Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-04772-5 OPEN

Wikipedia as a cultural lens: a quantitative approach for exploring cultural networks

Luis A. Miccio[1,2], Paschalis Agapitos[3,4], Carlos Gamez-Perez[5], Francisco González[6], Juan Luis Suarez[7] & Gustavo A. Schwartz[1,3]✉

The structure of cultural networks emerges from complex interactions among many elements related to people, ideas, and objects. However, these interactions can be very subtle and difficult to quantify, precluding a quantitative analysis of the cultural networks that can be crucial to understanding complex dynamics better. In this work, we propose a new approach that combines the formalism of complex networks, the structural relationship between nodes, and the corpus of Wikipedia to map and analyse the interactions among cultural entities. To test the proposed methodology, we study the case of the interdisciplinary cultural network connecting art, science, and philosophy in Europe in the seventeenth century. The results are aligned with well-established historical knowledge of the period and, more importantly, provide new insights to unveil how elements in these networks interact with each other. In particular, we found that nodes within a given cluster, related respectively to art, science or philosophy, interact with nodes in the same cluster following a core-periphery behaviour. In contrast, inter-cluster interactions across disciplines follow a power law distribution.

[1] Centro de Física de Materiales (CSIC-UPV/EHU), San Sebastian, Spain. [2] Institute of Materials Science and Technology (INTEMA-CONICET), Mar del Plata, Argentina. [3] Donostia International Physics Center, Donostia, Spain. [4] Department of Philosophy, University of the Basque Country UPV/EHU, Leioa, Spain. [5] Universitat de Barcelon, Departament d'Educació Lingüística, Científica i Matemàtica, Barcelona, Spain. [6] Universidad de Oviedo, Departamento de Filología Inglesa, Francesa y Alemana, Oviedo, Spain. [7] CulturePlex Lab, Western University, London, ON, Canada. ✉email: gustavo.schwartz@csic.es

Introduction

The development of revolutionary ideas, cultural breakthroughs, or paradigm shifts typically emerges from complex interactions among many people, concepts, and cultural objects, that is, within the cultural networks individuals and groups are part of and create to adapt to their changing environments. In this respect, we understand that culture is "information capable of affecting individual's behaviour that they acquire from other members of their species through teaching, imitation, and other forms of social transmission" (Richerson and Boyd 2006). We mostly focus on the informational, and not the behavioural, dimension of culture as manifested in the cultural network emerging from Wikipedia's internal links. Specifically, we focus on a subset of those internal links connecting 17th-century individuals described as scientists, artists, and/or philosophers by the very text of the English Wikipedia, with the purpose of characterizing this cultural network using some of the most common metrics used in network analysis. Therefore, in this paper we are reassembling (Latour, 2005) a cultural network from 17th-century Europe using the knowledge and informational network created by 21st-century authors and editors in Wikipedia. In order to gain access to this cultural network of the past, we deploy methods and tools from network theory and analysis with the purpose of identifying the cultural aspect of the network, that is, the meaning making inscribed in the network of interactions between these 17$^{\text{th}}$-century scientists, artists and philosophers (Suárez 2025; Suárez et al. 2015; Ibrus et al. 2021).

These interactions are difficult to quantify, and the increasing number of elements to be considered would eventually make it harder, if not impossible, to deeply understand some aspects of the cultural dynamics using only traditional tools. To overcome these limitations, some historiographical approaches have been implemented in the past to analyse and explain both dynamics concentrated over short periods, small communities, or small units of research (events, individuals, etc.) (Burke 1992; Hargadon and Wadhwani 2023; Magnússon 2020; Magnússon and Szijártó 2013), and large periods and/or formations whose dynamics span over decades and even centuries (Braudel 1994; Christian 2020; Turchin 2008; Villmoare 2022). Moreover, several quantitative methodologies have been proposed in the last few years to shed light on the emergence of cultural patterns. Schich et al. have analysed the migration patterns of notable individuals over 2000 years to understand how human culture is disseminated (Schich et al. 2014). Smolla and Akçay use complex network analysis to study cultural evolution and to connect individual behaviour with the emergence of population-level structures (Smolla and Akçay 2019). Brown et al. have used data mining and complex networks to model the social networks underpinning the early modern publication industry in the Spanish Golden Age (Brown et al. 2017). This kind of quantitative study belongs to what is known as cultural analytics (Manovich 2020), an emerging research field that has proven fruitful in showing mathematical approaches that help unveil patterns and regularities in problems traditionally tackled with humanities and social science methods, and that otherwise would remain hidden from our knowledge. Both the emerging discipline of cultural analytics and the approach proposed here align with Mesoudi's, who states that "culture is defined as information rather than behaviour (in anthropological jargon, it is an *ideational* definition of culture). Restricting our definition of culture to information does not mean to say that culturally acquired information does not *affect* behaviour." (Mesoudi 2011) Our approach also aligns with recent attempts across several disciplines to use large, real-world data sets and big-data analysis to better understand historical phenomena (Gao et al. 2012; Michel et al. 2011; Spinney 2012; Villmoare 2022)

Among the corpora for quantitative cultural research, Wikipedia has been extensively used (Arroyo-Machado et al. 2020; Eom et al. 2015; Eom and Shepelyansky 2013; Miccio et al. 2022; Schwartz 2021) in the fields of information retrieval, natural language processing and ontology building (for a systematic review, see Mehdi et al. (Mehdi et al. 2017)). More recently, the Wikipedia's corpus has been used in cultural research focusing on cultural heritage (Rantala et al. 2024), world literature (Fischer et al. 2023) and gender bias (Zheng 2022). Table SI1 in the Supplementary Information file summarizes relevant works that apply the complex networks formalism to Wikipedia. Although it was conceived of (and is massively used) as an encyclopaedia, Wikipedia's content (text, images, and links) provides a valuable corpus for researchers in a broad range of domains. The major advantage of using Wikipedia is that in addition to the explicit knowledge in its articles (text and images), a significant amount of implicit knowledge emerges from the underlying network of internal links. It has been recently shown that it is possible to convert this network of internal links into a meaningful network of knowledge (Schwartz 2021). Based on this methodology, it has been possible to quantitatively analyse the network that connects the cultural contexts of Pablo Picasso, Albert Einstein and James Joyce. This approach has been shown to successfully deploy a multiscale analysis that allows characterising individual nodes (degree and betweenness centrality, participation coefficient, etc.), clusters (size, density, openness, etc.) and the whole networks (modularity, assortativity matrix) these individuals belong to. The same approach has also been applied to study the interactions between Michelangelo, Copernicus, and Pico della Mirandola in the Italian Renaissance (Miccio et al. 2022).

Instead of focusing on specific cases, this work proposes a statistical approach to quantitatively analyse some characteristics of the cultural networks belonging to a given historical period. This aim is achieved by averaging many networks based on individuals from such epoch (each network is obtained from a triad, comprised of an individual from each of the mentioned disciplines, i.e., an artist, a scientist, and a philosopher). By doing this, we aim to establish the average properties of the period, unearth the structure that allows the flow of knowledge across different disciplines, quantify the behaviour of individual nodes, and describe the collective characteristics of clusters and networks. Although this methodology can be applied to any historical period (depending on the temporal window used when selecting the people), in this work we will focus, as a proof of concept, on the interdisciplinary cultural network that connects art, science, and philosophy in Europe in the seventeenth century.

Thus, by combining ideas borrowed from knowledge discovery in databases and complex networks theory, along with the proper contextualisation of the analytical results within the existing historiographical knowledge, this approach helps unveil the emergence of collective knowledge in and about the period. In addition, it also finds subtle connections between *cultural entities* (as defined in the next section) present on Wikipedia that would otherwise be difficult to detect. The presented results agree with very well-established knowledge and also provide new insights into the structure of cultural networks, revealing unknown characteristics of the studied historical period.

Methods

Structural relationship. Most of the research works that use Wikipedia's internal links network are based on the direct connection between articles (Eom et al. 2015; Eom and Shepelyansky 2013; Gabella 2019). Instead, here we use the concept of *structural relationship* between nodes. This means that two Wikipedia

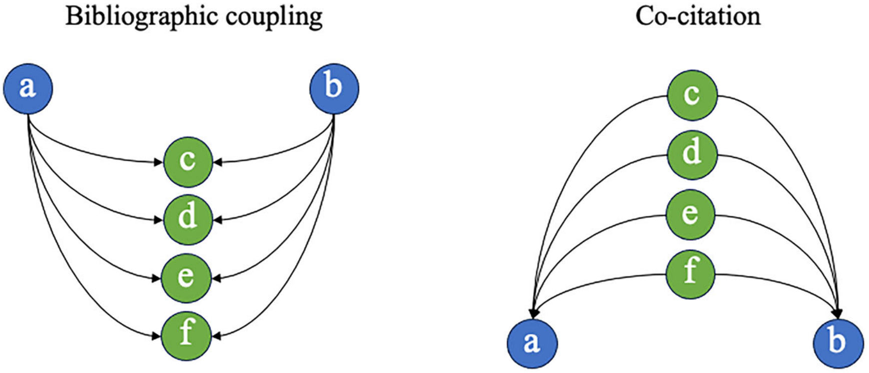


Fig. 1 Structural relationship between elements in a complex network. The distance (or relatedness) between two structurally related nodes can be measured using a metric inspired by the normalised Google distance (see text).

articles can be strongly related even if one does not link to the other. We say that two elements are structurally related if they are related to common elements or if there are common elements that are related to both (see Fig. 1). Furthermore, the strength of this structural relationship can be quantified using a proper metric. In this work, we use a metric inspired by the *normalised Google distance* (NGD) (Witten and Milne 2008), that is commonly used, for instance, in natural language processing to compute similarity between words or phrases (Cilibrasi and Vitányi 2007). The NGD is defined as:

$$d_{in/out}(a,b) = \frac{\log(\max(|A|,|B|)) - \log(|A \cap B|)}{\log(|W|) - \log(\min(|A|,|B|))}$$

where a and b indicate the two articles of interest, A and B represent the sets of nodes (Wikipedia articles) that link to/from ($d_{in/out}$) a and b, respectively, and W refers to the total number of nodes in the network. If $|A \cap B| = 0$, then the corresponding distance is infinite. We considered two distances between nodes a and b: one for nodes that *link to a* and b ($d_{in}(a,b)$) and another for nodes that are *linked from a* and b ($d_{out}(a,b)$). The final distance ($d(a,b)$) was taken as the harmonic mean between the in/out distances. Finally, the relatedness between nodes a and b is defined as $r(a,b) = \exp(-d(a,b))$, which is always in the range [0,1]. (See Supplementary Information, Section2, for more details).

