

## Research Article

Eva Vrangbæk, Christian Vrangbæk\*

# Modelling the Semantic Landscape of Angels in Augustine of Hippo

<https://doi.org/10.1515/opth-2025-0050>

received November 10, 2024; accepted May 28, 2025

**Abstract:** This contribution explores the application of computationally assisted modelling to analyse angelology in the works of Augustine. We provide a concise outlining and discussion of the strengths and weaknesses of these models. By investigating terms denoting angels and their co-occurring terms within a semantic network, we demonstrate the analytical process. The study begins by introducing the field of computationally assisted text analysis, followed by an outline of our methodology. Subsequently, we analyse angelologic terms such as *angelus*, *homo*, and *daemon* across Augustine's works using word-embedding techniques. We conclude that Augustine's angelology has an anthropological focus on human beings obtaining *aequalitas* with angelic, rational beings.

**Keywords:** Augustine, word-embedding, computational methods, hermeneutics, digital humanities

## 1 Introduction

In this contribution, we wish to analyse angelology in the works of Augustine. The aim is not on producing an overarching claim about Augustine's angelology, rather, in accordance with the aim of the special issue of mathematical models as assisting tools in textual analysis, we wish to focus on a short and concise showcase and discussion of the strengths and weaknesses of applying mathematical models via computationally assisted programming. The article thus both focuses on Augustine's angelology and the methodological aspects. We investigate terms that describe angels and terms co-occurring with angel-like terms. We showcase and discuss how such an analysis takes place. The article begins with introducing the field of computationally assisted text analysis; second, we outline the methodology; and third, we analyse angelologic terms as *angelus*, *homo*, and *daemon* across the collected works of Augustine on the basis of a word-embedding analysis. The central question to be answered is: Can angels in Augustine be mapped through word-embedding?

### 1.1 Computationally Assisted Text Analysis

In recent years, the employment of mathematical modelling via computational methods has risen not only in the humanities in general but also among classicists and New Testament scholars.<sup>1</sup> The variety of applications is almost as diverse as the different research questions, for which reason it becomes almost impossible to

---

<sup>1</sup> See e.g. Špiclová and Kaše, "Distant Reading," 423–39; Keersmaekers and Speelman, "Applying Distributional Semantics," 17–21; Baunvig, "A Computational Future?," 325–52; Vrangbæk et al., "Non-Canonical Acts," 1–10; Vrangbæk, *What is Love*, 12–32.

---

\* **Corresponding author: Christian Vrangbæk**, Department of Theology, Aarhus University: Aarhus Universitet, Aarhus, Denmark, e-mail: chv@cas.au.dk

**Eva Vrangbæk:** Department of Theology, Aarhus University: Aarhus Universitet, Aarhus, Denmark

speak of computational methods *per se*, and each study has its own research question and its own aim.<sup>2</sup> Within the humanities and theological disciplines when working computationally, i.e. analysing texts assisted by mathematical models, we need to be fully open about the parameters set in order for “non-computational,” colleagues to follow along. Our aim in this contribution is thus to analyse the angelology in Augustine with a method called word-embedding while prioritising a detailed description of the steps involved in conducting such an analysis. In accordance with the aim of this special issue of *Mathematical Modelling Early Christian Literature*, our contribution here will evaluate the strengths and weaknesses of this approach, specifically, the use of word-embedding. The computational methods are not a means in themselves, but assisting tools in analysing and unfolding insights of research questions, in this case, Augustine’s angelology, and nevertheless, we wish to prioritise space in the contribution for showcasing and outlining the methodology, as well as the “unstated” or hidden assumptions that lie behind the methodology in order to continue bridging a perceived gap between “digital” and “traditional” research methods.

Computational text analytical methods arise from the possibilities offered from “data mining” and “knowledge discovery in databases” which won terrain in the 1990s.<sup>3</sup> The Italian literary scholar Franco Moretti provocatively encouraged that, “we know how to read texts, now let’s learn *not* to read them.”<sup>4</sup> One of Moretti’s students, Matthew Jockers, followed the lines of Moretti, but advocated also for employing a “blended method.”<sup>5</sup> In recent years, manifold examples of computationally integrative research have been employed.<sup>6</sup>

One of the most recognised forms of computational methods is the stylometric approach of determining authorship or structural questions.<sup>7</sup> The methodological setup of stylometry is usually to take a debate embedded in a research tradition and reassess the question moving from a qualitative analysis to a quantitative stylometric “bottom-up”-analysis, where, for example, minor features of conjunctions, pronouns, word-sequences reveal “hidden” structures and can offer quantitative results.<sup>8</sup> Some computational approaches in recent years take a similar methodological setup, of reassessing a discussion embedded in the literature, for example, in discussing the structure of the Danish mediaeval work of Saxo Grammaticus (ca. 1180-1210), *Gesta Danorum*.<sup>9</sup> In this study, the methodological aim is to “combine automated techniques ... in order to reassess *Gesta Danorum*’s composition debate.” Moreover, it is stated in the study that their approach “is more akin to qualitative text analysis as it is practised in the humanities and social sciences, but it utilises the transparency and formal rigour of mathematics and computation.”<sup>10</sup> The foundational idea here is similar to the methodological approach we stress in this contribution, namely, that a) the need to embed methods in a research context and b) the combination and cross-fertilisation of computational and traditional, close-reading methods.<sup>11</sup>

## 2 Methodology

The basic idea of mathematical modelling via computational methods is to “convert textual data or natural language into structured formats before subjecting them to statistical, mathematical and numerical

<sup>2</sup> A large amount of digital classicist projects concern linguistics e.g. McGillivray et al., “A New Corpus Annotation,” 47–105, while others make modelling assist traditional close reading, e.g. Nielbo et al., “Automated Compositional,” 1–14; Vrangbæk and Nielbo, “Love Between,” 130–49.

<sup>3</sup> Fayyad et al., “From Data Mining,” 37; Nielbo et al., “Quantitative Text Analysis.”

