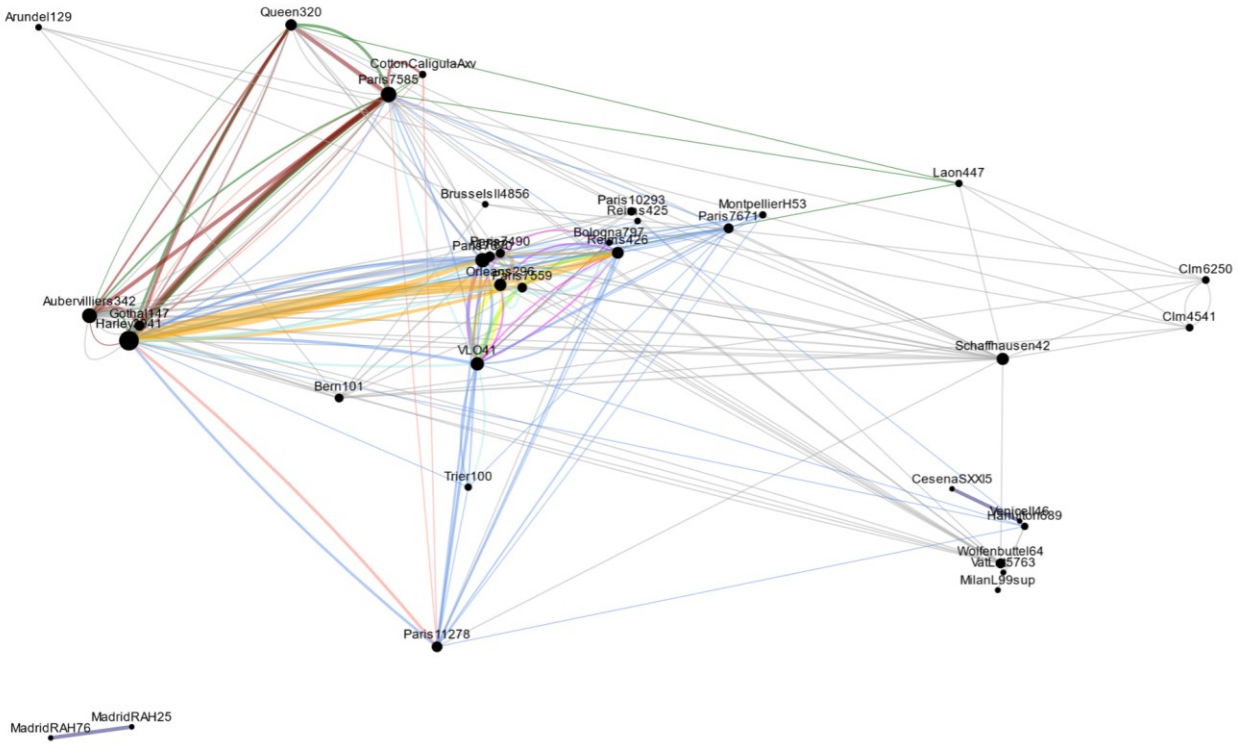
We opted for most cluster nodes to have different colours to distinguish individual clusters. Nevertheless, we also used the colouring as a means to show similarities between clusters. In the first place, two of our witnesses, Orleans 296 (containing only chapters 21-44) and Paris 7490 (containing only chapters 5-17), seem to contain a single layer of glosses because they were annotated in the same location and perhaps by the same group of glossators (they may even represent two fragments of a single damaged manuscript). We therefore assigned the same colour to the clusters that connect these two manuscripts to the same manuscripts (e.g., I1 connecting Orleans 296 with VLO 41 and I2 connecting Paris 7490 with VLO 41 are both yellow). Second, we coloured all ‘clusters of medium importance’ (T-W and Z) light blue. Finally, all micro-clusters constituted by one or two glosses of ranks 3 and 4 (C) are coloured gray. Users who want to delve into the details can access detailed information about the glosses

using the context menu on the manuscripts and glosses. As elsewhere, the detailed manuscript description in the project database is also accessible through the context menu.