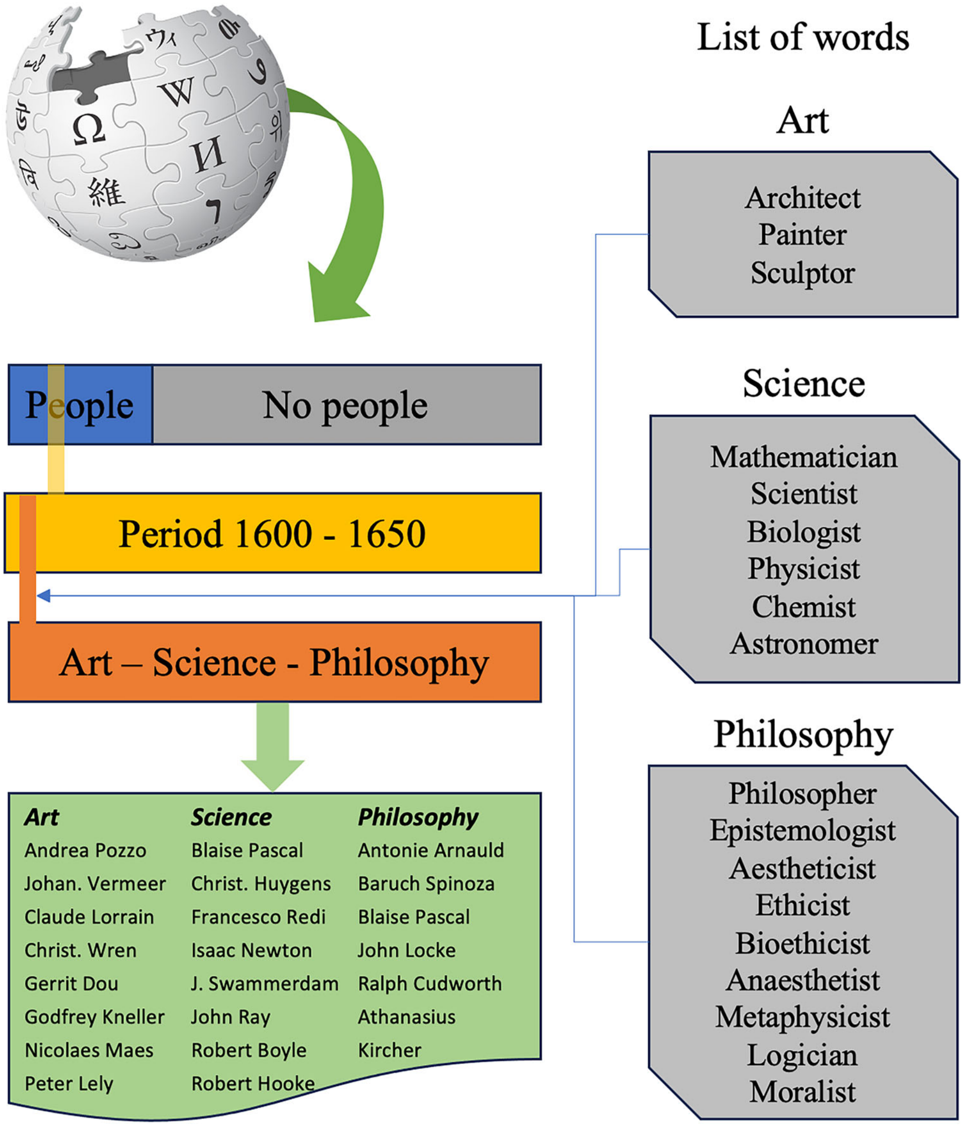
Using the NGD provides a much more stable approach than using direct links because at least two connections are necessary to change the distance between two articles. In addition, the fact that we only use the links between articles and not their content, reduces the impact of errors and biases present in Wikipedia as discussed later (see the end of the Results and Discussion section). Using internal links makes us miss the details but allows us to enhance the vision of the big picture.

Cultural entities. The fundamental idea of the proposed methodology is to use the links between Wikipedia's articles (a directed network) as an intermediate step to build meaningful, non-directed networks, where the NGD establishes the relationship between different elements. Each of these elements represents what we call hereafter "cultural entities" in the context of this work. This means that a given entry in Wikipedia, for example, "Isaac Newton", represents for us not the historical, "real" Isaac Newton but an abstract knowledge entity consisting of many layers, including but not necessarily limited to its social, scientific and political contexts, other cultural entities that influenced or have been influenced by it, etc. Thus, a cultural entity results from both the explicit knowledge generated by historians and other humanists/experts to be used and deployed in a specific context (i.e., education, historiography, cultural management, cultural transmission, etc.), and the implicit knowledge generated and discovered by the transformations of, in this case, the internal links in Wikipedia entries into a non-directed network that is analysed quantitatively. In the context of this work, cultural entities only emerge or happen within our non-directed weighted network and are depicted as labels such as "Isaac Newton" in lieu of nodes, links, or clusters. They help us refine the cultural lenses we use to depict and understand the historical periods we are studying by surfacing the structural relations of the network on which such depictions and understanding stand. It is worth noting that, in this case, we are studying a cultural network *as represented* on Wikipedia. Although this implies a partial and biased view of culture, the agreement between the results obtained using this methodology and the well-established cultural knowledge (as discussed later) makes this approach suitable for this kind of study.

People. As mentioned above, in this work we propose to average many networks based on triads of individuals from three areas of knowledge creation (Art, Philosophy, and Science). For this purpose, we need to extract a representative sample of the artists, scientists, and philosophers from Wikipedia for the studied period and generate all possible triads based on this data. We extracted people by parsing a publicly available dump from Wikipedia. Specifically, we used a snapshot of the May 1^{st}, 2022 English version. Thus, by detecting the tag 'Person' in the *infobox* (when available), we retrieved a sample of 501,046 persons. The corresponding Wikipedia articles were analysed using natural language processing techniques to obtain the birthdate, the number of links and their profession(s) (as stated in the first sentence of the corresponding biography's text).

Figure 2 shows the pipeline we used to generate the triads. From Wikipedia's dump, we extracted all the (tagged) people; then, we took a subset composed of all the people born in the period 1600–1650. From this subset, we selected artists, scientists, and philosophers based on whether (at least) one term in our predefined lists of words (see Fig. 2) appears as a profession in the corresponding biography. Although arbitrary, using lists of words allows for a systematic and well-defined classification of people into different disciplines and can be easily tuned according to the research's purpose.

Networks. We generate a network for each triad based on the relatedness among nodes as described in previous works (Miccio et al. 2022; Schwartz 2021). Briefly, for each one of the selected people (hereafter called 'seed'), we downloaded from the online English version of Wikipedia (January 29th, 2023) the



Art	*Science*	*Philosophy*
Andrea Pozzo	Blaise Pascal	Antonie Arnauld
Johan. Vermeer	Christ. Huygens	Baruch Spinoza
Claude Lorrain	Francesco Redi	Blaise Pascal
Christ. Wren	Isaac Newton	John Locke
Gerrit Dou	J. Swammerdam	Ralph Cudworth
Godfrey Kneller	John Ray	Athanasius
Nicolaes Maes	Robert Boyle	Kircher
Peter Lely	Robert Hooke	

Fig. 2 A schematic representation of the pipeline used to extract from Wikipedia all the artists, scientists, and philosophers for the studied historical period. First, we extract all the people from Wikipedia; then we take those belonging to the desired period (1600–1650); and finally, we classify people as artists, scientists, or philosophers depending on whether at least one word in the corresponding list of words appears in their biography.

corresponding content pages (nodes), the internal links (only those in the main text of an article that connect with other articles on Wikipedia) and their outgoing neighbours (with the corresponding internal links) within distance two from the seed. Thus, for each triad, we built a graph composed of all the nodes (Wikipedia articles) and internal links extracted for each person in the triad. As an example, for the triad Johannes Vermeer, Christiaan Huygens and Blaise Pascal, the corresponding graph contains 38487 nodes and 1471559 edges. The next step was to fix the redirects (pages that automatically send visitors to another page and do not have 'real content') by redirecting the source page to the corresponding target (the output of the redirect). We also excluded Wikipedia pages with no relevant information for our research purpose (i.e., redirect, disambiguation, category, and list pages). For the obtained graph, we extract the N_i (the out-degree of the seed i) most related nodes to each of the given seeds, as measured using the NGD metric. Finally, we converted the so obtained directed graph of internal links on Wikipedia into an undirected graph, where the NGD gives the distance (d) between elements and $r = e^{-d}$ ($r \in [0, 1]$) the relatedness.

