AIKoGAM: An AI-driven Knowledge Graph of the Antiquities Market: Toward Automatised Methods to Identify Illicit Trafficking Networks

RICCARDO GIOVANELLI 🆔

ARIANNA TRAVIGLIA 🆔

*Author affiliations can be found in the back matter of this article

]u[ubiquity press

ABSTRACT

The longstanding illicit trafficking of archaeological artefacts has persistently presented a global issue, posing a substantial threat to cultural heritage. This paper introduces an innovative automated system that utilises Natural Language Processing (NLP), Machine Learning (ML), and Social Network Analysis (SNA) to construct a Knowledge Graph for antiquities. The objective is to offer insights into the provenance of artefacts and identify potential instances of illicit trafficking. The paper delineates a comprehensive methodology, from the ontology to the Knowledge Graph. The methodology comprises four distinct phases: the initial phase involves tailoring existing ontologies to match project-specific needs; the second phase centres on selecting appropriate technologies, and scraping and text-mining tools are designed to extract pertinent data from textual sources; the third phase centres in the creation of a robust and accurate Knowledge Graph that captures artefact provenance. The paper suggests employing NLP models, specifically utilising Named Entity Recognition (NER) techniques. These models automatically extract relevant information from the unstructured provenance texts, organising them as events to which both objects and actors participated with their locations and dates. The final phase is concerned with defining and building the Knowledge Graph. The authors explore a property graph model that distinctively represents nodes and relationships, each augmented by associated properties. Employing an SNA approach, the model is projected in multiple network levels of ownership histories (actor-object network) or actor relationships (actor-actor network). This approach reveals patterns within the antiquities market. When integrated with the authors' recommended strategies such as crowdsourced ontology definition, collaboration with reputable organisations for quality sources, and the application of transfer learning techniques, the suggested approach holds promising implications for the protection of cultural heritage.

CORRESPONDING AUTHOR:
Riccardo Giovanelli

Università Ca' Foscari Venezia, Italy; Istituto Italiano di Tecnologia - Center for Cultural Heritage Technology, Italy

riccardo.giovanelli@unive.it

1 INTRODUCTION

The antiquities market, both legal and illegal, is a widely acknowledged phenomenon, prompting significant attention from national governments and international bodies aiming to regulate it. This commitment is reflected in dedicated funding programs aimed at comprehending and addressing the issue.

As emphasised by Graham et al. (2023), there is a growing trend towards digital and technological solutions, evidenced by European Commission funding for several initiatives. These include RITHMS (2023), a Multiplex SNA-based platform, and projects employing 3D photogrammetry, AI for border control (CORDIS 2022a), chemical marking, miniature devices, blockchain against illicit activities (CORDIS 2022b), and technologies for the identification, traceability, and provenance research of cultural goods (CORDIS 2022c; Patias & Georgiadis 2023).

Parallelly, the debate surrounding the antiquities ownership, market and cultural rights persists (Campfens 2020; Cuno 2012; Hixenbaugh 2019), with efforts to involve art market representatives in decision-making processes.

Despite claims by collectors, dealers, and auction houses of adhering to ethical practices, UNESCO (2022b) data indicates a shortfall in meeting standards for diligently vetted objects with solid ownership histories. The survey revealed a significant lack of due-diligence policies (15%), with the 33% neglecting provenance research for lower-priced items, and the 54% not giving attention to objects originating from conflict zones, leading to a market with a wide range of unlawfully sourced artefacts.

The 2022 'Operation Pandora VII', executed by Law Enforcement Agencies (LEAs), underscores this, with 8,495 online checks leading to the confiscation of 4,017 illicit goods (EUROPOL 2023; Giovanelli 2023). However, as Brodie et al. (2022: 117) argue, 'if the reported arrests and seizures are only the tip of a much larger dark criminal iceberg, then the indications are that the illicit trade in cultural objects is very poorly controlled'. Nonetheless, it is worth noting that most of the seizures and identifications were made possible through crosschecks against LEAs databases.

While databases of stolen goods exist (CECOJI-CNRS and EC-DGHOME 2019: 82–89; Oosterman 2019), their effectiveness is limited when dealing with illegally excavated items lacking prior documentation (Rush & Millington 2015; EC-DGEAC et al. 2019: 177; Roodt 2016: 221). These databases prove inadequate in such cases, with a few exceptions in which LEAs have managed to seize written and photographic evidence of the unlawful origin of cultural goods. Noteworthy instances include the integration of the 'Medici dossier' within the Italian Carabinieri's LEONARDO database (Watson & Todeschini 2007; Gill & Tsirogiannis 2016; CC-TPC 2023b; Pellegrini 2023) and the seizure of Almagi`a's Archives by the New York County District Attorney (Bogdanos & Iyer 2021; Giovanelli 2021b). In such scenarios, the monitoring of marketplaces by both LEAs (CC-TPC 2023a) and archaeologists/experts (Gill & Tsirogiannis 2016; Tsirogiannis 2021) has proven instrumental in successfully identifying looted cultural goods and eventually enabling their recovery.

Significant gaps persist in understanding the antiquities trade, particularly the 'demand' side (Brodie et al. 2022). While we concur with Brodie et al. (2022: 121) that many cultural goods appearing on the 'nearly infinite number of storefronts for cultural objects' being declared as looted stem from old illegal excavations, these marketplaces harbour valuable data yet to be fully utilised. Galleries, auction houses, and dealers catalogues and websites form extensive archives within a big data framework (Boyd & Crawford 2011), offering an overview, albeit partial (Brodie et al. 2022; Yates 2022), of the trade's current state and historical evolution.

Provenance statements, critical in investigations and in understanding trade dynamics, are increasingly scrutinized for their quality and verifiability (Albertson 2020a; Brodie 2015; Davis 2020; Fabiani & Marrone 2021; Fletcher 2020; Gill & Tsirogiannis 2016; Giovanelli 2018; Levine 2009; Mackenzie & Yates 2020; Tsirogiannis 2015a; Tsirogiannis 2015b; Tsirogiannis 2016; Tsirogiannis 2021). Provenance is crucial for unravelling 'an object's trajectory as it changes legal status while crossing international borders' (Gerstenblith 2020), and the examination of the 'provenance given by the art dealer' (UNESCO 2021) is a critical step in initiating a recovery procedure, as potential evidence to support restitution claims (UNESCO 2022a).

Acknowledging the significance of grasping market dynamics, the pivotal role of provenance, the growing emphasis on digital progress, and the challenges and resolutions emerging from extensive data exploration in the humanities (Smith 2018), this paper proposes a semi-automated system combining Natural Language Processing (NLP), Machine Learning (ML), and Social Network Analysis (SNA) to construct a Knowledge Graph (KG) of the antiquities market. Our aim is to generate fresh insights into the antiquities market dynamics and its actors, incorporating Artificial Intelligence (AI) technologies to correlate the market with existing trafficking networks knowledge.

2 BACKGROUND: ANTIQUITIES TRADE AS A COMPLEX NETWORK

2.1 NETWORK ANALYSIS

Our approach views the trade of cultural heritage as a complex network, essentially a graph of nodes and interconnected edges with non-trivial topological features (Barabási & Pósfai 2016; Newman 2003), aptly suited for

application to markets (Granovetter 1985; Jackson 2010) and arts and humanities studies (Brughmans, Collar, & Coward 2016; Brughmans & Peeples 2018; Brughmans & Peeples 2023; Collar et al. 2015; Meirelles, Schich, & Malina 2014).

The integration of network models to describe the antiquities trade, implicitly recognised as hierarchical during the 1990s Medici investigations (Watson & Todeschini 2007), has evolved. Preceding these developments, within the domain of visual and contemporary art markets, the application of network theories as effective models for analysing art trade began to emerge during the late 1960s (Yogev & Grund 2012). Yogev and Grund (2012) employed SNA to scrutinise galleries and artists in art fairs. Campbell (2013) described the illicit antiquities market as a 'fluid network model', aligning with criminological literature on criminal organisations' involvement (Williams & Godson 2002) and highlighting its transnational nature (Chappell & Polk 2011; Mackenzie 2011; Mackenzie & Davis 2014; Mackenzie, Brodie, et al. 2019; Palombo & Yates 2023).

Seminal works applying SNA include D'Ippolito (2014), who designed two models of the network involving Giacomo Medici, utilising the organigram found in the possession of Pasquale Camera (Isman 2009: 181–184, Apparati; La Corte Suprema di Cassazione, Sezione II penale 2011: 4; Tribunale di Roma 2004: 236–239), augmented by associated open-source data. The first model represents a structural network where nodes are categories of actors, while the second model focuses on Medici's ego-network (Granovetter 1983; Marsden 2005; Perry, Pescosolido & Borgatti 2018). By employing degree, eigenvector, betweenness and closeness centrality measures, D'Ippolito showcases how the network displays fluidity and a non-hierarchical structure. Additionally, the study highlights the considerable influence that Giacomo Medici exerted over his connections.

Tsirogiannis and Tsirogiannis (2016) developed a model that sheds light on the structural biases of the illicit trade network, which result from inaccurate information, missing links, and predefined boundaries. Their model considers the interactions among 98 actors and their relationships, as described in Watson and Todeschini (2007), which provides a detailed portrayal of the antiquities trade world from 1972 to 2001. Tsirogiannis and Tsirogiannis designed two representations of the network, one directed and one undirected, both sparse and consisting of two connected components. The authors introduced three innovative algorithms. The first two algorithms aimed to evaluate transaction paths within the network and were compared to Dijkstra (1959) Shortest Path methods. The third algorithm, which aimed to predict missing links, underwent testing against the Jaccard (1912) method. The authors' algorithms outperformed the existing methods. However, they acknowledged limitations when dealing with incomplete data (Ahmad et al. 2020; Lü

& Zhou 2011; Ryan & Ahnert 2021). Nevertheless, the findings underscore the significant contribution of the authors' innovative algorithms in enhancing the analysis of the illicit trade network.