<sup>4</sup> Moretti, *Distant Reading*, 48.

<sup>5</sup> Jockers, *Macroanalysis*, 26.

<sup>6</sup> Baunvig, “A Computational Future?,” 325–52; Nielbo et al., “Automated Compositional,” 1–14; Vrangbæk and Nielbo, “Love Between,” 130–49; Špiclová and Kaše, “Distant Reading,” 423–39.

<sup>7</sup> Zhang et al., “Did Gaius Julius,” 28–58; Deneire, “Filelfo,” 239–70.

<sup>8</sup> Zhang et al., “Did Gaius Julius,” 28–58; Jockers, *Macroanalysis*, 63–104; For a critique of bottom-up stylometry see Robertson, *Paul’s Letters*; and Robertson, *Statistical Approaches*.

<sup>9</sup> Nielbo et al., “Automated Compositional,” 1–14.

<sup>10</sup> Ibid.

<sup>11</sup> Ibid.; Vrangbæk, *What is Love*, 23–5.

analysis.”<sup>12</sup> However, behind such a description lies a process with three main phases: Text/data-preparation, modelling (in this case word-embedding), and interpretation. The text/data-preparation and the modelling are outlined in Sections 2.1 (Sections 2.1.1 and 2.1.2) and 2.2. The analysis and interpretation of Augustine’s angelology is presented in Section 3.

## 2.1 Text as Data and Its Preparation

### 2.1.1 Textual Database

The transformation of textual data relies on robust digital text editions. Projects employing computational methods frequently depend on freely available online texts, which tend to be older and lack critical text editing. Additionally, finding comprehensive scanned collections of required texts is often difficult. Consequently, this may lead to a somewhat randomly assembled collection of sources, with available texts not always thoroughly checked for scanning errors.<sup>13</sup> In some cases, if data are insufficient, then researchers now and then turn to augmentation of data.<sup>14</sup> In the earlier phases of computational methods, like in *knowledge discovery in databases*, it was a goal in itself to master “extracting high-level knowledge from low-level data.”<sup>15</sup>

In the case of Augustine studies, there exists a solid digital edition of Augustine’s complete works, namely, the Corpus Augustinianum Gissense (CAG).<sup>16</sup> The CAG edition is a fruitful conglomerate of the best text-critical versions from standard editions as *Corpus Scriptorum Ecclesiasticorum Latinorum* (CSEL) and *Corpus Christianorum Latinorum* (CCL), and *Source Chrétiennes* (SC).<sup>17</sup> That we can employ the textual data of the CAG edition provide that our texts (data) are of high quality in the sense that we have all of Augustine’s works in a solid digital state. Moreover, when computational techniques are applied to texts, it is important that we have enough textual material, following the principle of there are no data like more data. If we consider Augustine’s texts as “data” that have a high quality, then we are also more prone to work with them in traditional ways of qualitative text analysis in close interaction with computational methods.

### 2.1.2 Vectorisation

In order for modelling to commence, we need to convert the unstructured text into structural data. To enable a computer to process a text, the text must be structured and converted into numerical vectors (*vectorisation*). The manner in which this is executed naturally influences the outcome. Different methods are applicable to different experiments and modelling choices. In word embedding, an advanced and context-based vectorisation is needed to convert the text into a multidimensional vector space.<sup>18</sup>

Vectorisation of words is possible in several ways which depends on the type of investigation. Common vectorisation models are the *bag-of-words* (BOW)-model and *term frequency-inverted document frequency* (TF-IDF)-model. The bag-of-words (BOW) model functions like a basic word count. In contrast, the term frequency-inverse document frequency (TF-IDF) model is an enhanced version of the BOW model where weights are assigned based on a logarithmic scale. This scale considers term frequency (TF) and inverse document frequency (IDF), which indicates how unique or distinct a word is across the entire corpus. Words that are

<sup>12</sup> Nielbo et al., “Quantitative Text Analysis,” 2.

<sup>13</sup> Nielbo et al., “Mining the Past,” 95; Vrangbæk, *What is Love*, 32–3.

<sup>14</sup> Gupta, “Data Augmentation.”

<sup>15</sup> Fayyad et al., “From Data Mining,” 39.

<sup>16</sup> <https://www.augustinus.de/projekte-des-zaf/corpus-aug-gissense/cag-online/cag-online-english>.

<sup>17</sup> <https://www.augustinus.de/764-al-cag-online>.

<sup>18</sup> Studies on the choices about which vectorisation to choose are for example: Mikolov et al., “Efficient Estimation;” Sineglazov and Savenko IM, “Comparative Analysis;” Turian, “Word Representations.”

rare or distinctive are given greater importance in capturing the essence and information of a document.<sup>19</sup> The TF-IDF-model is a weighted occurrence model and useful, e.g., for calculating *lexical impact*. This is, however, not suitable for word embedding, which we apply here where the more sophisticated *context-based model* of CBOW (*continuous bag of words*) has been applied. The CBOW model is using “a simple neural network architecture that aims to predict the neighbors of a word.” This neural network holds information of the distributed representations of words, and it has been shown that these representations can capture linguistic regularities of the texts.<sup>20</sup> The context-based vectorisation preprocessing is what lies behind “word embeddings.”<sup>21</sup>

## 2.2 Word Embedding

After structuring and transforming the text into numerical data, the data are ready to be modelled. At this stage, a wide range of modelling are available for extracting patterns and information from the text, which depends on the research question at hand. Since we in this investigation focus on angelic terms in a large text corpus, we have selected the method of word embedding, which is suitable for this purpose. As mentioned earlier, word embedding vectorises text with a CBOW method, but unlike methods such as BOW and TF-IDF, CBOW does so through a “training” process involving machine learning principles.<sup>22</sup>