The final graph, shown in Fig. 3, has 1017 nodes and 126073 edges. Figure 3 was built using Gephi (Bastian et al. 2009), an open-source software for exploring and manipulating networks, and the Fruchterman-Reingold method (Fruchterman and Reingold 1991), a force-directed layout algorithm. The nodes were assigned to the different clusters (art, science, philosophy) according to the seed they were linked to in the directed Wikipedia graph. If a node was connected to many seeds, then it was assigned to the seed it was most related to.

In Fig. 3 we can see a relatively strong interaction between science and philosophy (Huygens- and Pascal-related clusters). The Vermeer-related cluster ('art' cluster) is more apart and less connected to the other two clusters. The only inputs of the algorithm are the three people of the triad; from there, only the internal links on Wikipedia and the selected metric (NGD) are used to generate the network, based (in this case) on Vermeer, Huygens, and Pascal. Considering this, it is noticeable how well the Fruchterman-Reingold algorithm localises the different elements (see Fig. 3). Thus, Isaac Newton is in the middle of the scientific cluster, Blaise Pascal is halfway between science and

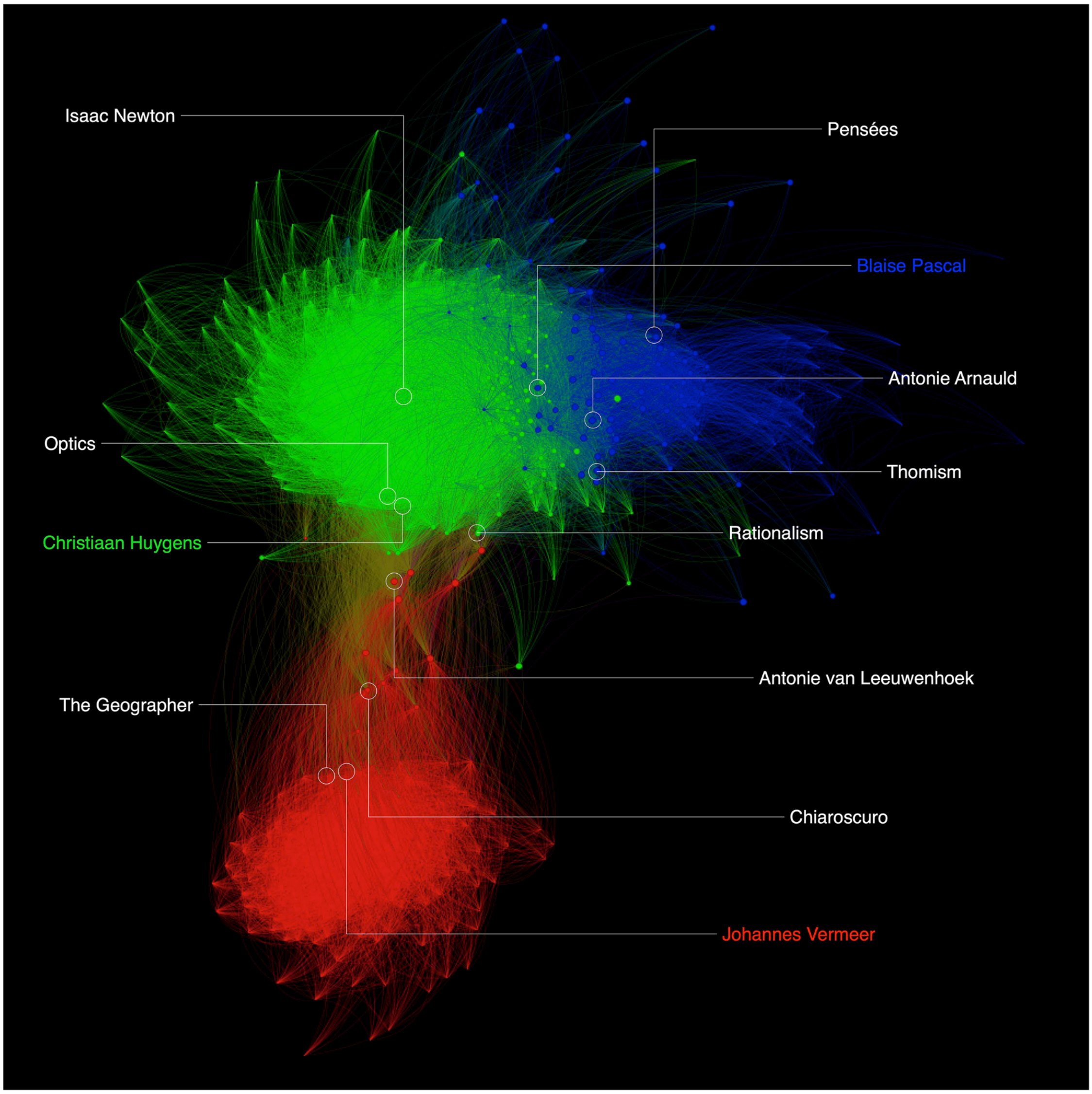


Fig. 3 Cultural network based on the seeds of Johannes Vermeer, Christiaan Huygens and Blaise Pascal. Each point represents a Wikipedia entry, and each line represents the relationship between those items. Note the strong interaction between science (green) and philosophy (blue) and the relative isolation of art (red).

philosophy, and *Chiaroscuro* is midway between *optics* and Vermeer.

Data sampling. To characterise a given historical period, we repeated the procedure of generating the network for many triads and then averaged the obtained values for the different coefficients. The number of triads to be analysed is an important factor to ensure a reliable mapping of the underlying cultural network of the studied period. This number is related to the sampled period's heterogeneity and the desired precision of the results according to the purpose of the research. The error of the statistical variables we used here goes as $\sigma/\sqrt{N_T}$, being σ the standard deviation and N_T the number of sampled triads. Thus, given the standard deviation (characteristic of the sample) and the required error (which depends on the research question), we can determine the necessary number of triads, as discussed in more detail in the next section. In the case of this work the number of triads was the highest possible (465) given the number of seeds we have.

Finally, only some of the selected people for the studied period were used to generate the triads for the networks based on the number of outgoing links (outdegree) to other articles. The threshold for the outgoing links has been set based on three criteria: (1) In this work, we focus on the most relevant people from a given period; although there is no way to determine such relevant people objectively, the outdegree can be taken as an indicator of relevance. (2) High values of outdegree also increase the probability of having high-quality Wikipedia entries. Typically, articles with more outgoing links were edited more times (and by more editors), increasing their quality (Ruprechter et al. 2020). (3) Finally, since our method is based on sampling, and given that the generated network's size depends on the seeds' outdegree, we need a minimum number of outgoing links for each selected seed on Wikipedia to ensure statistical representativity. We established an empirical safe threshold of 100 for the outdegree based on previous experience studying many networks of different historical periods (Miccio et al. 2022; Schwartz 2021). Figure SI2 in the Supplementary Information file illustrate and explain in detail each step of the proposed methodology.

Data analysis. Beyond the qualitative agreement shown in Fig. 3, we can use different coefficients, properties and tools from complex networks theory to quantitatively analyse the graphs. Thus, we can calculate global, cluster and node properties for each graph (made from each triad).

Modularity. Modularity provides a quantitative measure of the clusterisation of the network and, for a weighted network, is defined as (Newman 2003)

$$Q = \frac{1}{W}\sum_{C}\left(W_c - \frac{S_c^2}{4W}\right)$$

where W is the weight of all links in the network, W_c is the weight of all the internal links of cluster c, S_c is the weight (internal and external) of all nodes in c, and the sum runs over the three clusters in the network. The maximum possible Q value for each network has a non-trivial value given by $Q_{Max} = 1 - \sum_C (S_c^2/4W^2)$. Therefore, to compare the modularity across several networks, we used the normalised value given by $Q_N = Q/Q_{Max}$. This coefficient helps to identify how well a network is divided into clusters, with higher values indicating stronger, more distinct groupings of nodes.

Assortativity. Another relevant global/cluster property is the *assortativity matrix* A, where each element a_{ij} is defined as the sum of the weighted links connecting nodes from cluster i and j. As done for modularity, we can normalise this matrix, $A_N = A/||A||$, where $||A||$ represents the sum of all the elements in A. In this way, each element from A_N (a_{ij}^N) gives the fraction of weighted links connecting nodes from clusters i and j.

Openness. We can also define *openness* as the percentage of nodes (for each cluster) having an external strength larger than their internal strength. Openness (Op) is thus a measure of how many nodes are more connected with nodes in other clusters than in their clusters. This value varies between 0 and 100%; as expected, the lower the openness, the less interactive the corresponding cluster.