Fabiani and Marrone (2021) explored the circulating antiquities market, focusing on Christie's, Bonham's, and Sotheby's sales from 2007 to 2015. They present an insightful example of how auction catalogue text—even though reporting incomplete data—can be effectively transformed into a bipartite network (Liew et al. 2023; Newman 2003) of 'items' and auction houses, providing a visualisation of the flow of 33,709 antiquities between auction houses and dealers across time and space. The study offers numerous valuable results, which are too extensive to summarise here. Remarkably, the analysis emphasises the importance of 20 dealers within the network, with several of them having 'known links with the illicit trade' (Fabiani & Marrone 2021: 24). These findings, in conjunction with patterns of sales, further reinforce the doubts about sourcing methods employed by these dealers.

2.2 KNOWLEDGE GRAPHS

The concept of Knowledge Graph (KG) is relatively new, gaining popularity through Google's association (Singhal 2012), with no consensus regarding its definition (Ehrlinger & Wöß 2016; Fensel et al. 2020; Hogan et al. 2021). Originating in 1974 as a graph-based abstraction of the teaching-learning process (Marchi & Miguel 1974), where nodes represent individual pieces of knowledge and directed edges depict prerequisites, KGs have evolved with various interpretations and formal structures (Hogan et al. 2021: Appendix A). Like SNA, which represents actors as nodes and relationships as edges, KGs are large-scale graphs assembled from diverse sources to accumulate and convey real-world knowledge (Hogan et al. 2021), organising data with graph-based data models, ontologies or rules to represent quantified statements. The structure and rationale of KGs are part of, and align with, concepts in the Semantic Technologies, Linked Open Data (LOD), and ontologies (Gujral & Shivarama 2023), as they aim to organise specific domains of knowledge. Furthermore, KGs have the potential to accumulate new knowledge through inductive or deductive methods (Hogan et al. 2021), as well as to integrate naturally with ML (Yu et al. 2022).

As discussed, the matter of antiquities trade encompasses a variety of perspectives, methodologies, data sources, and network models from diverse fields of study. In this multidisciplinary realm, numerous approaches and viewpoints have been employed to analyse this intricate phenomenon. Given the demonstrated benefits of applying SNA in this field, it is only logical to explore graph theory as a whole and expand the research focus to include graph databases and KGs, as highlighted by Dörpinghaus et al. (2022).

Rapid progress in cultural heritage KGs (Pellegrino, Scarano & Spagnuolo 2022) contrasts with limited development in illicit trafficking of antiquities. Brennan and Tsirogiannis (2022) work on an interactive KG using the forensic data from Christos Tsirogiannis is one first example. Their web-based platform maps 48 trafficked objects, including those associated with Giacomo Medici (CC-TPC 2023b; Gill & Tsirogiannis 2016; Pellegrini 2023; Watson & Todeschini 2007), Gianfranco Becchina (Pellegrini & Rizzo 2021), and Edoardo Almagià (Giovanelli 2021a; Giovanelli 2021b), visualizing networks of artefacts, individuals, institutions, archives, and histories. The web-graph facilitates the visualisation of both the ego-network of each object and the entire network. Nonetheless, the lack of official documentation, an open repository, or a white paper for the graph hinders our comprehension. Consequently, the description provided relies solely on observations gleaned from the website.

Graham et al. (2023) constructed a KG using the 129 case study summaries that specifically address the illicit trafficking of cultural heritage, authored by domain experts, in Trafficking Culture (2023). The authors scraped the texts and manually tagged entities (individuals, businesses, locations, objects) and various types of relationships within the dataset using the INCEpTION semantic annotation tool (Klie et al. 2018). The resulting graph consists of subject-predicate-object triples. The authors thus used the Complex embedding module (Trouillon et al. 2016) from the Ampligraph Python Library (Costabello et al. 2019), projecting a vectored multidimensional representation of the graph on which to perform link predictions. The link prediction applied by Graham et al. (2023) was performed by hypothesising new triples combination between the existing entities. The experiments on the two-dimensional visualisation of the model revealed close proximity between Leonardo Patterson, a renowned dealer of pre-Columbian antiquities (Yates 2015), and the Brooklyn Museum. Subsequent investigations into these two entities unveiled the previously unnoticed fact that Leonardo Patterson donated two objects to the Brooklyn Museum in 1969. This significant discovery has paved the way for new avenues of research, underscoring the importance of the work conducted by the authors. However, the authors acknowledge certain limitations in their study. One major limitation is that the model's predictive power is constrained by the existing data. Moreover, the manual annotation process required for constructing the KG is recognised as a time-consuming undertaking.

To address the second limitation, Graham, Yates, and El-Roby (2023) have explored the potential of leveraging existing Large Language Models (LLMs) to automate the tagging and population of their KG. Building upon the same dataset, and adapting the work of Huang (2022), they developed a prompt to interact with the *OpenAI text-DaVinci-003 GPT-3.5* model (OpenAI 2023) API.

Through this automated process, they aimed to extract all possible directed relationships involving individuals, organisations, places, and objects. The extracted information, along with the connected nodes and their respective labels, were outputted as key-value pairs, from which they generated a set of triples, following the methodology employed in their previous work. Their approach, while less accurate, produced comparable results during the link prediction and significantly reduced the time required for the process.

3 METHODOLOGY

In line with current trends in computational research on the illicit trafficking of cultural property, our work focuses primarily on researching, exploiting, and refining various ML tools and algorithms. Our goal is to develop a semi-automated, interoperable, and accessible methodology that can benefit other research groups and LEAs alike.

This research initiative, named AIKoGAM (Giovanelli & Traviglia 2023), aims to establish a common ontology following the LOD and Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson et al. 2016) principles, and seeks to create an open data environment that fosters knowledge generation from existing resources and encourages data enrichment, sharing, and reuse.

3.1 INFORMATION EXTRACTION

3.1.1 Web harvesting

Online data accessibility for antiquities in the market has increased since Christie's introduced digital access to bidding in 2006 (Christie's 2012), further boosted by Sotheby's partnership with eBay in 2014 (eBay Inc. Staff 2014). Consequently, auction houses, galleries, and dealers have increasingly adopted digital platforms to showcase their catalogues.

The Information Retrieval (IR) module in AIKoGAM starts with identifying and examining data sources, including selecting target websites, locating pertinent data linked to archaeological objects and provenance, comprehending the structure, quality, and nature of these data, scrutinising data source usage policies, assessing data availability, considering access restrictions, and formulating a scraping strategy.

For auctions, specific sales and corresponding lots can usually be located through a straightforward website architecture. This typically involves navigating from the 'Home page' to the 'Past auctions page' or the 'Future auctions page', leading to a 'Sale page' (Figure 1) and finally reaching the 'Lot page'. For galleries, the website structure may vary, but it usually involves browsing through an 'inventory page' (Figure 2) displaying all the objects or categorised pages (Egyptian, Roman, Greek etc.) listing objects within each respective category.
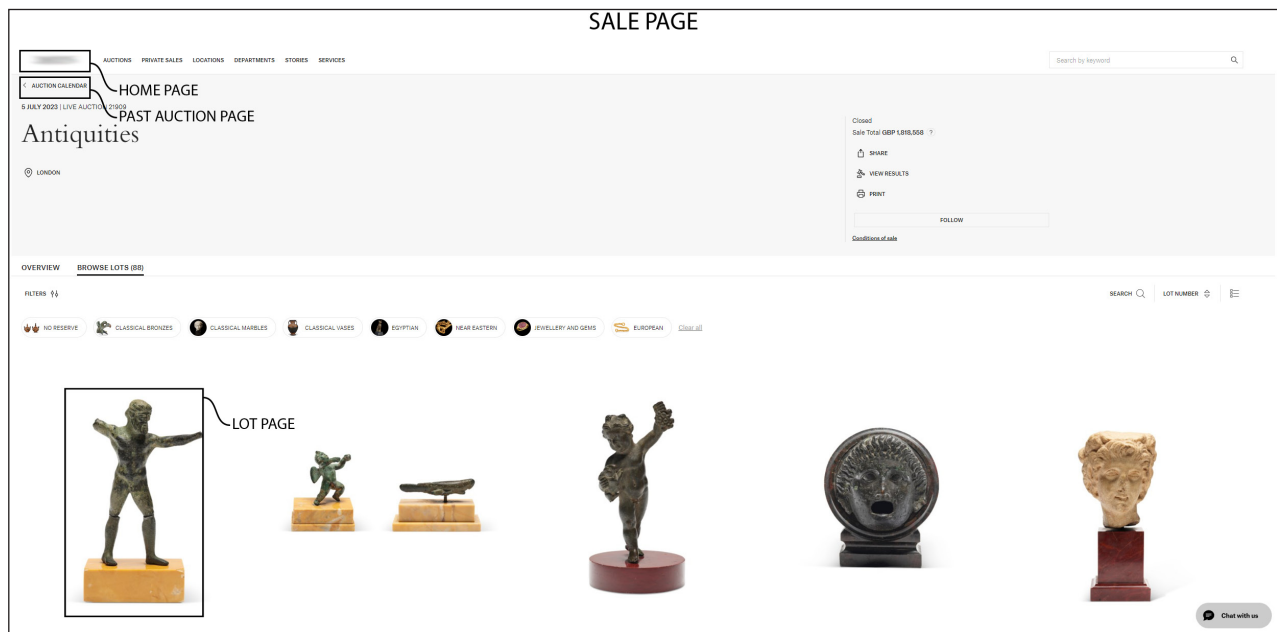
Figure 1 Example of a specific 'sale page', highlighting the straightforward structure of typical auction houses' websites. Screenshot: Christie's (2023b), edited by the authors.
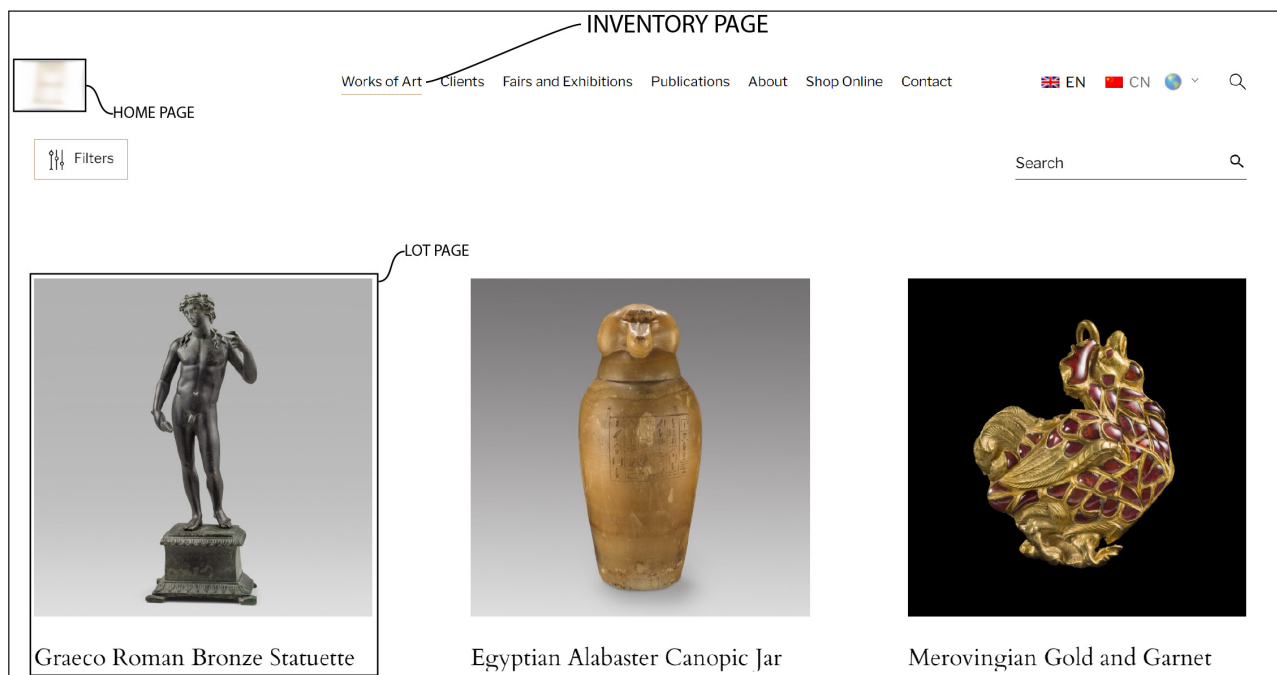