Word embedding involves “training” an artificial neural network on a large corpus of text. This training requires the algorithm to iterate over the text repeatedly, learning to predict the likelihood of a word based on its context. Through this process, it develops weights that determine the parameters of the resulting semantic vector space. The model is designed to assign similar vectors to words that appear in similar contexts, effectively capturing contextual or distributional semantics and relationships between words. Consequently, even if certain words are not common across documents, their vectors still hold dense and meaningful values derived from the contexts in which these words occur. The data size has been Augustine’s complete works which amounts to approximately 5 million tokens.<sup>23</sup>

Using word embedding focuses on the context, i.e. the internal context in a corpus. Words within the corpus are given weights, and thereby meaning, through their co-occurrences. Each word is assigned a value not just based on its frequency but also on its contextual meaning, derived from the words surrounding it during vectorisation. The representations and the semantic space they span is a simple implementation of a view on language called distributional semantics. Within distributional semantics, the meaning of a word is dependent on its distributional position in the semantic vector space, or in other words, that the meaning of a word is “characterised by the company it keeps.”<sup>24</sup> The “embedding” refers to the process of representing words as dense vectors in a semantic vector space, in other words, words are embedded in the contextual corpus.<sup>25</sup>

The computational word embedding analysis is unsupervised since there are no predetermined labels. The information and patterns are calculated solely within the corpus. The resulting semantic vector space contains information on every word across the entire corpus, and in order for Augustine scholars to survey, for example, angelology, we need to extract portions of the semantic vector space. The process is thus in two steps: the first with a computational unsupervised step, and the second a supervised step, where informed Augustine scholars query labelled seed-terms. While such a mix of unsupervised word embedding and

<sup>19</sup> For TF-IDF see e.g. Zhang et al., “A Comparative Study,” 2758–65; Albrecht et al., *A Blueprint*, 132–3.

<sup>20</sup> Mikolov et al., “Exploiting Similarities,” 3.

<sup>21</sup> Goldberg, *Neural Network*, 90–2; Albrecht et al., *A Blueprint*, 270–1; Vrangbæk, *What is Love*, 30–2.

<sup>22</sup> Sineglazov and Savenko IM, “Comparative Analysis,” Goldberg, *Neural Network*, 90–2.

<sup>23</sup> Baunvig and Nielbo, “Mermaids,” 24–5.

<sup>24</sup> Attributed to John Firth, cf. Harris, “Distributional Structure,” 146–62; Boleda, “Distributional Semantics,” 2.

<sup>25</sup> Jurafsky and Martin, “Speech and Language Processing,” Li and Yang, “Word Embedding,” 83–104; Mandikal and Mooney, “Sparse Meets Dense,” Goldberg, *Neural Networks*, 90ff; Vrangbæk, *What is Love*, 27–8.

supervised input, it is not strictly a semi-supervised method; however, our approach reveals a mix of unsupervised and supervised methods.<sup>26</sup>

The output of a word embedding analysis is a pattern of the so-called co-occurrence structure, i.e. how words co-occur based on their contextual embeddings, where distance/proximity is measured between vectors with cosine similarity.<sup>27</sup> When we query angelologic terms, we receive structural information concerning these terms. The results come in a list of words with first-level and second-level associations, which is visualised in semantic networks that gives a picture of the semantic “landscape” around the queried terms. The output of a word-embedding analysis needs to be evaluated, and further, it needs to be interpreted and integrated into the broader analysis.<sup>28</sup>

A first-level association is a term that is primarily associated with the seed term. An example from angelus’ network in our study could be *nuntius*, which is a first-level association to *angelus*. This means that the vector of *nuntius* has a low cosine distance, and therefore high similarity to the seed term, *angelus*, in the semantic vector space. A second-level association is a term that has the same relationship, not directly to the seed term, but directly related to one of the first-level associations. In the semantic kernel of *angelus*, *interpreter* is therefore not linked directly to the seed term *angelus* but to the first-level association *nuntius*. If we had *nuntius* as our seed term, then the second-level association *interpreter* would appear as first level.<sup>29</sup>

### 3 Analysis: Mapping Angelus

Even though the idea of angels may seem as pre-rational speculation, we would miss a key component of Augustine’s theology if we left it out of the equation of his thinking. Angels and their assigned roles in the cosmos were real phenomena to Augustine. The biblical stories about them were not allegories, but rather stories about real divine agents.<sup>30</sup> Most prominent in Augustine’s thinking stands his analyses of the angels’ position in the cosmos in his *De civitate Dei*, where angels are described as the ones who chose either the heavenly or the earthly city at their creation and remaining in that state until the end of time.<sup>31</sup> But angels are not only present in the cosmic scheme, they are also present as concrete beings and also on the more personal level.

In the following analyses, the method of word embedding is employed to map the semantic landscape of Augustine’s use of terms that would reveal information about his angelology. The aim is not to make a new and competing claim on Augustine’s view on angels, rather, in line with this special issue, and we wish to test the applicability of word embedding and explore what it reveals on Augustine’s angelology.

#### 3.1 Angelus

To investigate the angelology of Augustine, we begin by creating an individual semantic kernel with *angelus* as seed term. To create a comprehensive yet manageable network, we have chosen a network with six first-level associations and four second-level associations. More or less words could be included in the visualisation, since all vectors have been stored in the trained model. The result of this first word-embedding analysis with

<sup>26</sup> Jabin and Berlyand, “Supervised, Unsupervised, and Semisupervised Learning,” 19–23.