Nodes connectivity. At the level of individual nodes, we focus on analysing how they connect with other nodes, particularly the intra- and inter-cluster connectivity. Although the obtained network is an undirected weighted graph, we can define in- and out-strengths for each node as the sum of the respective weighted links connecting to other nodes in the same cluster (s_{int}) or any other cluster (s_{out}).

Bootstrapping. We used bootstrapping (with $N = 1000$) to estimate the different statistical coefficients (mean, variance, etc.) (Gel et al. 2017; Thompson et al. 2016). This methodology is commonly used as a valuable method for sampling, especially in those cases where the entire sample is virtually unavailable or inaccessible. (See Supplementary Information, Section2, for more details).

Results and discussion

The methodology we propose in this work is general and can be applied to a broad range of problems that analyse the interactions among different cultural entities in any field or historical period. In particular, we will focus on studying the relationships among art, science, and philosophy throughout the 17th century to test our approach. The present study is a proof of concept performed over 465 triads of people born between 1600 and 1650. This period is somehow between the historical ages analysed in previous works (Miccio et al. 2022; Schwartz 2021), allowing a comparison to prior results. In this section, we analyse the obtained cultural networks at three different scales: the global network (modularity and assortativity), the clusters and their interaction (assortativity matrix, openness, links' distributions), and individual nodes (strength distribution, relative frequency).

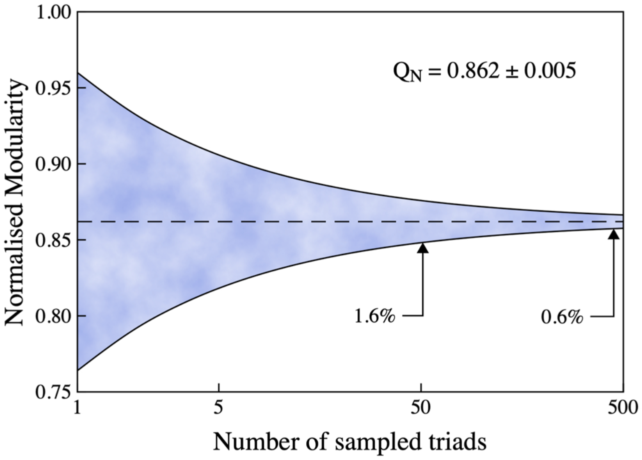


Fig. 4 Average normalised modularity (Q_N) and the corresponding error interval as a function of the number of triads. The relative error decreases upon increasing the number of triads, being 1.6% for 50 samples and 0.6% for 465 (the whole dataset). The dashed line indicates the mean value obtained with the bootstrapping method. Note the logarithmic scale on the *x-axis*.

Modularity of the period. The average normalised modularity $\left(\overline{Q_N}\right)$ allows us to characterise the period regarding the degree of clusterisation of the different disciplines. Figure 4 shows the calculated $\overline{Q_N}$ and the corresponding error interval as a function of the number of sampled triads (N_T). The estimated error for $\overline{Q_N}$ is about 1.6% if we take only 50 triads and diminishes below 1% for the entire dataset (465 triads). In addition, the modularity's standard deviation (σ) provides implicit information about the heterogeneity of the relationships between artists, scientists and philosophers during the studied period. The lower the value, the more homogeneous the corresponding interactions are. In this case, the calculated value of the average normalised modularity is 0.862 ± 0.005 and the standard deviation is 0.098, representing 11.4% of the total $\overline{Q_N}$.

By using the results from previous works, we observe that the modularity for this period is higher than that obtained for the interactions between art, science, and philosophy in the Italian Renaissance ($Q_N = 0.77$), studied through the particular lens of the relationships between Michelangelo, Copernicus, and Pico della Mirandola (Miccio et al. 2022). In contrast, when the relationships among art, science, and literature were analysed using Picasso, Einstein, and Joyce as seeds (Schwartz 2021), a value of $Q_N = 0.88$ was obtained. Although these are particular cases and not the average of the corresponding historical periods, they may suggest a progressive increment of the modularity throughout the centuries, which agrees with the well-known hypothesis of growth of the disciplinary specialisation of knowledge production over time (Yndurain 2016).

Intra- and interdisciplinary interactions. As mentioned above, the coefficients of the normalised assortativity matrix (a_{ij}^N) indicate the fraction of weighted links between clusters i and j and quantitatively determine the intra- and inter-cluster interactions. Coefficients a_{ii}^N ($i = 1, 2, 3$) give the strength of the intra-cluster

Table 1 Assortativity coefficients with their corresponding error as determined using the bootstrapping method.

Assortativity coefficients	Art	Science	Philosophy
Art	$\overline{a_{11}^N} = 0.285 \pm 0.004$ $\sigma_{11} = 0.085\,(30\%)$	$\overline{a_{12}^N} = 0.025 \pm 0.002$ $\sigma_{12} = 0.035\,(140\%)$	$\overline{a_{13}^N} = 0.017 \pm 0.001$ $\sigma_{13} = 0.020\,(117\%)$
Science		$\overline{a_{22}^N} = 0.268 \pm 0.003$ $\sigma_{22} = 0.074\,(28\%)$	$\overline{a_{23}^N} = 0.092 \pm 0.003$ $\sigma_{23} = 0.065\,(71\%)$
Philosophy			$\overline{a_{33}^N} = 0.313 \pm 0.005$ $\sigma_{33} = 0.115\,(37\%)$

The standard deviation for each distribution is also shown. The number in parenthesis indicates the relative weight of the standard deviation compared to the corresponding mean value (see text).

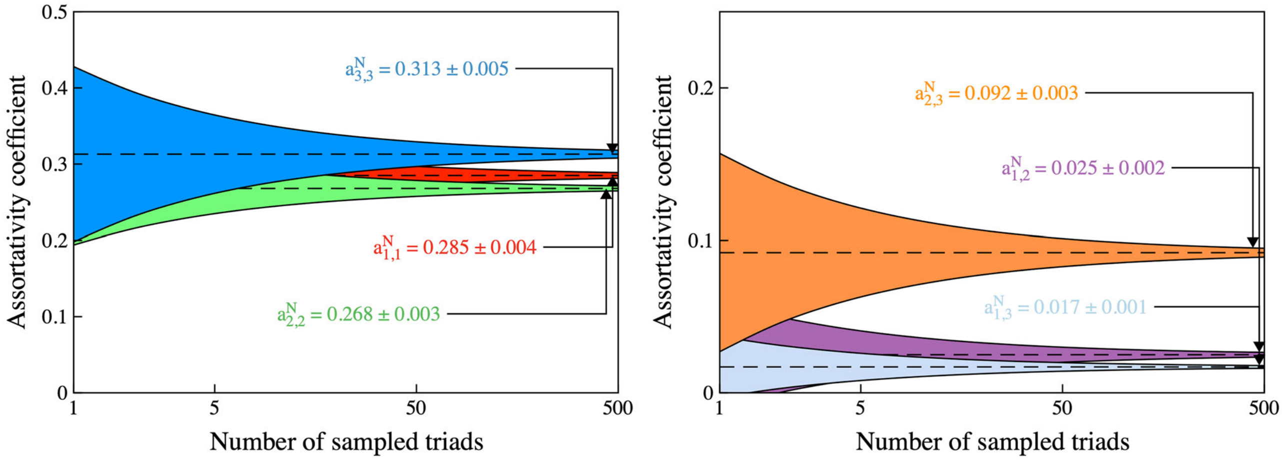


Fig. 5 (Left) Average values of the normalised assortativity coefficients in the diagonal (a_{ii}^N) as a function of the number of triads. These values represent the strength of the internal interactions within each discipline: art (red), science (green) and philosophy (blue). (Right) Average values of the out-of-diagonal ($a_{ij}^N\,(i \neq j)$) normalised assortativity coefficients as a function of the number of triads. These values represent the strength of the interdisciplinary interactions: art/science (violet), art/philosophy (light blue) and science/philosophy (orange).

relationships (see Table 1). Figure 5 (left) shows these coefficients' mean values and corresponding error intervals upon increasing the number of sampled triads (N_T).

The plot shows a behaviour similar to that observed for modularity (see Fig. 4). It is interesting to note that for 50 triads, the possible values for the three coefficients overlap. However, by increasing the number of triads up to 465, we observe that the error for each coefficient is smaller than the difference between their means. This allows us to precisely establish their values and (in this case) determine a difference across them. We obtained values of $\overline{a_{11}^N} = 0.285 \pm 0.004$ for the art cluster, $\overline{a_{22}^N} = 0.268 \pm 0.003$ for science and $\overline{a_{33}^N} = 0.313 \pm 0.005$ for philosophy (see Table 1). This means that whereas art and science present similar strengths of the interactions among the corresponding cultural entities within each cluster, philosophy shows a higher strength of 'connectivity'. Comparing these results with the previously studied particular cases, we observe that both in the Italian Renaissance ($a_{11}^N = 0.383$) and at the beginning of the 20th century ($a_{11}^N = 0.316$), at least for the studied cases, art was the most connected discipline (instead of philosophy) within its cluster.