Figure 2 Example of a specific 'inventory page' of an antiquity gallery, highlighting the straightforward structure of typical gallery websites. Screenshot: Phoenix Ancient Art (2023b), edited by the authors.

Although these web pages appear straightforward and present the relevant information about the objects in a human-readable fashion (see Figure 3), underlying Hyper-Text Markup Language (HTML) reveals significant differences in structure among websites. Some adopt structured graph databases, making data retrieval easy for both users and software thanks to their contextual schemas. Conversely, many websites rely on HTML elements (such as headings, paragraphs, lists, etc.) combined with Cascading Style Sheets (CSS) selectors for content organisation. CSS selectors enable selective styling of HTML elements or groups of elements, enhancing the visual appeal and layout.

Certain web pages are well structured, adhering to a consistent layout and appearance. They use designated HTML elements and CSS classes to maintain this predictability. Structured web pages facilitate information retrieval because data is clearly organised in an anticipated manner: product listings are structured using '<div>' elements for individual products, simplifying data extraction using CSS selectors.
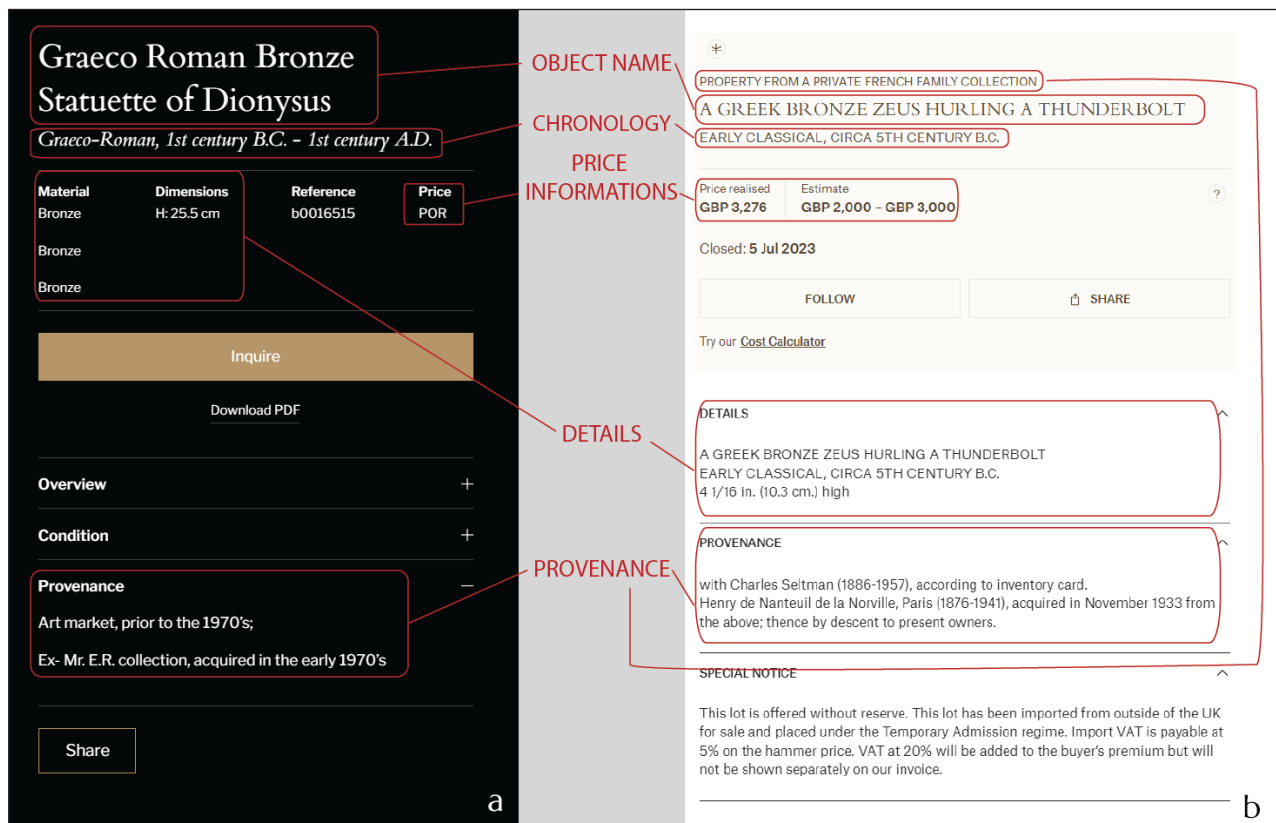
Figure 3 Examples of how relevant information are placed on lot pages of **a)** an antiquity gallery (Phoenix Ancient Art, 2023a) and **b)** an auction house (Christie's 2023a).

Some others do not have a strict structure, resulting in a free-form layout. Retrieving information from such pages can be challenging as content is not neatly organised. Information may be embedded within paragraphs or other textual content, necessitating more complex extraction methods.

Sotheby's utilises an Apollo GraphQL Database for systematic data organisation, allowing straightforward data access via the database Application Programming Interface (API), or through uncomplicated scraping methods. Christie's has a different structure where certain information like object names, estimates, or image URLs are retrievable via API calls, while other details, such as provenance, literature, description, and exhibitions, require HTML element detection with specific CSS selectors.

To handle these differences, the 'harvesting.py' script in Python was developed, utilising Beautiful Soup 4 (Richardson 2019) and Requests (Chandra & Varanasi 2015) libraries for structured data collection. Initially, manual fine-tuning is required to maximise data collection automation and minimise manual cleaning and restructuring of the data afterwards.

3.1.2 Ontology mapping and Knowledge Graph database building

To maximise interoperability and applicability across various websites, the harvesting module does not directly produce a uniform database due to varying keys in the harvested dictionaries. We developed an ontology and a vocabulary to map harvested data to specific entities and relationships. This facilitates future AI automation and reduces the need for manual cleaning. The vocabulary ensures the mapping of different keys pointing to the same logical instance, which lays the foundations for the KG and future interoperability with other ontologies and LOD.

For AIKoGAM, we adapted existing ontologies for general entities and proposed a new ontology for provenance reasoning, capturing both the synchronic relationship of each owner to the objects and the diachronic dimension (Figure 4).

A single provenance statement is treated as an event involving the object, the seller, the buyer, and the subsisting market event (such as an auction sale or the presence in a specific market environment, inferred from statements like 'Bought by the present owner in London antiquities market, 1980'). To build a Neo4j database instance with this schema, we developed the 'mapping.py' script. This script employs data-cleaning techniques and regular expressions and generates unique identifiers for each sale and lot utilising the Secure Hash Algorithm 512, which produces a fixed string of characters representing the data. Thanks to the algorithm's collision resistance properties (USDOC-NIST 2015), we can guarantee uniqueness and adhere to data indexing constraints.