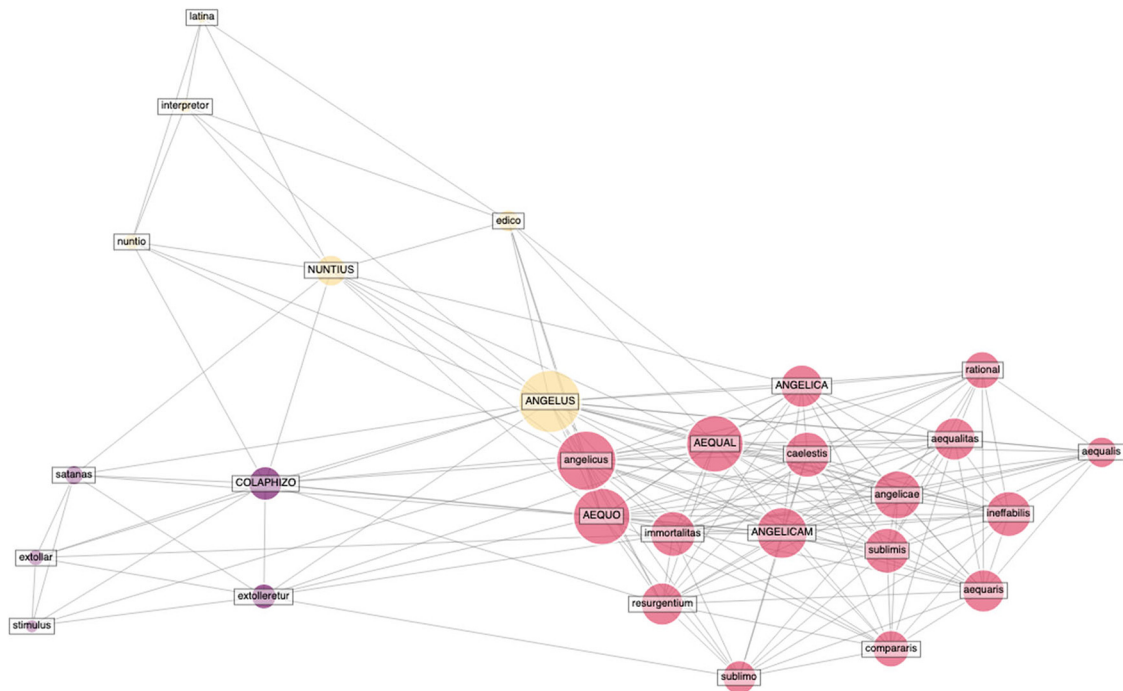
<sup>27</sup> Mikolov et al., “Efficient Estimation,” 5.

<sup>28</sup> Baunvig and Nielbo, “Mermaids,” 23–32; Vrangbæk 2024, 19–21.

<sup>29</sup> The description of word-embedding builds on previous analyses with word-embedding in Vrangbæk and Nielbo, “Love Between,” Vrangbæk, *What is Love*.

<sup>30</sup> Klein, *Augustine’s Theology of Angels*, 1–2; Muehlberger, *Angels*, 18; Ployd, “Participation and Polemics,” 421–2.

<sup>31</sup> Babcock, “The Human Fall,” 133–5; Vrangbæk, “The Fall of the Will,” 139–40.



**Figure 1:** Semantic network with “angelus” as seed term, and 6 first-level associations and 4 second-level associations.

angelus as seed term is a list of 28 words, of which 6 are first-level associations, and 21 second-level associations. The network is visualised in Figure 1.

The nodes and edges represent the distance and proximity between the terms, and the colouring of the clusters is based on an extra layer of community detection between the terms, which is built with the Louvain best partition method.<sup>32</sup> The semantic network of angelus reveals three clusters.

The largest and most dense is the red cluster, which contains terms such as aequalis, ANGELICAM, angelica, angelicus, angelica, AEQUO, caelestis, immortalitas, resurgendum, sublimo, ineffabilis, rationalis, a cluster that can be interpreted as a topic of become equal with angels and obtaining their heavenly status and immortal being.<sup>33</sup> In his writings, Augustine describes angels as having spiritual bodies (Gn.litt. 6,19; 6,24), which is rendered in the associative terms of caelestis, immortalitas, and sublimo, but what is more striking is the strong presence of the aequalitas theme. It is important to note that while such a topic is present in the *In Iohannis Evangelium Tractatus*, where Augustine writes about the aequalitas which human beings hope for (Io.eu.tr. 100,6), our network is a representation of all occurrences of these words in Augustine’s works, and at the same time, a passage from a particular text can, with our knowledge of Augustine, be linked as an illustrative passage that can be interpreted and contextualised with the computationally produced semantic network.<sup>34</sup> The topic of *aequalitas* is, however, needed to be compared with how Augustine views angels as static creatures. Many interpreters have shown how Augustine describes angels as having received a choice at their creation, and unlike human beings, angels are not able to make progress or regress in the salvific scheme. Angels remain the same status from creation, they are eternal and unchangeable beings. This insight needs to be included in the result so that the dynamic of *aequalitas* is not mistaken to concern angels, but rather concerning a perspective toward humans becoming like angels, and not *vice versa*.<sup>35</sup>

<sup>32</sup> <https://python-louvain.readthedocs.io/en/latest/api.html>.

<sup>33</sup> Ideally, we would only have the lemmatised form of angelicus and aequalis and not the inflected variants of the adjectives. This is a sign that no classical language computational lemmatisation is flawless, however, as we can see many of the terms spawn together and display a close semantic relationship, and, at the same time, gives due space to letting other content words.

<sup>34</sup> Mayer, “Aequalitas,” 142–6; van Fleteren, “Angels,” 20–2; Madec, “Angelus,” 302–15.



The second cluster is the yellow cluster. This contains fewer terms and is more sparse. It has the seed-term ANGELUS, followed by NUNTIUS, nuntio, edico, interpreter, latina. This is thus a cluster that can be interpreted along the lines of the communicative aspects of the angels. In his writings, Augustine notes that benevolent spiritual beings are called “angelus/angeli” from Greek, ἄγγελος, and nuntius/nuntii in Latin due to their role as conveyer and speaker of the divine message (Io.ev.tr. 24,7; trin 2,13).<sup>36</sup>