The out-of-diagonal elements of the normalised assortativity matrix $a_{ij}^N(i \neq j)$ represent the interactions between different disciplines. Figure 5 (right) shows (in violet) the corresponding interactions between art and science for the studied period. We obtained a value of $\overline{a_{12}^N} = 0.025 \pm 0.002$ which is substantially smaller than the intra-cluster counterparts for both art (0.285) and science (0.268). Comparing again with previous studies for particular cases, we observe similar values for the art/science interactions at the beginning of the 20th century (Schwartz 2021) (0.020) and a much higher level of interaction during the Italian Renaissance (Miccio et al. 2022) (0.056), which agrees with the idea of a stronger relationship between art and science in the latter period as stated in previous works (Bernal 1954; Kemp 1990; Wade 2017). Similar results are observed for the art/philosophy interactions (light blue) with an average value of $\overline{a_{13}^N} = 0.017 \pm 0.001$, whereas the corresponding value for the 15th century grows up to 0.130, almost eight times bigger.

Different is the case of the interactions between science and philosophy. Figure 5 (right) shows (in orange) the corresponding normalised assortativity coefficient upon increasing the number of sampled triads. On the one hand, the mean value is much higher than for the others out of diagonal coefficients ($\overline{a_{23}^N} = 0.092 \pm 0.003$) and also higher than the same coefficient at the Italian Renaissance (Miccio et al. 2022) (0.058). These results seem to corroborate Sandoz's argument (Sandoz 2021) on the history of scientific disciplines as it relates to philosophy's steady shifting at the beginning of the modern period from an interest in art and aesthetics to inquiries around epistemology, science and scientific knowledge. For Sandoz, the Baconian division of knowledge was very influential in England at the time, and to prove his point, he quotes Thomas Hobbes who, in his *Leviatán* (1651), states the following: "There are of knowledge two kinds, whereof one is knowledge of fact; the other, knowledge of the consequence of one affirmation to another. The former

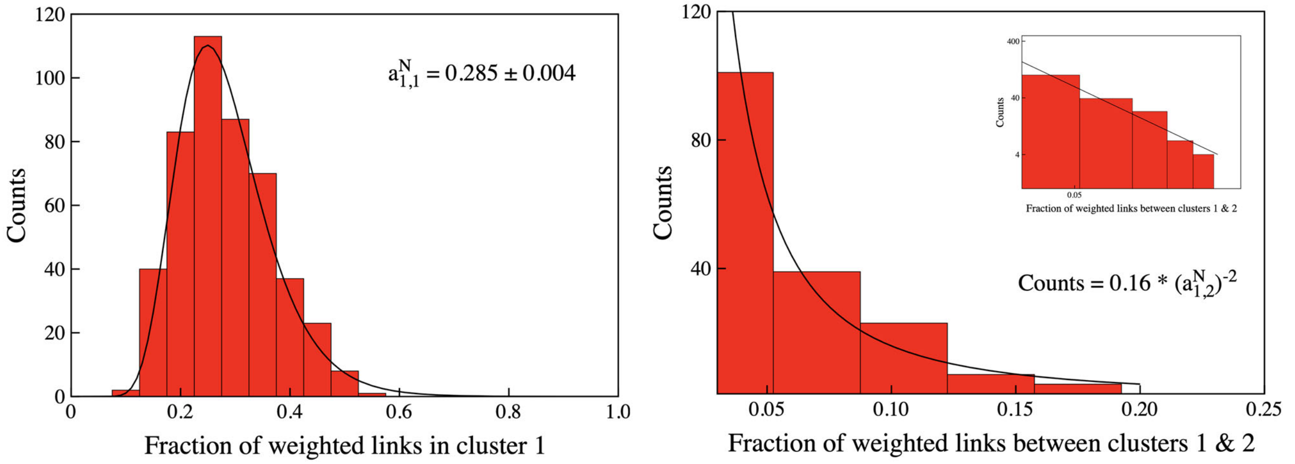


Fig. 6 Distribution of the values of the normalised assortativity coefficients for in diagonal element (a^N_{11}) (left) and out of diagonal (a^N_{12}) element (right). Solid lines represent the best fit to the histograms. A lognormal function for the a^N_{11} element and a power law for a^N_{12}. The inset shows a double log plot where the power law distribution is evident.

is nothing else but sense and memory, […] and this is the knowledge required in a witness. The latter is called science, […] and this is the knowledge required in a philosopher" (Sandoz 2021). The normalised assortativity matrix seems to show that this new knowledge required by the 17$^{\text{th}}$-century philosopher was being achieved at the expense of the efforts previously devoted to the matters of sensibility, art and direct human experience of the world.

Figures 4, 5 shows that the number of triads used in this study is enough to decouple the coefficients of the assortativity matrix. The corresponding errors are well below the difference between their mean values, giving a precise estimation of the coefficients. Table 1 shows the values for the coefficients of the assortativity matrix with the corresponding errors for 465 triads. The table also shows the standard deviation (σ) for each distribution, whereas the number in parentheses indicates its relative relevance regarding the mean value $(\frac{\overline{\sigma_{i,j}}}{\overline{a^N_{i,j}}}100)$. This last parameter indicates the dispersion for a given coefficient. Thus, we can observe that the elements in the diagonal, which correspond to the fraction of weighted links within each discipline, show similar dispersion (around 30%). However, this value substantially increases for the out-of-diagonal elements, going from 71% ($\overline{\sigma_{23}}$) to 140% ($\overline{\sigma_{12}}$). This means that the relationships between science and philosophy are much more homogeneous for the studied period than those between art and science. The interactions between art and philosophy are also rather heterogeneous (117%). Thus, the connections between science and philosophy are not only the strongest ones but also the most homogeneous among disciplines. This would help to quantitatively confirm the well-established historical facts about, first, the overlapping and, second, the slow separation of philosophy and science in this time around topics of the nature of the mind/matter relations, epistemology, and the detachment of the scientific method from its philosophical matrix, and the mechanisms that run and explain the universe (Blair 2007; Burke 2017; Freedman 1994). The establishment of separate art channels of production and distribution for Protestant and Catholic markets in Europe and Latin America (Suárez et al. 2012), on the one hand, and the role that local and national patrons played in the creation of the respective art fields, on the other, would also explain the heterogeneity of the connections that the art cluster show in this analysis with both philosophy and science (Tucker 2010).

Figure 6 (left) shows a histogram with the distribution of the values for the a^N_{11} coefficient for the 465 triads considered in this work. The solid line represents the best fit to the data using a lognormal distribution function. As shown in the Figure, the expected value for the mean of this distribution function is the same as that calculated using bootstrapping. Similar results were obtained for the rest of the coefficients in the diagonal. However, a different behaviour is observed for the distribution of the out-of-diagonal elements of the assortativity matrix. Figure 6 (right) shows the distribution of the values for the a^N_{12} coefficient for all the sampled triads. In this case, the distribution follows a power law with an exponent $\alpha = -2$. Thus, weighted links' intra- and inter-cluster distributions follow two different trends. We will further discuss this point in relation to the behaviour of individual nodes.

As previously defined, openness gives a relative measure of how connected the nodes in a given cluster are with nodes in *other* clusters. Table 2 shows the obtained values for some triads where we kept constant two seeds and varied the third one. In the case of Table 2 (first 8 lines) we fixed Isaac Newton and Baruch Spinoza, whereas the art's seed was varied. We can observe from the obtained openness that it keeps relatively constant for Newton- and Spinoza-related clusters: close to zero in the former case and close to 90% in the latter, regardless of the art's seed. This means that the Newton-related clusters are very isolated (the corresponding elements do not strongly interact with elements in other clusters). On the contrary, the nodes in the Spinoza-related clusters are strongly connected to nodes in different clusters; since its openness is relatively independent of the art's seed, we can assume that most of the interactions are with elements in the Newton-related clusters. This agrees with the high value observed in these particular networks (all triads in Table 2 (first 8)) for the corresponding assortativity coefficient $a^N_{23} > 0.203$ (compared to the period's average $\overline{a^N_{23}} = 0.092$).

The inbreeding of the Newton-related cluster is consistent with the evolution of Newtonianism in Europe. Due to the technical nature of the *Principia*, a book published in 1687, it took time to appreciate the relevance of the British scientist's ideas even in his homeland. As Paul Hazard has pointed out, the eighteenth century drew on Newton's ideas from the previous century's end and elevated them to paradigmatic status (Hazard 1961) (p.293). Only then, thanks to the contribution of a legion

Table 2 Openness values for each cluster for representative triads.

Triads' seeds			Openness (*Op*)		
Art	Science	Philosophy	Art	Science	Phil.
Nicolaes Maes	Isaac Newton	Baruch Spinoza	1.79	0	88.98
Claude Lorrain	Isaac Newton	Baruch Spinoza	3.81	0	88.70
Johannes Vermeer	Isaac Newton	Baruch Spinoza	3.05	0	90.91
Peter Lely	Isaac Newton	Baruch Spinoza	26.67	0.23	84.44
Christopher Wren	Isaac Newton	Baruch Spinoza	31.96	0.49	77.33
Gerrit Dou	Isaac Newton	Baruch Spinoza	3.75	0	90.91
Andrea Pozzo	Isaac Newton	Baruch Spinoza	4.08	0	87.50
Godfrey Kneller	Isaac Newton	Baruch Spinoza	37.25	0.42	88.71
Johannes Vermeer	Christiaan Huygens	Blaise Pascal	3.70	0	43.97
Johannes Vermeer	Isaac Newton	Blaise Pascal	2.50	0	54.55
Johannes Vermeer	Johannes Hevelius	Blaise Pascal	0.63	10.61	0
Johannes Vermeer	Jan Swammerdam	Blaise Pascal	1.22	26.19	0

(First 8) The seeds for science and philosophy were kept constant, whereas different artists were considered. We can observe that the values for the clusters related to Newton and Spinoza remain relatively constant, independent of the art's seed. (Last 4) In this case, we fixed the seeds for art and philosophy and varied that for science. We observe little changes when replacing Huygens with Newton but drastic changes when using Hevelius.