To enrich the provenance information of the objects, we have devised the 'event_extractor.py' script that leverages the Spacy library NER models (Montani et al.
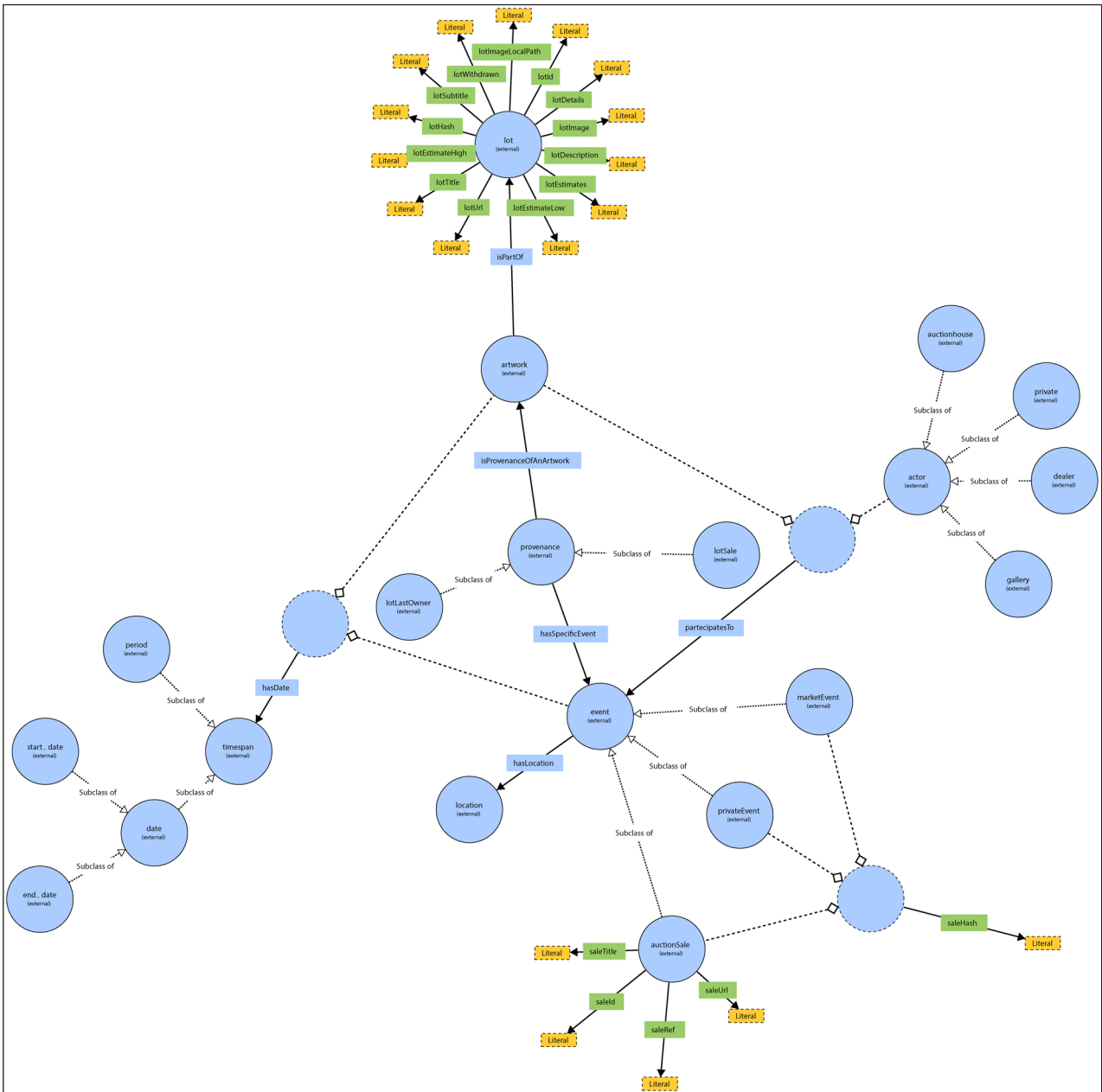
Figure 4 AIKoGAM partial ontology schema.

2020) to split provenance statements into 'events'. Each of these events ideally contains an actor, a location, and a time-span indication. Recognising linguistic diversity within our dataset, our 'detect_language_batch' function uses the LangID Python library (Heafield, Kshirsagar & Barona 2015), a high-speed language identification tool pre-trained across 97 languages. This function associates the language identifier with the corresponding SpaCy NER model, facilitating the categorisation and storage of the relevant SpaCy model variable within each statement. The script systematically processes each provenance statement, extracting NER entities based on the identified language. To address discrepancies in entity labels across distinct SpaCy models, we pass the recognised entities through a constant dictionary that maps entities to specific standardised labels, ensuring uniformity in the output. The script then stores within each dataset entry

the various 'events' in which the object participated. This information is structured as shown in the following example: the provenance statement 'Sotheby's Hong Kong 04 August 2022 lot 1095' is processed by the 'detect language batch' function into the dictionary:

> {'text': 'Sotheby's Hong Kong 04 August 2022 lot 1095', 'detected language': 'en',
> 'spacy model': 'en_core_web_md}.

Subsequently, The 'event_extractor.py' script stores the event within the entry as:

> {'label': 'Sotheby's Hong Kong 04 August 2022 lot 1095', 'ORG': 'Sotheby's Hong Kong',
> 'DATE': '04 August 2022',
> 'NUMBER': '1095'}.

The 'kg_construction.py' script then builds into a desktop instance of a Neo4j (2023b) the enhanced KG by automatically extracting events from each object and constructing new nodes based on them. These event nodes are connected to the corresponding artworks through a 'PARTICIPATED_TO_EVENT' relationship.

Through CYPHER queries (Neo4j 2023a), the native query language for Neo4j, the named entities are extracted from the event nodes properties and generates new nodes as specified in the ontology schema depicted in Figure 4. The event node remains central, receiving relationships from the artworks and projecting or receiving new relationships with other relevant entities.

To evaluate the general SpaCy models accuracy, we checked the results of a dataset sample (976 items). Of the 2671 entities extracted from these 976 items, 94 were incorrectly identified, resulting in an overall accuracy of approximately 96.49% (considering a correct identification if a collection name or an organisation name is labelled under 'PERSON' tag and vice versa, given the general tendency of having personal names in organisation names and the high chance of error. Given that our ontology mapping considers both 'PERSON' entities and 'ORG' entities as 'actor' entities, we consider both identifications as correct).

Examining the error rates across different models reveals variations in performance (see Tables 1 and 2).

The most frequent error involves entities better identifiable as 'OTHER' being misclassified as 'ORG', occurring 17 times. This pattern suggests a challenge where entities difficult to categorise are seen as organisations, indicating potential areas for model improvement.

Analysing misclassifications by entity type, 'PERSON' and 'ORG categories are often misclassification of 'OTHER' entities, while 'GPE' misclassifies 'PERSON' and vice versa. This highlights challenges in accurately identifying individuals and geopolitical entities in specific contexts.

While the overall accuracy of the NER models is commendable, there are notable variations in performance across different models, with specific challenges in accurately classifying certain entity types. Understanding these nuances is crucial for refining and optimising the models to enhance their performance in real-world applications.

4 PRELIMINARY RESULTS

Through the transformative process, we evolved from an original database with 30,954 entries (30,422 objects, 529 sales, and 3 entities) to a KG database containing 111,979 nodes and 237,394 relationships. This enhanced KG offers a detailed representation of provenance data, enabling us to extract valuable insights.

Figure 5 summarises the node and relationship counts in the KG: 13,482 actors involved in selling or acquiring

MODEL	ERROR RATE (%)
en_core_web_md	3.35
zh_core_web_sm	6.17
de_core_news_sm	6.25
es_core_news_sm	6.67
fr_core_news_sm	11.46
it_core_news_sm	16.67
pl_core_news_sm	20.00
nl_core_news_sm	32.36
lt_core_news_sm	33.33
pt_core_news_sm	33.33
da_core_news_sm	66.67

Table 1 Error Rates per Model.

PREDICTED	TRUE	COUNT
ORG	OTHER	17
PERSON	OTHER	16
GPE	OTHER	11
GPE	PERSON	10
PERSON	GPE	4
ORG	GPE	4
FAC	PERSON	4
GPE	ORG	4
ORG	DATE	3
NORP	OTHER	3
NORP	PERSON	2
GPE	DATE	2
FAC	ORG	2
FAC	GPE	2
MONEY	OTHER	1
PRODUCT	ORG	1
PRODUCT	GPE	1
NORP	ORG	1
PRODUCT	OTHER	1
WORK OF ART	ORG	1
NORP	DATE	1
WORK OF ART	DATE	1
WORK OF ART	OTHER	1

Table 2 All Errors and Their Counts.

30,422 antiquities dealt with by Christie's, Sotheby's, and Phoenix Ancient Art from 1999 to 2023.