The resulting network is split into three components. Filtering reveals that the largest component is itself divided into two interconnected segments: one smaller centered around clusters E and G (which corresponds to manuscripts with a connection to England) and a larger one around clusters A, F, O, and I (which happens to correspond to manuscripts with connection to northern France). The two segments are bridged by Harley 3941, a manuscript from Brittany that contains glosses from both clusters F and G. Filtering also makes clear which manuscripts are well-connected (Orleans 296, Harley 3941, VLO 41, etc.) and which are outliers with few connections to other manuscripts (Paris 7559 and Paris 7671).

4.6 Manuscripts by geographical location

The third visualization is a unimodal network that shows only the manuscripts and the relations between them as determined by clusters they share (Figure 5). We therefore assign edges a colour, which is identical to the colour of a cluster that joins them in the previous network visualization (see section 4.5). It is a multigraph, i.e., there can be multiple edges between a single pair of nodes. The node size is proportional to its degree, edge width is again determined by weight. The layout is determined on the basis of the geographical location where we assume the glossing took place. We try to avoid overlap between the nodes and labels of manuscripts from the same or similar locations by introducing some random variation.



[Figure 5. Curated display of the geography-based manuscript network]

The resulting graph groups together manuscripts from several regions where glossing took place: England, Brittany, Northern France, Italy, and Spain. Thus, it allows the users to explore regional and trans-regional connections between manuscripts and therefore to examine how glosses travelled between monastic centers in the early Middle Ages. For example, it shows that two important annotated manuscripts from Brittany, Harley 3941 and Gotha I 147, share no glosses, even though they were annotated in the same region (and perhaps in close geographical proximity), but are rather closely related to manuscripts from northern France and manuscripts from England respectively. Users can also learn that manuscripts from England are related to each other, and to manuscripts from Brittany, but less so with manuscripts from other regions, while annotated manuscripts from Italy and Spain are more or less isolated. The hub of glossing as well as of gloss exchange, as this graph indicates, was in Northern France. Those who are familiar with the landscape of early medieval Europe may also note that German-speaking regions are absent from this graph: the first book of the *Etymologiae*, it seems, was not glossed in this area.

5 Potential applications elsewhere

5.1 In the domain of glosses

The editorial model provided by our prototype edition of the glosses to the first book of the *Etymologiae* can be applied to other organic corpora of glosses. The minimal condition is that the given corpus contains glosses shared by multiple witnesses that can be assumed as edges of a network. It has been, indeed, noted that other corpora of glosses contain 'parallel glosses' that could be used to explore them as networks (Lambert, 1983, Wieland 1985, Bauer 2019a, Bauer 2019b, Bauer 2019c). Even the corpora of glosses that have been partially probed with the genealogical approach (Keskiaho 2019, Aubé-Pronce and Pollard 2021) could benefit from a complementary network display that could transcend the limits of the genealogical method. The three network displays we propose in this article - one to relate glosses to the structure of the glossed text, one to relate manuscripts to certain batches of glosses (clusters, families, sets, etc.), and one that shows the mutual relationships between manuscripts based on their geographical provenance - should be well-transferable to other corpora of glosses.

The network approach can be further expanded across multiple corpora of glosses, as long as these share certain similarities that allow them to be considered shared glosses even if appearing in texts of different authors. A good example are the glosses to grammatical authors in the early Middle Ages. Grammar was a highly regularized subject taught in schools. Since many of the early medieval grammars follow a single model (the *artes* of Donatus), the text of early medieval grammatical manuscripts tends to be similar or identical (to the extent that if preserved fragmentarily, researchers may have a problem discerning them). As a result, many early medieval annotated grammatical manuscripts likely contain similar or identical glosses. Glosses to early medieval grammatical texts were never examined as a single megacorpus,

chiefly due to the scholarly paradigm inherited from the genealogical method: to treat each text separately as an enclosed tradition. The network approach could help us open up this paradigm, because, as was argued above, it is position and context agnostic, and it can, therefore, be used to seek similarity and relationships even across manuscripts that transmit different texts.[18] A network visualization could be even used to display gloss-based relationships between annotated texts as virtual entities rather than between annotated manuscripts as their witnesses.