Table 3 Average openness for each cluster as calculated using bootstrapping.

Art	Science	Philosophy
$\overline{Op} = 5.0 \pm 0.4$ $\sigma = 8.1\,(162\%)$	$\overline{Op} = 15.5 \pm 1.0$ $\sigma = 22.1\,(143\%)$	$\overline{Op} = 13.4 \pm 1.1$ $\sigma = 23.8\,(178\%)$

We can observe a strong variability (high values of σ) since openness is very sensitive to the seeds in the triad.

of popularisers (Bentley already in 1692, Fontenelle, Voltaire, etc.), the ideas of the English scientist would be widely distributed and would then penetrate all domains, as Morris Kline (Kline 1972) (p.319) has shown, from philosophy to theology, from art to poetry, even multiplying significantly the poems dedicated to science and Newton. In the continent, still deeply imbued with Cartesianism, resistance to Newtonism survived until the 1740 s. It is not surprising, therefore, that the influence of Newton's ideas took longer to be felt by Europe-based artists than by England-based ones. The values in Table 2 (first 8 lines) seem to confirm this, as the highest values by far are given by Peter Lely (a Dutch artist who settled in London), Godfrey Kneller (a German artist who settled at the English court) and Christopher Wren (an English architect and scientist). In continental Europe, the rest of the artists (from Holland, France, Italy) show much lower values.

Table 2 (last 4) shows a different scenario. In this case, we have fixed Johannes Vermeer and Blaise Pascal and varied the scientists. We can observe that replacing Newton with Huygens slightly changes the openness for the philosophy cluster (keeping the other two openness values almost unchanged). However, changing Newton by Hevelius (or Swammerdam) drastically changes the openness in both science and philosophy clusters. These results have two implications: on the one hand, this means that the nodes in the clusters related to Pascal are mainly (and strongly) connected to nodes in both Huygens- and Newton-related clusters (Jorink and Maas 2012; Leer and Boers 2022; Snelders 1989). On the other hand, Vermeer- and Pascal-related clusters weakly interact with each other ($Op = 0$), whereas Hevelius- and Swammerdam-related clusters are more prone to interact with nodes other than theirs.

Finally, we focus on the average value of the openness for the whole dataset (465 triads) and the corresponding distributions.

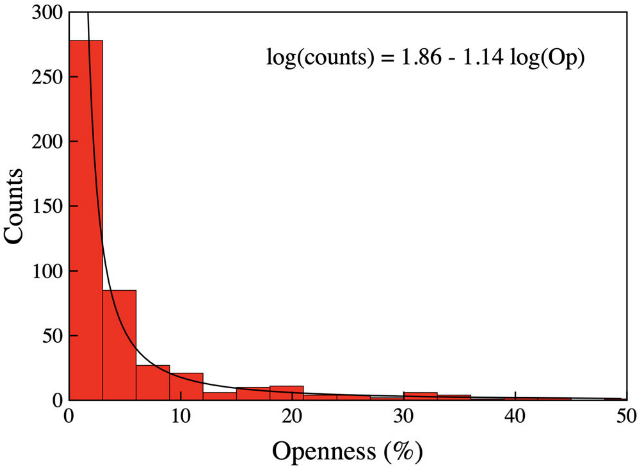


Fig. 7 Distribution of the openness values for the art cluster. The strong tendency towards low values is evident from the plot. The label shows the parameters for the best fit using a power law function.

As with the previous coefficients, we used bootstrapping to estimate the average, the standard deviation, and the openness error. Table 3 shows the average value for each cluster. These results show that art, in this period, is the most 'endogamic' cluster, whereas science and philosophy have an almost three times higher openness.

Concerning the distribution of the openness, Fig. 7 shows a histogram of the corresponding values for the art cluster. We can observe an evident trend towards low values, which is typical for all clusters due to the homophilic behaviour of their components. This means that nodes of a given cluster tend to be more strongly connected to other nodes in the same cluster than to nodes in different clusters. The distribution of the openness values follows a power law with exponent α = –1.14. A similar tendency was observed for the science cluster (α = –0.95) and the philosophy cluster (α = –1.09). It is easy to see that the lower the exponent, the higher the average openness, as the values are more uniformly distributed.

Individual nodes. At the level of individual nodes, we can use the previously defined coefficients (s_{int}, s_{out} and the corresponding distributions) to get insight into the relevance and role of the distinct cultural entities emerging in the studied period. For this period, we have 3585 unique nodes in the 465 networks, and we should note that individual nodes not only refer to people but also to objects (*microscope, Galilean moons, books, paintings*) or ideas (*rationalism, energy, stoicism*). These nodes (cultural entities associated with objects, ideas, or people) can be connected with others in the same or different clusters. The relative strength of these connections (intra- or inter-cluster) will determine the network structure and, consequently, the characteristics of the corresponding historical period.

Table 4 People and concepts with the highest relative frequency in the different triads.

Node's name	Birth - Death	Discipline	*N*
Pierre Gassendi	1592–1655	Sci/Phil	355
Robert Boyle	1627–1691	Science	351
Mechanical Philosophy			338
René Descartes	1596–1650	Sci/Phil	326
Pierre Bayle	1647–1706	Philosophy	313
Nicolas Malebranche	1638–1715	Philosophy	313
Evangelista Torricelli	1608–1647	Science	309
Rationalism			307
Giambattista Vico	1668–1744	Philosophy	299
Thomas Hobbes	1588–1679	Philosophy	296
Michel de Montaigne	1533–1592	Philosophy	296
Scientific revolution			294
Blaise Pascal	1623–1662	Sci/Phil	290
Marin Mersenne	1588–1648	Science	276
Thomism			275
Baruch Spinoza	1632–1677	Philosophy	274
Gottfried Wilhelm Leibniz	1646–1716	Sci/Phil	263
Antoine Arnauld	1612–1694	Philosophy	263
Jean le Rond d'Alembert	1717–1783	Sci/Phil	262
Christiaan Huygens	1629–1695	Science	261
Frans van Schooten	1615–1660	Science	256
Gerolamo Cardano	1501–1576	Sci/Phil	255
Francis Bacon	1561–1626	Philosophy	252
Johannes Kepler	1571–1630	Science	251
David Teniers the Younger	1610–1690	Art	249
Willem van de Velde the Younger	1633–1707	Art	241
Juan Caramuel y Lobkowitz	1606–1682	Philosophy	234

For people, birth and death dates are shown, as well as the corresponding discipline. The last column shows in how many triads (of 465) these nodes appear.

Maybe the most elementary analysis we can do is to establish the relative frequency of each node in the triads. Table 4 shows those nodes that appear in more than half of the triads. Interestingly, the most frequent nodes correspond to cultural entities related to people (Leibniz, Huygens, Kepler) or ideas (mechanical philosophy, rationalism, scientific revolution), but not to objects (microscope, telescope, vacuum pump). Moreover, the majority of the most frequent nodes are not seeds of any triad. This means the proposed method unveils the most relevant cultural entities even if they are not explicitly present in the original data set (starting triads), confirming the usability of the method to retrieve implicit knowledge in a well-known knowledge database such as Wikipedia by quantifying the connections across cultural entities as defined above.

From Table 4 it is clear that science and philosophy have an outstanding presence among the most frequent nodes of the studied period. It is not only the people but also the four most relevant concepts or ideas that belong to science and philosophy.

Now, we move to analyse the strength distribution of the internal links (s_{int}) connecting to other nodes in the same cluster. Following the same reasoning as for modularity and assortativity coefficients, we can average the individual coefficients by considering, for each node, only the triads in which it appears. Figure 8 (left) shows a histogram with the distribution of the average internal strength for the 3585 nodes in the 465 sampled triads. The distribution closely follows a sum of two Gaussian functions (solid line in the Figure). This behaviour is compatible with a core-periphery distribution composed of a dense, highly connected group of nodes, called the *core*, and a sparse and outlying group of nodes, called the periphery (Barucca et al. 2016; Gurugubelli and Chepuri 2022). Similar distributions were found for individual clusters. This finding reveals a substructure of the clusters that deserves a more in-depth analysis and will be further studied in an upcoming article.