After a manual examination of the data, duplicate occurrences in the 'actors' became apparent. We
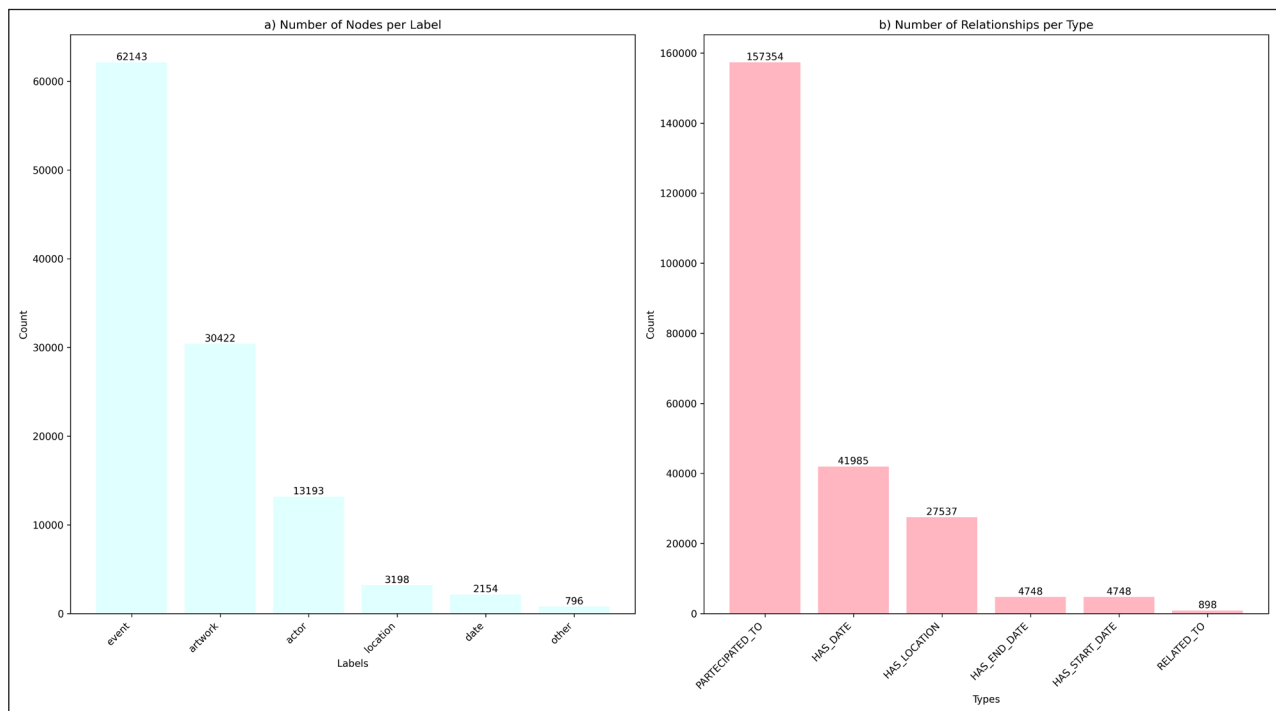
Figure 5 Nodes **(a)** and Relationships **(b)** counts by their labels and types.

introduced the 'similarity.py' script with a 'find_similar_nodes' function, identifying similar 'actor' nodes. The function stores pairs of nodes with similar names along with their similarity score. The 'ngram_index' function generates the index for identifying similar strings by decomposing them into smaller units known as n-grams (Krátký et al. 2011). The function retrieves the name of each 'actor' node, normalises it, and prepares it for subsequent comparisons. Within the loop that iterates over all the nodes again, the function systematically compares the normalised names of the two nodes. With an exact match the function categorises the two nodes as similar and the shorter of the two names is added to the list of merged strings. When the normalised names do not align exactly, the function employs fuzzy string matching (SeatGeek 2023) to calculate the similarity score. If this score exceeds the 0.9 threshold, the node pair is appended to the list of similar nodes. For instances where the similarity score is a perfect match (score 1.00), a similar consolidation process to the exact match scenario is executed.

The 'semi_auto_evaluation.py' script addresses occurrences of nodes representing the same entities but perceived as different due to variations in naming conventions. The comparison is performed on single words composing the string of the node name, through the 'process_name' function. The automated portion of the script identifies nodes with single words that match precisely and employs fuzzy matching (threshold 0.72) to calculate similarity scores. This accommodates cases where names are similar but not identical, addressing variations like 'thomas', 'Thmas', and 'Tommaso'. Where exact fuzzy matches are not found, the script checks for length consistency to filter out minor discrepancies and

considers a specific lesser similarity ratio (0.68 to 0.72). This is needed due to the heterogeneity of the data sources and languages, where substantial changes in name representation may occur. The script recognises potential matches by comparing trailing names, a mechanism useful for dealing with variations in naming conventions such as differences in titles, initials, and other suffixes, as well as residual contextual information. In cases where no duplicates are identified, the output is automatically generated. When matches are identified, the script prompts the user for manual confirmation based on the presented potential matches. This human intervention enables the mapping of nodes with strongly diverse representations. As a result, a dictionary is stored with the first encountered name as the key and the names to be merged as values. Subsequently, the head node is connected to the neighbours of the nodes to be merged, transferring their relevant edges and attributes, and then they are removed from the graph. This ensures the preservation of the graph structure and essential relationships during the clean-up.

The process identified 2,540 potentially similar nodes, with 791 (5.87% of the total actors) automatically merged. The 'semi_auto_evaluation.py' implementation facilitated the merging of an additional 907 actors (6.72% of the total actors), averaging 2.2 duplicates per node. This process resulted in a total of 1,698 nodes representing alternative spellings of the same entity. Nodes mislabelled or too general accounted for 523. These nodes were removed from further projections and re-labelled onto the Neo4j graph instance. In summary, from the initial 13,482 actors identified by SpaCy, our refined scripts and processes reduced their number to 11,479 single actors.
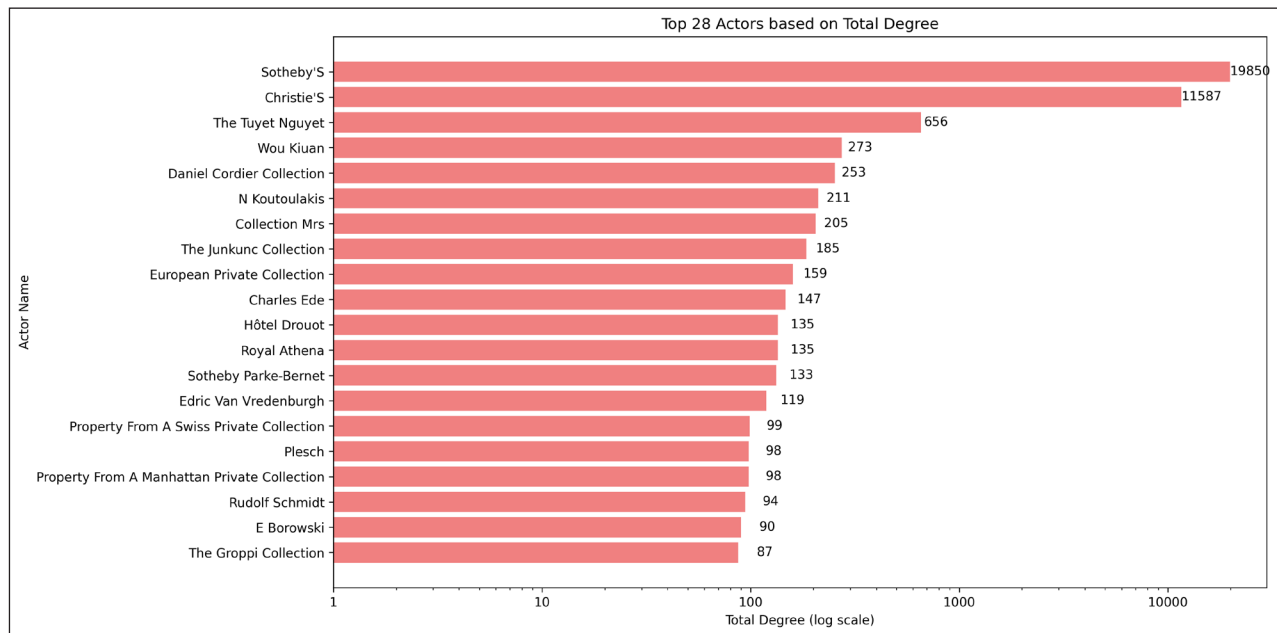
Figure 6 Nodes with the highest degree in the actors-artwork bipartite network.

For SNA on the KG, we experimented with an artworks-actor bipartite projection (Newman 2003; Liew et al. 2023). The top centrality degrees (Freeman 1978) in Figure 6 show that aside from Christie's and Sotheby's there are other notable actors. A prominent figure is Stephen Junkunc, a well-known collector of Chinese art, whose collection is highly desirable and has gained prominence in the market (Sotheby's 2019). Notably, in 2018 Sotheby's withdrew the sale of a statue of Buddha suspected to be a looted relic from the Longmen Grottoes and coming from his collection (The Value 2018).

Noteworthy names already connected with cases of illicit trafficking of antiquities appear, such as Royal Athena (Galleries), N(icholas) Koutoulakis, and E(lie) Borowski, a piece of evidence that very well aligns with the finding of Fabiani and Marrone (2021), whose datasets cover sales up to 2015. These three actors have been significantly involved in numerous cases of cultural object trafficking, as documented in previous works (Chasing Aphrodite 2012; Gill 2019; Tsirogiannis 2021).

To gain a deeper understanding, we build an additional layer, generating a monopartite undirected weighted network of actors (see Figure 7). In this process, we excluded the nodes representing Christie's and Sotheby's to avoid bias in our analysis. This allowed us to focus on the relationships among other actors within the network. Removal of out-of-scale high-degree nodes does not lead to network fragmentation, as demonstrated in previous research (Nguyen et al. 2021).

We refactored the artworks connected to actors in our previous layer, transforming those nodes into undirected edges between actors (Ramasco & Morris 2006): an edge between $actor_i$ and $actor_j$ has thus weight w_{ij} equal to the number of artworks both dealt.

METRIC	VALUE
Nodes	6159.0000
Relationships	22390.0000
Average Degree (AD)	7.2707
Average Degree Centrality (ADC)	0.0012
Average Eigenvector Centrality (AE)	0.0012
Average Betweenness Centrality (AB)	0.0007
Average Closeness Centrality (AC)	0.2123
Average Path Length (APL)	4.8148
Average Clustering Coefficient (ACC)	0.7161
Graph Edge Density (GED)	0.0012
Graph Diameter (GD)	14.0000

Table 3 Summary of Network Measures.