There are many medieval technical disciplines that resemble grammar in being highly standardized (e.g., medicine, computus, music, law, exegesis). Editions of corpora of glosses to these texts would benefit from a network approach, as it could dissolve the textual boundaries that are too readily assumed and allow us to examine medieval glosses perhaps closer to the medieval context of their origin and use. Above all, this approach could help us to elucidate the question of transmission of glosses and the character of medieval education: was it possible to transfer glosses from one text to another, and if so, how often did such transfer occur? When and where did it occur most frequently, and what does it reveal about medieval educational strategies and techniques?

5.2 In other fields

The network editorial framework is likely to be suitable for editing other historical documents and textual traditions that may resist the application of the genealogical method, especially as they have the form of chains of small textual entities and are characterized by a low degree of cohesion, integrity and sequentiality. These will be found mostly on the side of the pragmatic and technical literature, among various organic and systematic collections (e.g., of letters, sermons, miracles, lecture notes, excerpts, medical recipes, prayers, liturgical formulas, and technical instructions).

A good example from the realm of the early medieval Latin and vernacular literature are computistic handbooks, which are well-known for consisting of many small textual units (algorithms) that can be assembled together in many different sequences. It has already been suggested that in order to study these handbooks, it is essential to focus on these algorithms and understand how they are shared by manuscript witnesses and circulated (Warntjes 2020). A different example is provided by medieval sermon collections which have been likewise recently approached through network analysis (Boodts 2019). Finally, the network approach has

[18] This being said, we acknowledge that building such an environment would entail additional technical challenges and the model we propose in this article is not alone suitable for such a task. For example, an edition of a megacorpus of glosses to early medieval Donatus-based grammars could not work solely with unique identifiers tying glosses to text sections as in our edition of the glosses to the first book of the *Etymologiae*, but would have to apply a different system of identificators (e.g., based on their alphabetical order). Similarly, the contextualized display of shared glosses may prove challenging as it is bound to entail several, even many, different grammatical texts side by side.

been recently used for the analysis of relationships, although not for the editing, of the books of hours (Boillet, Bonhomme, Stutzmann, Kermorvant 2019).

If we move beyond networks as an editorial framework to networks as a tool that could enrich digital editions more broadly, the obvious application would concern social networks. Many historical documents presuppose a strong social context in which they emerged or circulated and therefore would naturally benefit from network visualizations that would make these contexts obvious. Collections of letters are a good example of digital publications where network graphs can reveal properties of the underlying correspondence network. Graphs can be used to visualize the amount of correspondence exchanged between two persons (or cities) as we already saw in the example of Letters 1916-1923. Much work has been done on network visualisations in correspondence projects that could easily be integrated into editions (e.g. Floyd 2019, Beshero-Bondar and Donovan-Condron 2017). Moving more into interpretative territory, visualization of character networks in novels and plays could be an obvious enhancement in many digital editions (Schöch et al 2019, Fischer and Skorinkin 2021, Baird 2021). And interpretation need not concern itself only with social relationships: story summary graphs (Shahbazi 2019), topic networks (Hall 2021) and text networks (Paranyushkin 2019) are other examples of network diagrams that could enhance digital editions. Depending on the genre of the text, its research interest and the level of encoding, for any text multiple relevant network visualizations could be generated.

6 Challenges and conclusion

While we note many advantages of using a network approach in editing and the construction of digital editions in this article, we are also aware of the challenges and limits of this method and want to mention some of them.