Finally, we move to analyse how individual nodes are connected with nodes in other clusters (s_{out}). Figure 8 (right) shows the histogram for the distribution of the average external strength for all the nodes in all the triads (3585 nodes). The distribution follows a power law with an exponent α = −0.74. A similar behaviour is observed when we plot this distribution for

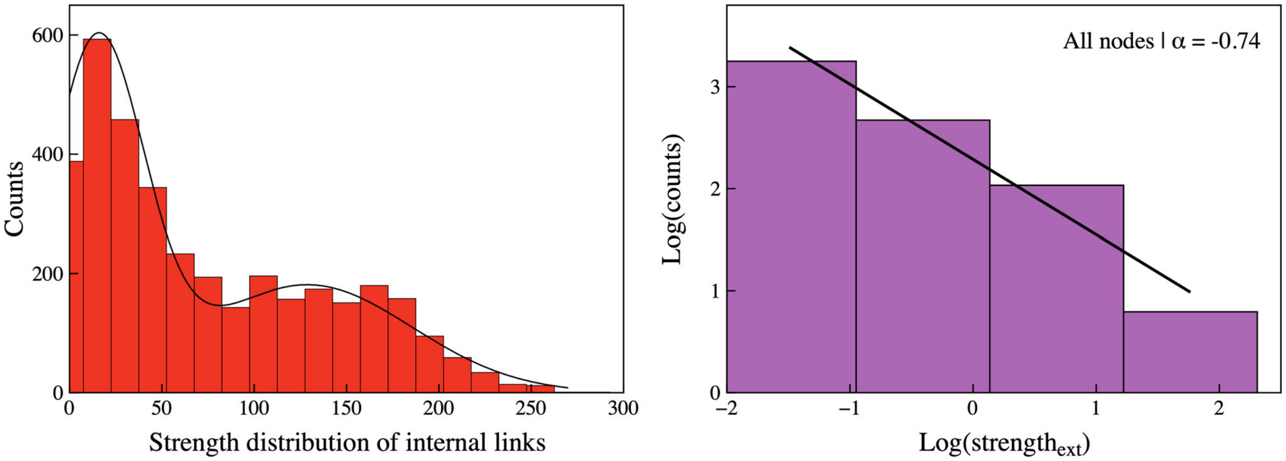


Fig. 8 (Left) Histograms for the distribution of the average internal strength for all the nodes. The solid line is a guide for the eyes and represents the sum of two Gaussian functions. (Right) Histograms for the distribution of the average external strength for all the nodes. Note the double logarithmic scale and the logarithmic binning. Solid line represents the best fit to a power law function.

each discipline. As expected, most nodes in a given discipline are weakly connected to nodes in other fields, and only a few nodes are strongly related to elements in different disciplines. We call those nodes having strong connections with other clusters *generalists*. On the other hand, those weakly related to different disciplines are called *specialists*. We have, in general, more specialists than generalists; this is something that we knew, but now, we can also quantify this ratio. The scale in Fig. 8 (right) shows that the proportion can be as big as one generalist per 200 specialists. The slope of the line in Fig. 8 (the exponent of the power law) determines the proportion between them: the more negative the exponent, the more specialists per generalist. This finding allows for a comparison of this ratio across different periods to understand the role of specialists and generalists in developing cultural dynamics and the generation of knowledge across disciplines.

The method we presented here is based on a simple premise: how articles are connected on Wikipedia can unveil relevant information about cultural networks by quantifying the connections across the cultural entities these networks are made up of. The main findings relate to the ability to rigorously use Wikipedia as a cultural lens into historical periods while elucidating the knowledge structure that supports those depictions of specific periods that come to the average reader through the texts of Wikipedia articles. In this sense, the correspondence, shown through various examples in this article, between the historical findings that emerged through our network analysis and the existing knowledge created by historians through traditional methods, helps support the validity of our method and the tools used to describe the structure of the historical networks we analysed.

Beyond the qualitative agreement between our results and previous historical knowledge, this methodology provides new quantitative insights into the studied historical period. These insights correspond to results for which there are not yet previous studies to compare with. Here is where our approach also proposes to uses Wikipedia as a cultural lens that shows possible research questions and areas to explore. Thus, we were able to determine the *modularity* of the period, related to the degree of clusterisation among art, science, and philosophy; we also found that philosophy was the most 'active' discipline (in terms of internal links) and we can quantify such activity, as well as the interaction across disciplines. Another remarkable finding is the quantitative determination of the ratio between specialists and generalists, which opens the door to understanding their role in knowledge generation and cultural dynamics. These findings need to be confirmed by traditional methods. This validation will boost the proposed method as a powerful tool for analysing cultural networks.

In addition, we found that the average linking behaviour of individual nodes seems to follow two different laws: on the one hand, the distribution of the links among the nodes inside a given cluster is compatible with the random growth model proposed by Erdős and Rényi (Barabási 2016; Erdos and Rényi 1960); on the other hand, the observed power law distribution of the links with nodes in other clusters, agrees with the preferential attachment model proposed by Barabási and Albert (Barabási 1987; Barabasi and Albert 1999). These findings allow for an understanding of how these cultural networks grow and evolve and how individual behaviour correlates with cluster and global properties. This would make it easier for the researcher to classify and quantify specific cultural entities in networks, depending on whether they belong to the core or the periphery in each cluster. We will explore this venue deeply in a future work.

The work presented in this article is a proof of concept of a more general methodology. This approach can be used to analyse any historical period; this depends entirely on the temporal window defined when selecting the people to be used as seeds. Additionally, a geographical window can be implemented in case one wants to study cultural networks located in specific geographical regions (outside the Western World). In this work, we focused the analysis on the relationships among art, science and philosophy, but any other relationship can be analysed using this methodology. For example, one can study the relationships between religion and politics or between different languages or between American and European films and so for. The possibilities are endless.

Finally, we should note that Wikipedia has many biases (gender, race, cultural, etc.) that could impact the results of any research based on its content. In the case of the present study, we are working with the English version of the Wikipedia and therefore, it represents a partial view of the world history. Additionally, most of Wikipedia's editors are white males, which induces race and gender biases in the articles' content. To mitigate the effect of these biases, we decided to use an NGD-inspired metrics. Although links between articles are not immune to the mentioned biases, they are less affected by them because most of the prejudices and partialities are in the content of the articles. Furthermore, since the NGD is based on many links, this diminishes even more the impact of the biases on the results. In addition, we observed an under-representation of non-Western artists, scientists, and philosophers in the English version of Wikipedia. This is why we wrote in the abstract that "*we study the case of the interdisciplinary cultural network connecting art, science, and philosophy in Europe in the seventeenth century*." To study non-Western figures or more local people, such as non-universal French or Italian artists, we suggest using the version of Wikipedia in the corresponding language. Although we miss a more complete picture, the findings presented in this work are still valid for the well-known figures in the Western world. Lastly, it is important to emphasise that this approach is not limited to using the Wikipedia corpus (see SI, Section 4, for more details). Using natural language processing and other artificial intelligence tools would allow extending the methods presented here to any database, provided it contains enough linked information, and compare the results for the same periods or cultural networks across different datasets.

Conclusions

In this work, we proposed a new approach to quantitatively analyse the structure of cultural networks as represented on Wikipedia. Based on ideas borrowed from knowledge discovery in databases and complex networks theory, this approach is geared towards unveiling of collective, implicit knowledge buried in the network of internal links on Wikipedia. The proposed approach also introduces the concept of *cultural entities*, which are related to the context and the different layers associated with a person, an object, or an idea, and the ways in which explicit and implicit knowledge build upon each other in a corpus of linked information. The results show an excellent agreement with well-established historical knowledge, but, more importantly, they also reveal new possibilities for understanding cultural networks and how cultural entities are connected among them. In particular, it is found that nodes within a given cluster interact following a core-periphery behaviour, whereas inter-cluster interactions follow a power law distribution. These findings provide new insights into the structure of cultural networks, help reveal some of their characteristics inscribed in existing but implicit knowledge, and show the potential of applying quantitative approaches to refine the lenses we use to study culture. Thus, by looking at the historical periods of interest, their protagonists, and also the networks of knowledge generation that have produced the resulting depictions, we can increase our understanding of both the structure and dynamics of cultural networks.

Data availability

The supported data sources are publicly available and their sources are presented in the 'Methods' section.

Received: 6 June 2024; Accepted: 17 March 2025;
Published online: 01 April 2025

References

Arroyo-Machado W, Torres-Salinas D, Herrera-Viedma E, Romero-Frías E (2020) Science through Wikipedia: A novel representation of open knowledge through co-citation networks. PLoS ONE 15(2):e0228713. https://doi.org/10.1371/journal.pone.0228713

Barabási AL (1987) Network science. Philos Trans R Soc A: Math, Phys Eng Sci 371:20120375. https://doi.org/10.1098/rsta.2012.0375

Barabási AL (2016) Network science. Cambridge University Press

Barabasi A (1999) Emergence of scaling in random networks. Science 286(5439):509–512. https://doi.org/10.1126/science.286.5439.509

Barucca P, Tantari D, Lillo F (2016) Centrality metrics and localisation in core-periphery networks. J Stat Mech: Theory Exp 2016(2):023401. https://doi.org/10.1088/1742-5468/2016/02/023401

Bastian M, Heymann S, Jacomy M (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. Proc Int AAAI Conf Web Soc Media 3(1):361–362. https://doi.org/10.1609/icwsm.v3i1.13937

Bernal JD (1954) Science in History I. Cameron Associates, New York

Blair AM (2007) Organisations of knowledge. The Cambridge Companion to Renaissance Philosophy. Cambridge University Press

Braudel F (1994) A history of civilisations. New York: Allen Lane

Brown DM, Soto-Corominas A, Suárez JL (2017) The preliminaries project: Geography, networks, and publication in the Spanish Golden Age. Digit Scholarsh Humanit 32(4):709–732. https://doi.org/10.1093/llc/fqw036