The monopartite projection, as indicated by the metrics in Table 3, appears relatively large, characterised by significant complexity and interconnection. The Average Degree (AD) suggests moderately linked individuals, fostering a decentralised pattern of influence. Centrality measures reinforce this, indicating a lack of concentrated control. Substantial mutual connectivity among entities facilitates efficient communication, as reflected in the Average Closeness Centrality (AC). A relatively short Average Path Length (APL) and a high Average Clustering Coefficient (ACC) highlight tightly interconnected clusters with cohesive subgroups. Despite sparse overall connections (GED), the network has a substantial diameter (GD), with long-distance connections contributing to the overall complexity.
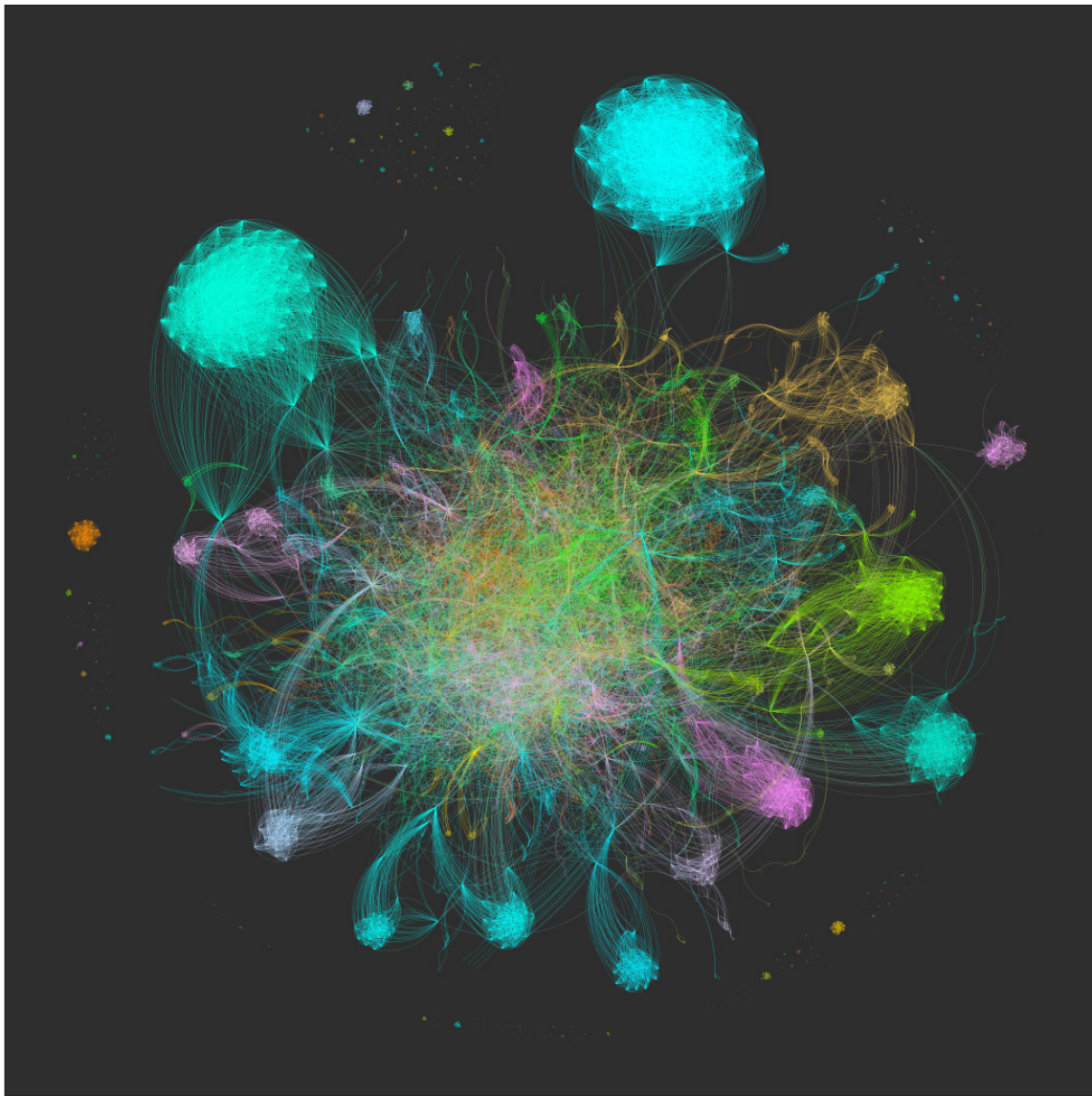
Figure 7 Visualisation of the actor-to-actor graph. The colours are based on the inferred community IDs. Graph designed with Gephi (Bastian, Heymann & Jacomy 2009).

To better understand the intricacies of the network, we employ various centrality measures. Figure 8 showcases scatter plots and linear regression lines, offering a clinical exploration of the distribution of nodes of the network's inherent structure. The network manifests a moderate closeness, emphasising closely linked nodes, especially at lower centrality levels (Figures 8b, 8d, 8f). Parsing the correlations between centrality measures provides deeper insights into the roles and relationships among network entities. The substantial positive correlation (Pearson's r of 0.715) between degree and betweenness centralities (Figure 8a) underscores the pivotal roles of highly connected nodes. Nodes with elevated degrees are crucial intermediaries, fostering efficient communication pathways and facilitating the seamless flow of information across disparate sections of the network. This interconnectedness is supported by the moderate relationship (Pearson's r of 0.268) between degree and eigenvector centrality (Figure 8c). Nodes with high-

degree centrality are likely, although not necessarily, to exhibit high eigenvector centrality, solidifying their status as influential 'hubs' adept at efficiently connecting with other significant nodes.

The predictive relationship (z-score of 33.20) observed in betweenness centralities (Figure 9c), along with a low standard deviation and an average of 0.00074, indicates a consistent pattern with a modest prevalence of intermediary nodes, underscoring the significance of specific nodes playing crucial roles. Beyond being well-connected, these nodes also tend to forge connections with other influential nodes, solidifying their central position in the network.

In real-world networks, nodes with high betweenness centrality can act as 'bottlenecks' for information or resources, playing a crucial role in maintaining efficient communication and coordination (Goh, Kahng & Kim 2003). The minimal correlation observed between betweenness and eigenvector centrality (Figure 8e) suggests that
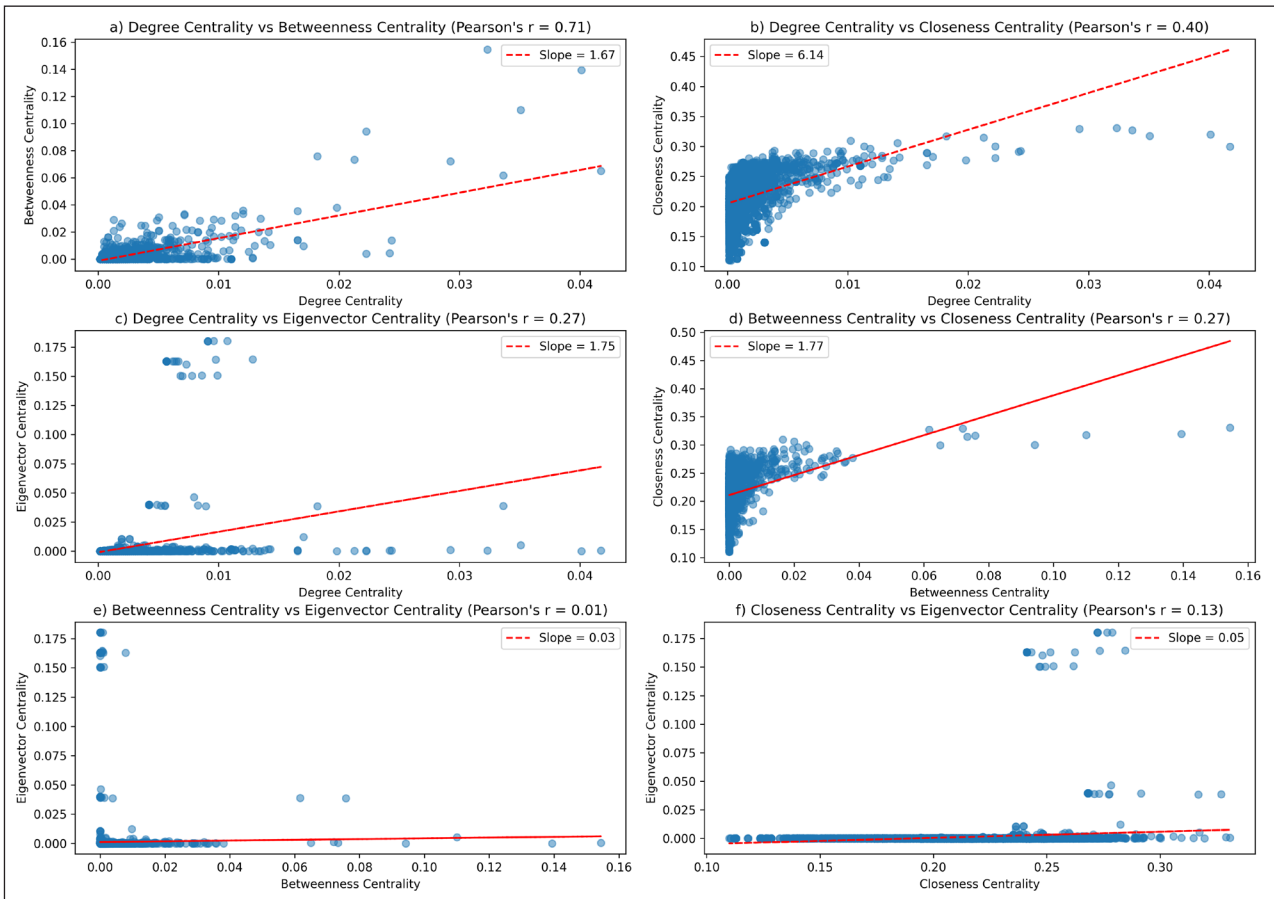
Figure 8 Scatter plots of centrality measures and linear regression for **(a)** degree and betweenness, **(b)** degree and closeness, **(c)** degree and eigenvector, **(d)** betweenness and closeness, **(e)** betweenness and eigenvector, **(f)** closeness and eigenvector.
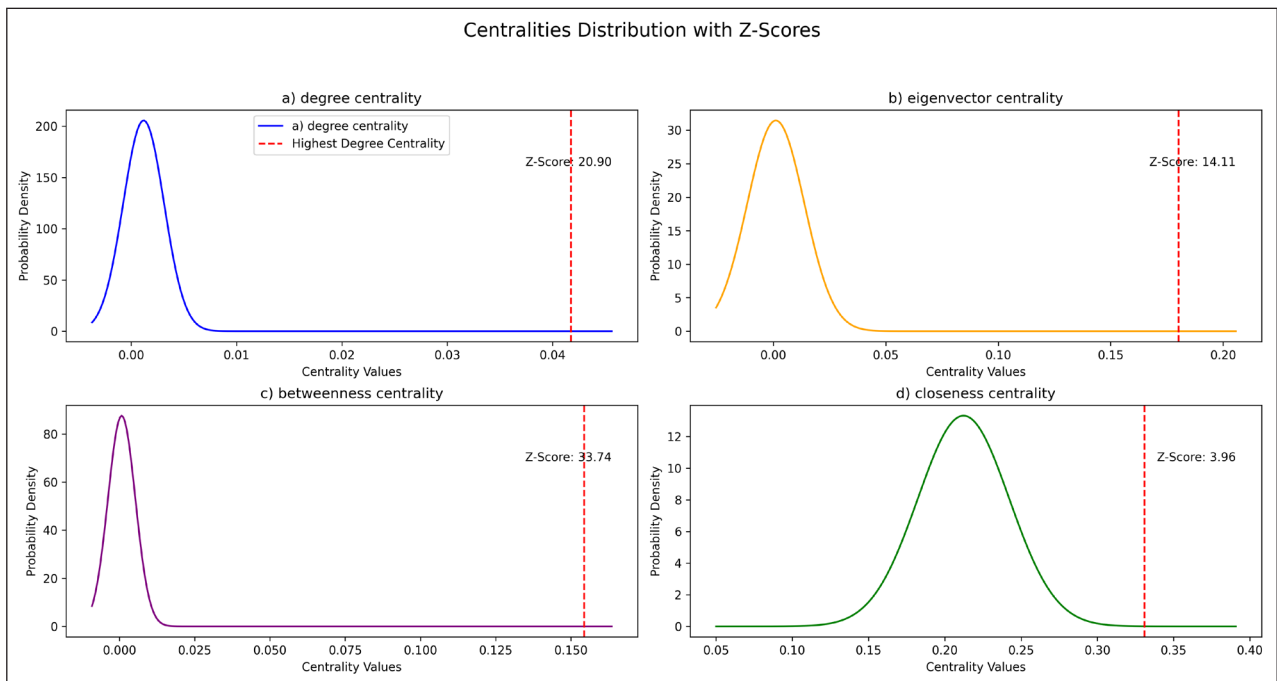
Figure 9 Gaussian plots of centralities values, with highest values in red and z-Score of each centrality: **(a)** degree centrality, **(b)** eigenvector centrality, **(c)** betweenness centrality, **(d)** closeness centrality.