From the perspective of editing the glosses, the network framework cannot compete with the genealogical method in sophistication and interpretative faculties. In particular, the method we describe in this article cannot be used to establish patterns of transmission and identify lost or hypothetical intermediaries. The users of our proof-of-concept edition may notice many intriguing relationships within the edited corpus, but the network visualizations will not help them to interpret them. For this, traditional analysis is still needed. Moreover, even with a (non-)triviality ranking, there is no reliable method to winnow out glosses that look similar from those that were genuinely shared, because the brevity and simplicity of glosses increases the probability that they were coined multiple times. Thus, even though the most important clusters should be considered genuine relics of transmission and therefore of links between witnesses (and certain monastic centers), a degree of ‘noise’ can never be filtered out from our model. Finally, as yet none of the network visualizations employed in our edition has a chronological dimension as a stemma would. The edition unwittingly obliterates the fact that the examined layers are not coeval (the difference between them is in some cases several centuries). Some witnesses clearly contain several batches of glosses added by different hands. This chronological insensitivity of our edition could be remedied by discerning multiple layers of

glosses per witness in accordance with the glossing hands rather than treating each witness as a single layer (we already laid the basis for such treatment in our XML file by encoding hands). We could also visualise these temporal layers in the edition.

From the perspective of the explorative character of the edition, we feel that while we certainly set steps in the direction of integrating explorative research tools in its interface, more could be done. Even though many aspects of the edition's functioning are dynamic and would be impossible on a printed page, many other aspects are still static. In the context of this proof of concept edition we limited ourselves to a two-panel display. The user can open a list of glosses by manuscript from the visualizations, but this hides the visualization. A more flexible layout with multiple panels that the user can arrange according to her interests would certainly be desirable. While to some these features may seem merely cosmetic, we believe intuitiveness and responsiveness of the interface are essential in its use as a cognitive tool. And there are many other ways in which the edition could support exploration. For instance, the table of contents of the chapters could provide (visual) information about the length of the chapter and the number of glosses. Instead of a side-by-side-view of main text and glosses we could have a view where the glosses to a lemma 'fly in' as the user hovers over a lemma. Filtering in or out manuscripts could be accomplished by dragging and dropping thumbnails rather than by clicking checkboxes. We now focused on networks, but many relevant aspects could be handled using maps, bar charts, or timelines - there is certainly room left for more experimentation in this respect.

In spite of these limitations, we believe this contribution shows that the network paradigm is suitable for editing glosses. An edition that uses the network paradigm can show meaningful relations between the manuscripts without pressing these into a stemmatological model. It can reveal the presence of clusters of glosses shared between manuscripts that remained invisible in conventional editions of glosses. When the manuscripts' relations are expressed in interactive network diagrams, these diagrams can become essential tools for interacting with the underlying material. Thus, editions of glosses can become an example for the explorative editions of the future, digital editions that are more dynamic, intuitive, tactile and flowing than those we know today.

Literature

Aris, M.-A., Wiener, C., Hellmann, M., Posselt, B., and Ullrich, S. (2010). Projekt “EDITING GLOSSES”. Glossae in Persium. Ludwig-Maximiliansuniversität München. www.persius.mueze.lmu.de

ARTFL Project. 2021. 'The Intertextual Hub: Search and Navigation Across Digital Collections ', University of Chicago. https://intertextual-hub.org/.

Aubé-Pronce, J.-F., and Pollard, R. M. (2021). Annotating Flavius Josephus in the Early Middle Ages: Early Impressions from Thousands of Notes. Medievalia et Humanistica 46, 167-200.

Baird, I. (2021). Outliers, Connectors, and Textual Periphery: John Dennis's Social Network in The Dunciad in Four Books. In I. Baird (Ed.), Data Visualization in Enlightenment Literature and Culture (pp. 265-308): Palgrave MacMillan

Baldzuhn, M. (2011). Schoolbooks. In A. Classen (Ed.), Handbook of medieval studies. Terms - methods - trends, 2061-2069. Berlin/Boston: De Gruyter.