Burke P (1992) New Perspectives on Historical Writing. Cambridge: Polity Press

Burke P (2017) Orders of knowledge in early modern Europe. Asiat Stud - Études Asiat 71(3):993–1002. https://doi.org/10.1515/asia-2017-0019

Christian D (2020) Maps of Time. An Introduction to Big History. Press U of C, editor Berkeley

Cilibrasi RL, Vitányi PMB (2007) The Google similarity distance. IEEE Trans Knowl Data Eng 19:370–383

Eom Y-H, Aragón P, Laniado D, Kaltenbrunner A, Vigna S, Shepelyansky DL (2015) Interactions of cultures and top people of wikipedia from ranking of 24 language editions. PLoS ONE 10(3):e0114825. https://doi.org/10.1371/journal.pone.0114825

Eom Y-H, Shepelyansky DL (2013) Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles. PLoS ONE 8(10):e74554. https://doi.org/10.1371/journal.pone.0074554

Erdös P, Rényi A (1960) On the evolution of random graphs. Math Inst Hungarian Acad Sci (5):17–61

Fischer F, Blakesley J, Jäschke R, Wojcik P (2023) Wikipedia, Wikidata, and World Literature. J Cultural Analytics 8(2):1–137

Freedman JS (1994) Classifications of Philosophy, the Sciences, and the Arts in Sixteenth- and Seventeenth-Century Europe. Mod Sch 72(1):37–65

Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. Softw: Pr Exp 21(11):1129–1164. https://doi.org/10.1002/spe.4380211102

Gabella M (2019) Cultural structures of knowledge from Wikipedia networks of first links. IEEE Trans Netw Sci Eng 6(3):249–252. https://api.semanticscholar.org/CorpusID:3710764

Gao J, Hu J, Mao X, Perc M (2012) Culturomics meets random fractal theory: Insights into long-range correlations of social and natural phenomena over the past two centuries. J R Soc Interface 9(73):1956–1964. https://doi.org/10.1098/rsif.2011.0846

Gel YR, Lyubchich V, Ramirez LLR (2017) Bootstrap quantification of estimation uncertainties in network degree distributions. Sci Rep 7(5807). https://doi.org/10.1038/s41598-017-05885-x

Gurugubelli S, Chepuri SP (2022) Generative models and learning algorithms for core-periphery structured graphs. arXiv preprint arXiv:2210.01489

Hargadon AB, Wadhwani RD (2023) Theorising with Microhistory. Acad Manag Rev 48(4):681–696. https://doi.org/10.5465/amr.2019.0176

Hazard P (1961) La crise de la conscience européenne 1680-1715. Paris: Fayard

Ibrus I, Schich M, Tamm M (2021) Cultural Science Meets Cultural Data Analytics. Cultural Science Journal 13(1). https://doi.org/10.2478/csj-2021-0001

Jorink E, Maas A, editors (2012) Newton and the Netherlands. How Isaac Newton was Fashioned in the Dutch Republic. Amsterdam: Leiden University Press

Kemp M (1990) The Science of Art. Optical Themes in Western Art. Yale University Press

Kline M (1972) Mathematics in Western Culture. New York: Penguin Books

Leer K, Boers H (2022) Huygens and Hofwijck: The Inventive World of Constantijn and Christiaan Huygens. Amsterdam University Press

Latour B (2005) Reassembling the social: An introduction to actor-network-theory. Oxford University Press

Magnússon SG editor (2020) Emotional Experience and Microhistory: A Life Story of a Destitute Pauper Poet in the 19th Century. Routledge

Magnússon SG, Szijártó IM editors (2013) What is Microhistory?, Theory and practice. Group T&F

Manovich L (2020) Cultural Analytics. Cambridge, Massachusetts: The MIT Press

Mehdi M, Okoli C, Mesgari M, Nielsen FÅ, Lanamäki A (2017) Excavating the mother lode of human-generated text: A systematic review of research that uses the wikipedia corpus. Inf Process Manag 53(2):505–529. https://doi.org/10.1016/j.ipm.2016.07.003

Mesoudi A (2011) Cultural Evolution, How Darwinian Theory Can Explain Human Culture and Synthesise the Social Sciences

Miccio LA, Gámez-Pérez C, Suárez JL, Schwartz GA (2022) Mapping the Networked Context of Copernicus, Michelangelo, and della Mirandola in Wikipedia. Adv Complex Syst 25(05n06):2240010. https://doi.org/10.1142/S0219525922400100

Michel JB, Shen YK, Aiden AP, Veres A, Gray MK, Team TGB et al. (2011) Quantitative analysis of culture using millions of digitised books. Science 331(6014):176–182. https://doi.org/10.1126/science.1199644

Newman MEJ (2003) The structure and function of complex networks. Siam Rev 45(2):167–256. https://doi.org/10.1137/S003614450342480

Rantala H, Leskinen P, Peura, L, Hyvönen E (2024) Representing and Searching Associations in Cultural Heritage Knowledge Graphs Using Faceted Search. in Salatino A et al. (eds) (2024) Knowledge Graphs in the Age of Language Models and Neuro-Symbolic AI. Studies on the Semantic Web 60: 420-435. https://doi.org/10.3233/SSW240033

Richerson PJ, Boyd R (2006) Not By Genes Alone. How Culture Transformed Human Evolution. Chicago University Press. 5

Ruprechter T, Santos T, Helic D (2020) Relating Wikipedia article quality to edit behavior and link structure. Appl Netw Sci 5(61). https://doi.org/10.1007/s41109-020-00305-y

Sandoz R (2021) Thematic reclassifications and emerging sciences. J Gen Philos Sci 52(1):63–85. https://doi.org/10.1007/s10838-020-09526-2

Schich M, Song C, Ahn Y-Y, Mirsky A, Martino M, Barabási A-L et al. (2014) A network framework of cultural history. Science 345(6196):558–562. https://doi.org/10.1126/science.1240064

Schwartz GA (2021) Complex networks reveal emergent interdisciplinary knowledge in Wikipedia. Humanit Soc Sci Commun 8(127). https://doi.org/10.1057/s41599-021-00801-1

Smolla M, Akçay E (2019) Cultural selection shapes network structure. Sci Adv 5(8):eaaw0609. https://doi.org/10.1126/sciadv.aaw0609

Snelders HAM (1989) Christiaan Huygens and Newton's Theory of Gravitation. Notes Rec R Soc Lond 43(2):209–222. https://www.jstor.org/stable/531383

Spinney L (2012) Human cycles: History as science. Nature 488(7409):24–26. https://doi.org/10.1038/488024a

Suárez JL, Sancho F, Rosa J, de la (2012) Sustaining a global community: Art and Religion in the Network of Baroque Hispanic-American Paintings. Leonardo 45(3):281. https://doi.org/10.1162/LEON_a_00374

Suárez JL, McArthur B, Soto-Corominas A (2015) Cultural networks and the future of cultural analytics. Proceedings of the International Conference on Culture and Computing. Kyoto University, Japan. https://doi.org/10.1109/Culture.and.Computing.2015.37

Suárez JL (2025) Reassembling and generating cultural networks: A digital humanities research agenda Metode Sci Stud J 14:25–29. https://doi.org/10.7203/metode.15.27784

Thompson ME, Ramirez LLR, Lyubchich V, Gel YR (2016) Using the bootstrap for statistical inference on random graphs. Can J Stat 44(1):3–24

Tucker R (2010) The Patronage of Rembrandt's Passion Series: Art, Politics, and Princely Display at the Court of Orange in the Seventeenth Century. Seventeenth Century 25(1):75–116. https://doi.org/10.1002/cjs.11271

Turchin P (2008) Arise "cliodynamics. Nature 454(7200):34–35. https://doi.org/10.1038/454034a

Villmoare B editor (2022) The Evolution of Everything: The Patterns and Causes of Big History. Press CU

Wade D (2017) Geometry & Art. Shelter Harbor Press

Witten IH, Milne DN (2008) An effective, low-cost measure of semantic relatedness obtained from Wikipedia links: 25-30

Yndurain D (2016) El fin del humanismo tradicional. Universidad de Huelva

Zheng X, Chen J, Yan E, Ni C (2022) Gender and country biases in Wikipedia citations to scholarly publications. J Assoc Inf Sci Technol 74:219–233. https://doi.org/10.1002/asi.24723

Acknowledgements

GAS acknowledges the financial support from the Spanish Government, project PID2023-146348NB-I00 financed by MICIU/AEI/10.13039/501100011033 and by FEDER (UE). GAS and PA acknowledge the financial support from the Donostia International Physics Centre (Programa Mestizajes) as well as the support of NVIDIA Corporation with the donation of a Quadro RTX 6000 GPU used for this research.

Author contributions

LAM performed the data mining and calculations. PA participated in the data mining and contributed to the data analysis. CGP and FG contributed to the data analysis and the writing. JLS participated in the data interpretation, writing and supervision. GAS proposed the idea, supervised the calculations and the data analysis, participated in the writing, coordinated the work and got the funding.

Competing interests

The authors declare no competing interests.

Ethical statement

This article does not contain any studies with human participants performed by any of the authors.

Informed consent

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-04772-5.

Correspondence and requests for materials should be addressed to Gustavo A. Schwartz.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.