nodes acting as intermediaries do not necessarily have a high influence on the broader network. Seemingly, the positive correlation (Pearson's r of 0.271) seen in Figure 8d implies that nodes facilitating communication pathways have reasonably efficient access to the broader network. The interplay of these centrality measures, as revealed by variations in Pearson's correlation coefficients across scatter plots, underscores the nuanced dynamics within the network. Positive correlations suggest a general trend of nodes with higher centrality in one measure exhibiting higher centrality in others. Importantly, these correlations extend beyond mere structural relationships, as evidenced by the scattered distribution of data points (Kanyou, Kouokam & Emvudu 2022).

Tables 5, 6, 7, and 8 provide a comprehensive overview of the top 50 nodes based on their centrality values. Additionally, Table 4 presents nodes that consistently appear between the top 50 centralities.

Revisiting previously discussed actors and introducing other well-known figures, Table 4 sheds light on the actors' dynamics. Robin Symes, whose 'role as a liaison for antiquities smuggling networks across the world spans multiple decades and jurisdictions' (Akers 2023), emerges as a key player. His centrality scores of 0.300 (closeness), 0.024 (betweenness), and 0.011 (degree) affirm his influential role and position as a bridge or intermediary within the network. The efficient flow of information or transactions reflected by Symes' central

presence within the network very well aligns with the global influence he once had as a dealer (Frammolino & Flech 2005). Fortuna Fine Arts, with a degree of 0.011, betweenness of 0.017, and closeness of 0.284, is another actor, well-known due to its recurrent associations with illicit antiquities cases (Albertson 2020b), that stands out: while its degree centrality suggests a moderate level of direct connections, the betweenness centrality indicates a role in facilitating the flow of information between other nodes. The closeness centrality suggests a relatively efficient connection to other nodes in the network, emphasising its importance in the overall structure. Recent legal developments, including charges related to provenance forgery and impersonation of deceased collectors (Rogers, Vere-Hodge & Dimmek 2021), emphasise the nature of its engagements.

Similar to Fortuna Fine Arts' in centrality scores is N(icolas) Koutoulakis: both entities exhibit a moderate degree centrality and a high closeness centrality, suggesting their involvement in connecting diverse entities, while in betweenness centrality N Koutoulakis has a significantly higher value, suggesting a more critical role in connecting different parts of the network and controlling the flow of information. Royal Athena (Galleries), highlighted for its high degree in the bipartite projection (see Figure 6) and its connection to illicit antiquities cases, maintains prominence with a substantial closeness centrality of 0.329, a high betweenness of 0.072 and one of the

LABEL	CLOSENESS	DEGREE	BETWEENNESS
Sotheby Parke-Bernet	0.331	0.032	0.154
Royal Athena	0.329	0.029	0.072
Parke-Bernet	0.327	0.034	0.062
Hˆotel Drouot	0.320	0.040	0.139
European private collection	0.317	0.035	0.110
Spink	0.317	0.018	0.076
Bonhams	0.315	0.021	0.073
The Devoted Classici…	0.309	0.010	0.016
Property from a New…	0.306	0.014	0.020
Charles Ede	0.300	0.022	0.094
Collection Mrs	0.300	0.042	0.065
Robin Symes	0.300	0.011	0.024
Ariadne	0.296	0.012	0.021
J-D Cahn	0.292	0.010	0.017
Merton D Simpson	0.288	0.009	0.023
N Koutoulakis	0.285	0.012	0.033
Fortuna Fine Arts	0.284	0.011	0.017
Mathias Komor	0.278	0.010	0.029

Table 4 Table comparing intersecting nodes between the top 50 centralities measures. In bold, nodes that are discussed in the paper.

NODE ID	SCORE	NAME
afa43...	0.042	Collection Mrs
64045...	0.040	Hotel Drouot
379f0...	0.035	European Private Collection
2947d...	0.034	Parke-Bernet
315bc...	0.032	Sotheby Parke-Bernet
7b56c...	0.029	Royal Athena
00cbd...	0.024	Marianne Dreesmann
5705d...	0.024	Anton Cr Dreesmann
74ade...	0.022	Collection Professor Dr Drs An...
e9d66...	0.022	Charles Ede
e2ce0...	0.021	Bonhams
6e833...	0.020	Edric Van Vredenburgh
f758b...	0.018	Spink
7d374...	0.017	2Nd Earl
428a9...	0.017	Abraham Rosman
4be3a...	0.017	Paula Rubel
e4041...	0.017	Marguerite Riordan
2bfc7...	0.014	Sangiorgi
578b5...	0.014	Property From A New York City...
2f2a4...	0.014	Anderson
5dd68...	0.013	Collection Privee Francaise
63089...	0.013	Rugby School
86f9d...	0.013	Collection Edric Van Vredenbur...
06a32...	0.013	Gavin Hamilton
c7a1b...	0.013	5Th Duke
925fc...	0.013	Mh Bloxam
cff18...	0.012	Wou Kiuan
1fa15...	0.012	N Koutoulakis
2f7be...	0.012	Ariadne
80093...	0.011	John
40cfd...	0.011	Fortuna Fine Arts
dbab2...	0.011	Robin Symes
6dd43...	0.011	Harewood
98164...	0.011	David Bromilow
60489...	0.011	George Spencer
e0a1f...	0.011	John Winston Spencer-Churchill
eda69...	0.011	Julia Harriet Mary Jary
81961...	0.011	Bitteswell Hall
0c4d0...	0.011	The Metropolitan Museum Of Art
3d9af...	0.011	Lawrence
62272...	0.010	American Art Association
40284...	0.010	The Devoted Classici The Priva...
37a47...	0.010	Brabourne
1459e...	0.010	Lansdowne House
45532...	0.010	Salomon Reinach
aa10f...	0.010	Vermeule
c2e41...	0.010	J-D Cahn
a2acb...	0.010	Mathias Komor
d148b...	0.009	Merton D Simpson
502a7...	0.009	Art Of The Ancient World

Table 5 Top 50 Highest Degree Nodes.