Finally, we have the purple cluster, which is slightly more connected than the yellow, nuntius-cluster. The purple cluster has COLAPHIZO, satanas, extollar, stimulus, and extolleretur. This colaphizo-cluster represents the darker side of angelology, with evil cosmic and moral adversaries, satan, and colaphizo. This last cluster seems to be a topic developed due to that Augustine frequently relies on 2 Cor 12:7: “datus est mihi stimulus carnis meae, angelus satanae, qui me colaphizet,” which Augustine cites, e.g. in *Enarrationes in Psalmos* (130,7) and *De gratia Christi et peccato originali* (12,xi).<sup>37</sup> However, it is important to note that a dualistic interpretation of this finding would be misfitting for Augustine.<sup>38</sup> The spiritual struggle consists of the “temptation, and behind every temptation is the demonic inducement to join the proud,” a struggle which is primarily against self-centeredness and pride.<sup>39</sup> The battle is not fought in dualistic terms, but with Christ’s sacrifice as the model.<sup>40</sup>

From this network with only one seed-word, we are able to discern three aspects of Augustine’s angelology: the angels as messengers (nuntius), the angels as related to evil (satanas) and moral internal combat (stimulus, colaphizo), and finally, the heavenly, rational aspect of angels which aequalitas is desired by human beings. The network is a representation of all occurrences of these words in Augustine’s works, and at the same time, a passage from a particular text can, with our knowledge of Augustine, be linked as an illustrative passage that can be interpreted and contextualised with the computationally produced semantic network.

The prominence of “nuntius” may communicate with Ellen Muehlberger’s analysis, which shows how Augustine reinterpreted the Old Testament by removing divine actors and replacing them with angels as messengers, thereby resolving theological issues concerning the materiality and visibility of God.<sup>41</sup> Elizabeth Klein has shown further, how Augustine interprets divine apparitions that are described in the Old Testament, that “angels act in persona Dei, representing God,” and compares with the roles of the prophets. This will corroborate why our finding has “nuntius” as a prominent co-occurring word to angelus.<sup>42</sup>

Concerning the topic of “aequalitas,” it is important to note that this topic is present in the *In Iohannis Evangelium Tractatus*, where Augustine writes about the aequalitas which human beings hope for (Io.eu.tr. 100,6).<sup>43</sup> The topic of *aequalitas* is, however, needed to be compared with how Augustine views angels as static creatures. Many interpreters have shown how Augustine describes angels as having received a choice at their creation, and unlike human beings, angels are not able to make progress or regress in the salvific scheme. Angels remain the same status from creation, they are eternal and unchangeable beings. This insight needs to be included in the result so that the dynamic of *aequalitas* is not mistaken to concern angels, but rather concerning a perspective toward humans becoming like angels, and not *vice versa*.<sup>44</sup>

In the following, we will investigate two new networks that enable us to dive deeper into the interconnections of angelus by investigating homo and daemon in connection to angelus.

35 Muehlberger, *Angels in Late Ancient Christianity*, 29–32; Ployd, “Participation and Polemics,” 421–39.

36 van Fleteren, “Angels,” 20–2.

37 Madec, “Angelus,” 312.

38 Klein, *Augustine’s Theology of Angels*, 167–8.

39 Ibid., 157–8.

40 Ibid., 168.

41 Muehlberger, *Angels in Late Ancient Christianity*, 87–8.

42 Klein, *Augustine’s Theology of Angels*, 114.

43 Mayer, “Aequalitas,” 142–6; van Fleteren, “Angels,” 20–2; Madec, “Angelus,” 302–15.

44 Muehlberger, *Angels in Late Ancient Christianity*, 29–32; Ployd, “Participation and Polemics,” 421–39.

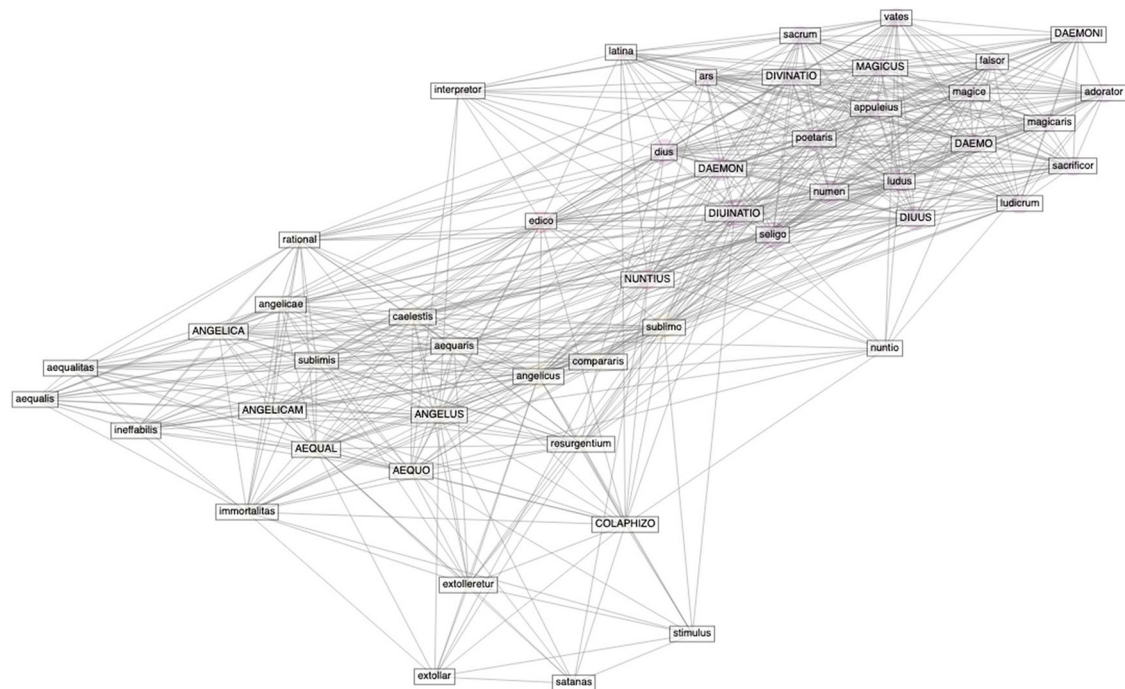


Figure 2: Angelus-Daemon 6-4.