Bauer, B., Hofman, R., and Moran, P. (2017). St. Gall Priscian Glosses, version 2.0. www.stgallpriscian.ie

Bauer, B. (2019a). Venezia, Biblioteca Marciana, Zanetti Lat. 349: an isolated manuscript? A (network) analysis of parallel glosses on Orosius's *Historia adversus paganos*. Etudes Celtiques 45, 91-105.

Bauer, B. (2019b). The Celtic Parallel Glosses on Bede's "De natura rerum". Peritia 30, 31-52.

Bauer, B. (2019c). The interconnections of St Gall, Stiftsbibliothek, ms 251 with the Celtic Bede manuscripts. Keltische Forschungen 8, 31-48. doi:10.25365/kf-8-2019-31-48

Bauman, S., et al. (2012). Text Analysis Meets Text Encoding. Digital Humanities 2012. Hamburg.

Beshero-Bondar, E. (2020). An Introduction to Network Analysis and Cytoscape for XML Coders. Retrieved from https://ebeshero.github.io/thalaba/cytosc.html

Beshero-Bondar, Elisa, and Kellie Donovan-Condron. 2017. 'Modelling Mary Russell Mitford's Networks', Women's Literary Networks and Romanticism:" a Tribe of Authoresses", 1: 137.

Bingenheimer, Marcus, Jen-Jou Hung, and Simon Wiles. 2011. 'Social network visualization from TEI data', Literary and Linguistic Computing, 26: 271-78.

Boillet, M., Bonhomme, M.-L., Stutzmann, D., Kermorvant, Ch. (2019). HORAE: an annotated dataset of books of hours. In Proceedings of the 5th International Workshop on Historical Document Imaging and Processing. doi:10.1145/3352631.3352633

Boodts, S. (2019). Navigating the Vast Tradition of St. Augustine's Sermons: Old Instruments and New Approaches. Augustiniana 69, 83-115.

Carter, D. (2015). Encoding and Representing Repetition in Lyn Hejinian's My Life. Scholarly Editing, 36.

Cinato, F. (2015). Priscien glosé: l'Ars grammatica de Priscien vue à travers les gloses carolingiennes. Turnhout: Brepols.

Conti, A. (2020). The stemma: A typology of variation and error. In P. Roelli, T. Andrews (Eds.), Handbook of Stemmatology (pp. 242-253). Berlin/Boston: De Gruyter. doi:10.1515/9783110684384

Dahlström, Mats. 2000. 'Drowning by versions', Human IT: Journal for Information Technology Studies as a Human Science, 4.

Dalbello, Marija. 2011. 'A genealogy of digital humanities', Journal of Documentation, 67: 480-506.

Dionisotti, A. C. (1996). On the nature and transmission of Latin glossaries. In J. Hamesse (Ed.), Les Manuscrits des lexiques et glossaires de l'antiquité tardive à la fin du moyen âge (pp. 202-205). Leuven: FIDEM.

Dogrusoz, U., Giral, E., Cetintas, A., Civril, A., and Demir E. (2009). "A Layout Algorithm For Undirected Compound Graphs", *Information Sciences*, 179, pp. 980-994, 2009

Dogrusoz , U., A. Karacelik, I. Safarli, H. Balci, L. Dervishi, and M.C. Siper, "Efficient methods and readily customizable libraries for managing complexity of large networks", PLoS ONE, 13(5): e0197238, 2018.

Eisenhut, H. (2010). Die Glossen Ekkeharts IV. im Codex Sangallensis 621. University of Zürich. orosius.monumenta.ch

Fischer, Frank, and Daniil Skorinkin. 2021. 'Social Network Analysis in Russian Literary Studies.' in, The Palgrave Handbook of Digital Russia Studies (Palgrave Macmillan, Cham).

Flight, C. (1992). Stemmatic theory and the analysis of complicated traditions. Manuscripta 36:1, 37-52.

Floyd, Caroline. 2019. 'A Social Network and Text Analysis of Charles Darwin's Correspondence, 1835-1842', University of Guelph.