### 3.2 Angelus, Homo, Daemon, and Their Relation

In this next part of the analysis, we will focus on two new networks, where we query, first, *angelus* together with *daemon*, and, second, *angelus* together with *homo*. In the first, we analyse *angelus* and *daemon*, and the results are shown in Figure 2.

The network forms two discernable clusters, one with angelus and its three subclusters as we saw in the previous analysis, and one with daemon. The terms that are contained in the daemon-cluster are as follows: DIUINATIO, DAEMO, DAEMONI, DIVINATIO, seligo, DIUUS, falsor, dius, appuleius, ars, poetaris, ludicrum, sacrificor, magiaris, adorator, numen, MAGICUS, magice, vates, sacrum, ludus.

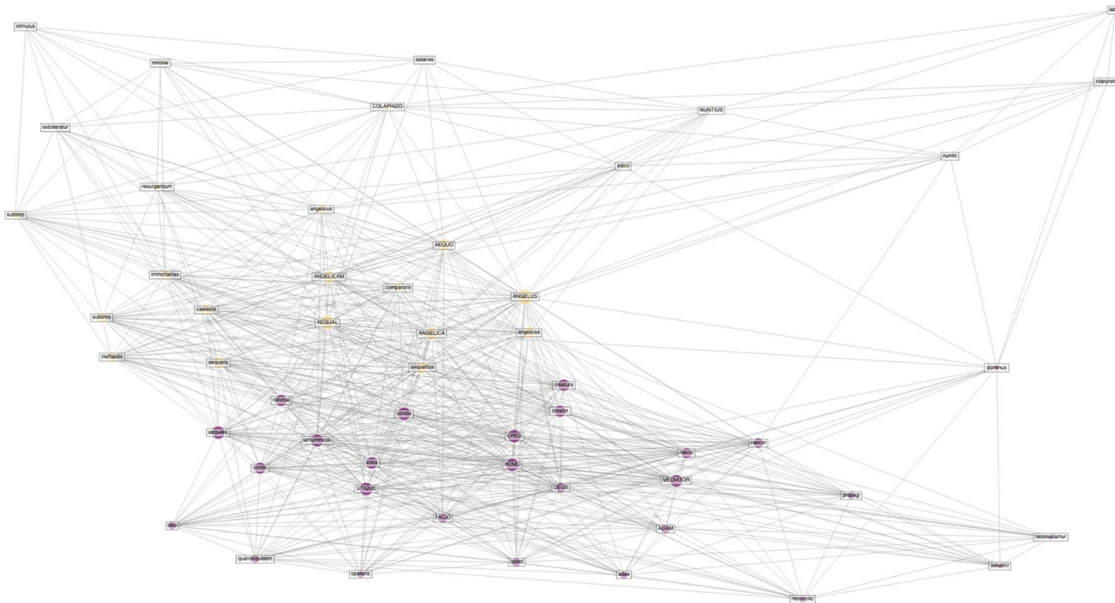
The first observation is that the daemon-cluster is more dense than the angelus-cluster. Moreover, concerning the relation to terms in the angelus-cluster, the term *satanas* is not directly linked to daemon (it is neither linked to *diabolus*, which we tested, but not visualised). This makes an interesting case, namely that daemon/daemones in Augustine's work is limited to the treatment of daemones in esp. *De divinatione daemonum*, which concerns the magical ability of daemones to predict (divinatio) future events. We observe in this context also ludus, ars, ludicrum, and poetaris as terms that fall under this topic. Moreover, Augustine's discussions of daemones in *De civitate dei* chaps 8–10 concerning communication with false daemones are also reflected in the terms falsor, adorator.<sup>45</sup> Again, if we test the connection with malus, noceo, and other “negative” terms, daemon still forms a separate, dense cluster as in Figure 2. Finally, the position of nuntius is remarkably right between daemon and angelus, which reflects how Augustine discusses the role of daemones in the Greco-Roman world as messengers between gods and humans (Figure 3).<sup>46</sup>

In the following investigation, we queried *angelus* together with *homo*. This network has the same parameters, 6 first-level associations and 4 second-level associations. Here, we can also see two-coloured clusters, the purple coloured is around *homo*, and the yellow is around *angelus*. Even though the Louvain best-partition-layer (cf. note 32) has coloured the groups and thus detects two separate groups, it is nevertheless remarkable

<sup>45</sup> van Fleteren, “Demons,” 266–7.

<sup>46</sup> Ibid.; de civ. 8, 20.





**Figure 3:** Homo and angelus 6-4.

how homo and angelus are much closer in the semantic network than the relation between angelus and daemon in Figure 2.

The homo-cluster contains creator, condio, ADAM, CREO, FACIO, MEDIATOR, UTIQUE, DEUS, opero, salvator, and reconciliator. We notice the bridging terms of creator and condio between angelus and homo, both are created beings. This reflects several aspects of Augustine's discourse on creation, e.g. how the angels were witness to creation and how angels understand the relation between incarnation and creation.<sup>47</sup>

We notice how the proximity of aequalitas and related terms, and most remarkable, that aequalis and rationalis, which were in the outskirts of the network of angelus, are now clustering closer with the homo-cluster. This adherence of rationalis and aequalis in the homo-cluster nuances the topic of aequalitas: Initially, it is not only concerning angelus/angeli, and, further, it does not only describe a topic common to both humans and angels, but rather it is telling of the anthropological focus of Augustine, where aequalitas concerns how humans as rational beings in the end will be transformed to an angel-like state. This transformation is supported by Augustine's reading of Luke 20,36 concerning becoming angel-like; however, it is not a corporeal likeness, but a spiritual likeness where humans share the love and true worship with angels toward God.<sup>48</sup>

## 4 Concluding Remarks

In this contribution, we have examined, with a focus on an exploratory aspect (Can we use word-embedding to map angelus?), the semantic landscape of angelus across the works of Augustine. Within the scope of evaluating the strengths and weaknesses in the application of mathematical modelling Early Christian literature, we outlined and exemplified the use of one specific technique, namely word embeddings. On the one hand, the weakness is that lemmatisation and preprocessing can involve some inaccuracies, however, on the other hand, this has in our case not significantly affected the outcome, since we have been able to detect meaningful relations between terms which can foster interpretation. We have mapped Augustine's angelology based on three semantic networks with angelus,

<sup>47</sup> Klein, *Augustine's Theology of Angels*, 12–56; 131–2.

<sup>48</sup> Ibid., 190–1.

angelus-daemon, and angelus-homo as the bearing seed-terms. Angels, as revealed from the network of angelus, contained three aspects, the celestial aspect, the messenger aspect and the adversary, both cosmic (satanas) and individually (colaphizo). When we investigated daemon, it was connected to angelus; however, it appeared also to be an independent topic. The most significant find appeared, querying homo together with angelus, and revealed that Augustine's angelology, despite its obvious celestial and quasi-divine connotations, clearly has an anthropological focus on humans obtaining *aequalitas* with angelic, rational beings.

**Acknowledgments:** We would like to thank *Zentrum für Augustinforschung* and *Schwabe Verlag* for the use of *Corpus Augustinianum Gissense* (CAG). We also wish to acknowledge the work of the Centre for Humanities Computing, Aarhus University.

**Funding information:** The research in this article was funded by the Carlsberg Foundation in the Semper Ardens: Accelerate-project "Computing Antiquity: Computational Research in Ancient Text Corpora."

**Author contributions:** CHV: writing – original draft, writing – review and editing, methodology, and analysis; EEHV: conceptualisation, writing – review and editing, and database building and hosting.

**Conflict of interest:** Authors state no conflict of interest.

**Data availability statement:** The code and applications are built on an algorithm called Word2Vec from the Gensim-library and visualised with Plotly, and it is accessible at [https://github.com/eehvrangbaek/aug\\_will\\_love/tree/main](https://github.com/eehvrangbaek/aug_will_love/tree/main).

## References

- Albrecht, Jens, Sidharth Ramachandran, and Christian Winkler. *Blueprints for Text Analytics Using Python: Machine Learning-Based Solutions for Common Real World (NLP) Applications*. 1st ed. Sebastopol, CA: O'Reilly Media, 2021.
- Babcock, William S. "The Human and the Angelic Fall: Will and Moral Agency in Augustine's City of God." In *Augustine from Rhetor to Theologian*, edited by Joanne McWilliam, 133–49. Waterloo: Wilfrid Laurier University Press, 1992.
- Baunvig, Katrine F. and Kristoffer L. Nielbo. "Mermaids Are Birds Embedding NFS Grundtvig's Bestiary." *Proceedings*, 2022. <http://ceur-ws.org>. ISSN 1613:0073.
- Baunvig, Katrine Frøkjær. "A Computational Future? Distant Reading in the Historical Study of Religion." In *Stepping Back and Looking Ahead: Twelve Years of Studying Religious Contact at the Käte Hamburger Kolleg Bochum*, 325–52. Leiden: Brill, 2023.
- Boleda, Gemma. "Distributional Semantics and Linguistic Theory." *Annual Review of Linguistics* 6:1 (2020), 213–34. doi: 10.1146/annurev-linguistics-011619-030303.
- Deneire, Tom. "Filelfo, Cicero and Epistolary Style: A Computational Study." In *Francesco Filelfo, Man of Letters*, edited by Jeroen de Keyser, 239–70. Leiden: Brill, 2018.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "From Data Mining to Knowledge Discovery in Databases." *AI Magazine* 17:3 (1996), 37–7. doi: 10.1609/aimag.v17i3.1230.
- van Fleteren, Frederick. "Angels." In *Augustine through the Ages an Encyclopedia*, edited by Allan D. Fitzgerald, 20–2. Grand Rapids, Mich: Eerdmans, 2009a.
- van Fleteren, Frederick. "Demons." In *Augustine through the Ages an Encyclopedia*, edited by Allan D. Fitzgerald, 266–8. Grand Rapids, Mich: Eerdmans, 2009b.
- Goldberg, Yoav. *Neural Network Methods for Natural Language Processing*. Cham: Springer Nature, 2022.
- Gupta, Rahul. "Data Augmentation for Low Resource Sentiment Analysis Using Generative Adversarial Networks." *arXiv*, 2019. doi: 10.48550/arXiv.1902.06818.
- Harris, Zellig S. "Distributional Structure." *Word* 10:2–3 (1954), 146–62. doi: 10.1080/00437956.1954.11659520.
- Jabin, Pierre-Emmanuel and Leonid Berlyand. "Supervised, Unsupervised, and Semisupervised Learning." In *Mathematics of Deep Learning*, 19–23. Germany: Walter de Gruyter GmbH, 2023. doi: 10.1515/978311025551-005.
- Jockers, Matthew. *Macroanalysis: Digital Methods and Literary History*. Illinois: University of Pennsylvania Press, 2013.
- Jurafsky, Daniel and James H. Martin. "Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models," 3rd edn. Online manuscript released January 12, 2025. <https://web.stanford.edu/~jurafsky/slp3>.