Fraistat, N., & Flanders, J. (2013). Introduction: Textual Scholarship in the Age of Media Consciousness The Cambridge Companion to Textual Scholarship (pp. 1-15). Cambridge: Cambridge University Press.

Franz, M. et al. (2016) "Cytoscape.js: a graph theory library for visualisation and analysis." Bioinformatics (Oxford, England) vol. 32,2: 309-11. doi:10.1093/bioinformatics/btv557

Grifoni, C. (2003). Otfridi Wizanbvrgensis Glossae in Matthaevm. Corpus Christianorum. Continuatio mediaevalis 200. Turnhout: Brepols.

Hadden, Richard. 2016. "More than a pretty picture: network visualisation as an interface for Digital Scholarly Editions." In Digital Scholarly Editions as Interfaces, 12. Graz.

Hall, B. (2021). Text Mining and Data Visualization: Exploring Cultural Formations and Structural Changes in Fifty Years of Eighteenth-Century Poetry Criticism (1967–2018). In I. Baird (Ed.), Data Visualization in Enlightenment Literature and Culture (pp. 153-195): Palgrave McMillan.

Hoenen, A. (2020). Computational methods and tools: History of computer-assisted stemmatology. In P. Roelli, J. van Zundert (Eds.), Handbook of Stemmatology (pp. 294-302). Berlin/Boston: De Gruyter. doi:10.1515/9783110684384

Jung, B., Brohan, J., Maher, R., O'Shea, L., & Wade, V. (2003). Stirring XML: Visualisations in SVG. Paper presented at the Conference Proceedings SVG Open.

Keskiaho, J. (2019). Copied Marginal Annotations and the Early History of Augustine's "De civitate Dei". Augustiniana 69, 277-298.

Kostoff-Kaard, J. L. (2015). The “Glossa Ordinaria” on Ecclesiastes: A Critical Edition with Introduction. PhD Thesis, University of Toronto.

Lambert, P.-Y. (1983). Les commentaires celtiques a Bède le vénérable. Etudes celtiques 20, 119-143.

Lindsay, W.M. (ed.) (1911). Isidori Hispalensis episcopi Etymologiarum sive Originum libri XX. Oxford: Clarendon Press. Retrieved from http://www.thelatinlibrary.com/isidore/1.shtml.

Mahlberg, M., Stockwell, P., Joode, J. d., Smith, C., & O'Donnell, M. B. (2016). CLiC Dickens: novel uses of concordances for the integration of corpus stylistics and cognitive poetics. Corpora, 11(3), 433-463.

Moran, P. (2022). Gloss Corpus, version 0.4.http://www.glossing.org/glosscorpus/.

Morard, M., Bataillon L.-J., Bériou, N., Burghart, M., and Dahan, G. (2016). Glossae scripturae sacrae electronicae. Institut de recherche et d’histoire des textes. https://gloss-e.irht.cnrs.fr

Moretti, Franco. 2013. 'Conjectures on world literature.' in, Distant Reading (Verso: London; New York).

Myrick, L., & Ohge, C. (2017). Mark Twain: April Fool, 1884. Scholarly Editing, 38.

O’Sullivan, S. (2004). Early Medieval Glosses on Prudentius' Psychomachia: The Weitz Tradition. Leiden: Brill.

O’Sullivan, S. (2010). Glossae aevi Carolini in libros I-II Martiani Capellae De nuptiis Philologiae et Mercurii. Corpus Christianorum. Continuatio mediaevalis 237. Turnhout: Brepols.

Palumbo, G. (2020). The genealogical method: Criticism and controversy. In P. Roelli, O. E. Haugen (Eds.), Handbook of Stemmatology (pp. 88-108). Berlin/Boston: De Gruyter. doi:10.1515/9783110684384

Paranyushkin, Dmitry. 2019. "InfraNodus: Generating insight using text network analysis." In The world wide web conference, 3584-89.