- Klein, Elizabeth. *Augustine's Theology of Angels*. Cambridge: Cambridge University Press, 2018. doi: 10.1017/9781108335652.
- Keersmaekers, Alek and Dirk Speelman. "Applying Distributional Semantic Models to a Historical Corpus of a Highly Inflected Language: The Case of Ancient Greek." *Glottometrics* 55 (2023), 17–43.
- Li, Yang and Tao Yang. "Word Embedding for Understanding Natural Language: A Survey." In *Guide to Big Data Applications*, edited by S. Srinivasan, Vol. 26, 83–104. Studies in Big Data. Cham: Springer International Publishing, 2018. doi: 10.1007/978-3-319-53817-4\_4.
- Madec, Goulven. "Angelus." In *Augustinus-Lexikon*, edited by Karl Heinz Chelius, 303–15. Basel: Schwabe, 1986. <http://www.adwmainz.de/projekte/augustinus-lexikon/informationen.html>.
- Mandikal, Priyanka and Raymond Mooney. "Sparse Meets Dense: A Hybrid Approach to Enhance Scientific Document Retrieval." *arXiv.Org* (2024). <https://arxiv.org/abs/2401.04055v1>.
- Mayer, Cornelius. "Aequalitas." In *Augustinus-Lexikon*, edited by Karl Heinz Chelius, 142–6. Basel: Schwabe, 1986. <http://www.adwmainz.de/projekte/augustinus-lexikon/informationen.html>.
- McGillivray, Barbara, Daria Kondakova, Annie Burman, Francesca Dell'Oro, Helena Bermúdez Sabel, Paola Marongiu, and Manuel Márquez Cruz. "A New Corpus Annotation Framework for Latin Diachronic Lexical Semantics." *Journal of Latin Linguistics* 21:1 (2022), 47–105. doi: 10.1515/joll-2022-2007.
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. "Efficient Estimation of Word Representations in Vector Space." *arXiv*, 2013. <http://arxiv.org/abs/1301.3781>.
- Mikolov, Tomas, Quoc V. Le, and Ilya Sutskever. "Exploiting Similarities among Languages for Machine Translation." *arXiv*, 2013. doi: 10.48550/arXiv.1309.4168.
- Moretti, Franco. *Distant Reading*. London: Verso Books, 2013.
- Muehlberger, Ellen. *Angels in Late Ancient Christianity*, edited by Ellen Muehlberger. Oxford: Oxford University Press, 2013. doi: 10.1093/acprof:oso/9780199931934.003.0001.
- Nielbo, Kristoffer L., Folger Karsdorp, Melvin Wevers, Alie Lassche, Rebekah B. Baglini, Mike Kestemont, and Nina Tahmasebi. "Quantitative Text Analysis." *Nature Reviews Methods Primers* 4:1 (2024), 25.
- Nielbo, Kristoffer L., Mads L. Perner, Christian Larsen, Jonas Nielsen, and Laursen Ditte. "Automated Compositional Change Detection in Saxo Grammaticus' Gesta Danorum." In *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference, Copenhagen, Denmark, March 5–8, 2019*, edited by Costanza Navarretta, Manex Agirrezabal, and Bente Maegaard, 14. DHN 2019.
- Nielbo, Kristoffer L., Ryan, Nichols, and Edward, Slingerland. "Mining the Past – Data-Intensive Knowledge Discovery in the Study of Historical Textual Traditions" *Journal of Cognitive Historiography* 3:1–2 (2016), 93–118. doi: 10.1558/jch.31662.
- Ployd, Adam. "Participation and Polemics: Angels from Origen to Augustine." *Harvard Theological Review* 110:3 (2017), 421–39. doi: 10.1017/S0017816017000165.
- Robertson, Paul. *Paul's Letters and Contemporary Greco-Roman Literature: Theorizing a New Taxonomy*. Leiden: BRILL, 2016.
- Robertson, Paul. *Statistical Approaches to Paul's Letters: Distributions, Visualization, Cluster Mapping, and Topology*. Leiden: Brill, forthcoming.
- Sineglazov, Victor M. and I. M. Savenko IM. "Comparative Analysis of Text Vectorization Methods." (2023). <https://er.nau.edu.ua/handle/NAU/61250>.
- Špiclová, Zdeňka and Vojtěch Kaše. "Distant Reading of the Gospel of Thomas and the Gospel of John: Reflection of Methodological Aspects of the Use of Digital Technologies in the Research of Biblical Texts." *Open Theology* 6:1 (2020), 423–39. doi: 10.1515/opth-2020-0111.
- Turian, Joseph, Lev-Arie Ratinov, and Yoshua Bengio. "Word Representations: A Simple and General Method for Semi-Supervised Learning." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, edited by Jan Hajič, Sandra Carberry, Stephen Clark, and Joakim Nivre, 384–94. Uppsala, Sweden: Association for Computational Linguistics, 2010. <https://aclanthology.org/P10-1040>.
- Vrangbæk, Eva Elisabeth Houth. "The Fall of the Will: An Investigation of the Will of Man Before and After the Fall in De civitate Dei." In *Explorations in Augustine's Anthropology*, edited by Fabio Dalpra and Anders-Christian Jacobsen, 131–52. Berlin: Peter Lang, 2021. <https://www.peterlang.com/document/1140471>.
- Vrangbæk, Christian, Eva Vrangbæk, Márton Kardos, Kristoffer Nielbo, and Jacob Mortensen. "Non-Canonical Acts and Their Topical Distribution." In *Proceedings of the First International Workshop of Semantic Digital Humanities (SemDH 2024)*, edited by Oleksandra Bruns, Andrea Poltronieri, Lise Stork, and Tabea Tietz. Vol. 3724. CEUR Workshop Proceedings. Herssonissos, Greece: CEUR, 2024. <https://ceur-ws.org/Vol-3724/#short4>.
- Vrangbæk, Eva Elisabeth Houth and Kristoffer Nielbo. "Love Between Desire and Will: An Investigation of Augustine's Concept of Love Assisted by Computational Methods." In *Augustine and Ethics*, edited by Sean Hannan and Kim Paffenroth, 130–49. Lanham: Lexington Books, 2023.
- Vrangbæk, Eva Elisabeth Houth. *What is Love Other than Will? An Investigation of Augustine's Concept of Love and Will Assisted by Computational Methods*. PhD-Dissertation. Aarhus University, UnPublished, 2024.
- Zhang, Olivia R., Trevor Cohen, and Scott McGill. "Did Gaius Julius Caesar Write De Bello Hispaniensi? A Computational Study of Latin Classics Authorship." *Human IT: Journal for Information Technology Studies as a Human Science* 14:1 (2018), 28–58.
- Zhang, Wen, Taketoshi Yoshida, and Xijin Tang. "A Comparative Study of TF\* IDF, LSI and Multi-Words for Text Classification." *Expert Systems with Applications* 38:3 (2011), 2758–65.