Ravenek, W., van den Heuvel, C., & Gerritsen, G. (2017). The ePistolarium: origins and techniques. In J. Odijk & A. Van Hessen (Eds.), CLARIN in the low countries (pp. 317-323). London: Ubiquity Press. doi:10.5334/bbi.26

Robinson, P. (2004). Where we are with electronic scholarly editions, and where we want to be. Jahrbuch für Computerphilologie, 5, 125-146.

Robinson, P. (2005). 'Current issues in making digital editions of medieval texts—or, do electronic scholarly editions have a future?', Digital Medievalist, 1.

Schöch, Christof, José Calvo Tello, Ulrike Henny-Krahmer, and Stefanie Popp. 2019. 'The CLiGS Textbox: Building and Using Collections of Literary Texts in Romance Languages Encoded in TEI XML', Journal of the Text Encoding Initiative.

Shahbazi, Behnam. 2019. 'StoryMiner: An Automated and Scalable Framework for Story Analysis and Detection from Social Media', PhD, UCLA.

Siemens, R. with the TaPOR Community. (2005). Text Analysis and the Dynamic Edition? A Working Paper, Briefly Articulating Some Concerns with an Algorithmic Approach to the Electronic Scholarly Edition. CH Working Papers, A.36.

Steinová, E. (2021). Innovating Knowledge Manuscript database. https://innovatingknowledge.nl/?page_id=33.

Steinová, E. (2020). “Annotation of the Etymologiae of Isidore of Seville in Its Early Medieval Context”, Archivum Latinitatis Medii Aevi 78, 5-81.

Teeuwen, M. (2008a). Carolingian Scholarship and Martianus Capella: The Oldest Commentary Tradition. Huygens Institute. martianus.huygens.knaw.nl

Teeuwen, M. (2008b). “The Pursuit of Secular Learning: The Oldest Commentary Tradition on Martianus Capella.” The Journal of Medieval Latin 18, 36-51.

Teeuwen, M. (2016). Writing in the Blank Space of Manuscripts: Evidence from the Ninth Century. Ars Edendi Lecture Series 4, 1-25.

TEI Consortium. (2021). TEI P5. Guidelines for Electronic Text Encoding and Interchange. Retrieved from https://tei-c.org/release/doc/tei-p5-doc/en/html/

Unsworth, John. 2002. "What is Humanities Computing and What is Not?" In Jahrbuch für Computerphilologie, 71-84.

Unsworth, J., & Mueller, M. (2009). The MONK project final report.

Warntjes, I. (2020). Isidore of Seville and the Formation of Medieval Computus. In A. Fear and J. Wood (Eds.), A Companion to Isidore of Seville, 457-523. Leiden/Boston: Brill.

Wieland, G. R. (1983). The Latin glosses on Arator and Prudentius in Cambridge University Library, MS Gg.5.35. Turnhout: Brepols.

Wieland, G. R. (1985). The glossed manuscript: classbook or library book? Anglo-Saxon England 14, 153-173.

Evina Steinová studied Classical and medieval Latin at the Masaryk University in Brno (Czech Republic) and Medieval Studies at Utrecht University (Netherlands). She received her PhD in Medieval Studies from Utrecht University in 2016. In 2018-2021, she worked as a postdoctoral researcher at the Huygens Institute in Amsterdam leading the *Innovating Knowledge* project dealing with the transmission and reception of the *Etymologiae* of Isidore of Seville in the early Middle Ages. Her scholarly monograph, *Notam superponere studui* (2019) deals with the use of annotation symbols in Late Antiquity and the early Middle Ages. She is also actively engaged in scientific communication in the domain of medieval history and collaborated on three popularizing books on history published in Slovakia.

Peter Boot studied Mathematics (Leiden) and Dutch literature (Utrecht). He wrote a thesis about the possibilities for annotation in digital scholarly editions: *Mesotext. Digitised Emblems, modelled annotations and humanities scholarship*. He works at the Huygens Institute (Amsterdam) where he oversaw the creation of the digital edition of Vincent van Gogh's letters. He works as a consultant on several digital edition projects. His main research interests are the digital scholarly edition and computational literary studies.