NODE ID	SCORE	NAME
315bc...	0.154	Sotheby Parke-Bernet
64045...	0.139	Hotel Drouot
379f0...	0.110	European Private Collection
e9d66...	0.094	Charles Ede
f758b...	0.076	Spink
e2ce0...	0.073	Bonhams
7b56c...	0.072	Royal Athena
afa43...	0.065	Collection Mrs
2947d...	0.062	Parke-Bernet
6e833...	0.038	Edric Van Vredenburgh
cff18...	0.036	Wou Kiuan
e4041...	0.036	Marguerite Riordan
1fa15...	0.033	N Koutoulakis
bb363...	0.033	Gunter Puhze
7a85d...	0.033	Property From A Manhattan Priv...
80093...	0.032	John
120da...	0.031	French Private Collection
5dd68...	0.030	Collection Privee Francaise
a2acb...	0.029	Mathias Komor
10b9b...	0.029	Mallett Son Ltd
1be38...	0.029	Collection Privee Americaine
399e6...	0.028	Plesch
37a47...	0.028	Brabourne
c51ee...	0.026	Heidi Vollmoeller
ba195...	0.025	Wilhem Horn
0e56f...	0.025	Property From A New York Priva...
49b25...	0.025	Jj Klejman
37fbd...	0.024	Steinitz
dbab2...	0.024	Robin Symes
d148b...	0.023	Merton D Simpson
0ae30...	0.022	Property From An Important Pri...
ae3eb...	0.021	E Borowski
2f7be...	0.021	Ariadne
2a5a2...	0.021	Alan Steele
814da...	0.021	Albert Museum
8b875...	0.021	Loo
77452...	0.021	Property From An American Priv...
63089...	0.020	Rugby School
bb92f...	0.020	Queen Victoria
578b5...	0.020	Property From A New York City ...
01ab3...	0.020	Eskenazi Ltd
66e07...	0.020	Property From An Important Eur...
08294...	0.019	Safani Gallery
a8489...	0.018	The Tuyet Nguyet
27185...	0.017	The Junkunc Collection
dbc68...	0.017	Galerie Flak
0c4d0...	0.017	The Metropolitan Museum Of Art
40cfd...	0.017	Fortuna Fine Arts
c2e41...	0.017	J-D Cahn
40284...	0.016	The Devoted Classici The Priva...

Table 6 Top 50 Highest Betweenness Centrality Nodes.

NODE ID	SCORE	NAME
315bc...	0.331	Sotheby Parke-Bernet
7b56c...	0.329	Royal Athena
2947d...	0.327	Parke-Bernet
64045...	0.320	Hotel Drouot
379f0...	0.317	European Private Collection
f758b...	0.317	Spink
e2ce0...	0.315	Bonhams
40284...	0.309	The Devoted Classici The Priva...
578b5...	0.306	Property From A New York City ...
e9d66...	0.300	Charles Ede
dbab2...	0.300	Robin Symes
afa43...	0.300	Collection Mrs
2f7be...	0.296	Ariadne
49b25...	0.293	Jj Klejman
66e07...	0.293	Property From An Important Eur...
00cbd...	0.293	Marianne Dreesmann
c2e41...	0.292	J-D Cahn
c33c4...	0.292	Hamilton
5705d...	0.291	Anton Cr Dreesmann
9b7d0...	0.290	Jay C Leff
5e5a6...	0.289	Drouot Richilieu
2bda0...	0.289	Rupert Wace
428a9...	0.289	Abraham Rosman
4be3a...	0.289	Paula Rubel
bb363...	0.289	Gunter Puhze
d148b...	0.288	Merton D Simpson
1fa15...	0.285	N Koutoulakis
502a7...	0.285	Art Of The Ancient World
06a32...	0.285	Gavin Hamilton
66cde...	0.284	Property From A North American...
40cfd...	0.284	Fortuna Fine Arts
e9b7e...	0.283	Ancient Sculpture Works Of Ar...
10773...	0.283	Arte Primitivo
7d374...	0.283	2Nd Earl
9a5c6...	0.282	Donna Jacobs Gallery
2bfc7...	0.281	Sangiorgi
74ade...	0.281	Collection Professor Dr Drs An...
120da...	0.281	French Private Collection
c7a1b...	0.281	5Th Duke
77452...	0.280	Property From An American Priv...
a817a...	0.280	Collection Louis Carre
741f3...	0.279	The Ernest Brummer Collection
0f1e8...	0.279	Property From The Harer Family...
817cf...	0.278	Beazley
dc9e0...	0.278	Shelburne
bc156...	0.278	Drouot Montaigne
a2acb...	0.278	Mathias Komor
b5262...	0.278	Comtesse Martine-Marie-Octavie
df6df...	0.278	Hubert De Ganay
98c6a...	0.278	Collection De Martine

Table 7 Top 50 Highest Closeness Centrality Nodes.

NODE ID	SCORE	NAME
aa10f...	0.180	Vermeule
a8566...	0.180	American Journal Of Archaeolog...
39645...	0.180	Adolf Michaelis
950f1...	0.180	Michaelis
06a32...	0.165	Gavin Hamilton
45532...	0.164	Salomon Reinach
ec288...	0.163	Archaologischer Anzeiger
345df...	0.163	Ann Arbor
b5697...	0.163	California Private Collection
c541b...	0.163	Georg Lippold
87aac...	0.163	Whittall Collection
7115d...	0.163	The Ancient Marbles
a93d7...	0.163	Greek Sculpture And Roman Tast...
fa5de...	0.163	Legendary Hollywood Decorator
cd934...	0.163	William Haines
66eed...	0.163	Classical Marbles
de379...	0.163	Elizabeth Angelicoussis
742ca...	0.163	Jimmie Shields
87cd6...	0.163	Comte De Clarac
aed9a...	0.163	Hermann Winnefeld
fd02a...	0.163	Pierre Gusman
5f1e1...	0.163	Walter Trillmich
b1507...	0.163	Madrider Mitteilungen
c92df...	0.163	Joachim Raeder
547d8...	0.163	Peter Schifando
61be2...	0.163	Jean H Mathison
6684d...	0.163	Sascha Kansteiner
e8244...	0.163	Gnomon
f480d...	0.163	Pantanello
ac171...	0.160	Arthur H Smith
a7bab...	0.151	William Petty
1459e...	0.151	Lansdowne House
cb7ed...	0.150	Hadrian
0afb5...	0.150	Henry Petty Fitzmaurice
1bd54...	0.150	Lansdowne
dc9e0...	0.046	Shelburne
87b42...	0.040	Douglas Hamilton
49835...	0.040	Archibald Hamilton
0edbb...	0.040	Hamilton Palace
bede3...	0.040	Alexander Hamilton
891d4...	0.040	Charles Townley
ac145...	0.040	Scot
f572e...	0.040	Gustav F Waagen
4f2cb...	0.040	International Studio Art Corpo...
016d3...	0.040	Private Collection, By Descent...
4468f...	0.040	Capitol
4a2ab...	0.040	Hamilton Archive
f1d40...	0.040	Singleton Abbey Collections
c33c4...	0.039	Hamilton
1bccc...	0.039	Brandon

Table 8 Top 50 Highest Eigenvector Centrality Nodes.

highest degree centralities (0.029). Royal Athena appears closely connected to other entities, facilitating efficient and direct interactions and playing a role of intermediary influencing the network dynamics. With one of the highest degree centralities, Royal Athena exhibits extensive direct connections, emphasizing its influential position. This suggests that Royal Athena is a central figure in terms of connecting with other entities and is involved with a notable number of direct relationships.

The eigenvector centrality offers unique insights into the structural dynamics of the network (Table 8). Unlike other centrality measures, eigenvector centrality places equal importance on the quality of connections, specifically those linked to other influential nodes. Among the top 50 nodes, a distinctive pattern emerges where individuals such as scholars (e.g., Salomon Reinach, Adolf Michaelis, and Gavin Hamilton) and entities not exclusively characterised as dealers but rather as figures intricately involved with artworks in other ways are featured. This observation reinforces the role of experts within the network, suggesting that their influence is not solely derived from transactional engagements but also from their connections with and knowledge of other influential nodes (Lawler 2005; Brodie 2009; Brodie 2011). Their significance extends beyond transactional networks, encompassing the broader and more nuanced web of interactions that shape the cultural heritage landscape.

5 DISCUSSION AND FUTURE WORK

The preliminary results of our work demonstrate the high potential of describing and investigating the antiquities market using AIKoGAM's methodology. Further analysis of ego networks of specific actors, encompassing all the possible layers that the KG offers, might lead to new knowledge and to transaction patterns previously unnoticed, as demonstrated by Graham et al. (2023).

While the potential of this application is evident, there are still several issues we need to address. General models like the one provided by SpaCy fail to identify certain entities correctly, leading to erroneous entities and relationships, especially with non-English statements. For example, sometimes the adopted model recognises entities such as 'Lyon' as actors in our network, instead of as locations. One possible solution is to manually tag a large dataset of provenance statements for NER and Relationship Extraction (RE) and fine-tune transformers models, such as RoBERTa for RE and DeBERTa-CRF for NER due to their performance on short texts (Wang et al. 2021), in order to make them domain-specific.

Another approach, as demonstrated by Graham, Yates, and El-Roby (2023), involves developing prompts for LLMs. Within our specific work, where we often encounter juxtaposed short phrases lacking evident contextual clues, this approach has varied success rates. Some

prompts adeptly extract entities and links, while others stumble when devoid of supplementary contextual data. A promising solution could involve the training or fine-tuning of a LLM using domain-specific literature.

As demonstrated, our research faced a challenge due to diverse raw data in terms of quality, language, and structure, specifically regarding inconsistencies in the representation of the names of individuals or organisations. With the developed scripts, we achieved significant improvements in data quality and accuracy by integrating automated scripts with manual inspection. This allowed us to detect and resolve duplicate occurrences, showcasing a deep understanding of the complexities arising from naming conventions and linguistic variations across various datasets. The application of advanced techniques along with flexibility in handling diverse spellings, played a crucial role in identifying nodes that likely represented the same 'actor' entity. Human-assisted fine-tuning enhanced the precision of the KG, while building a mapping dictionary for future automatising.

Looking ahead, the mapped dataset serves as a valuable resource for training and refining advanced NLP models, and opens avenues for more detailed and insightful analyses of the KG, enabling trend predictions based on enriched entity relationships.

The scripts can serve as templates for methodologies in similar projects dealing with diverse and heterogeneous data. They can be adapted to handle complex datasets where fully automated processes may encounter challenges. AIKoGAM modules are versatile and can be adjusted and expanded for similar projects, ensuring a balance between automated efficiency and human expertise.

The semi-automated data collection shows promise, but initial fine-tuning of each different website's HTML structure remains necessary. Research is ongoing on the feasibility of applying AI language models to HTML parsing and unsupervised information extraction (Gur et al. 2023; Zhang et al. 2023). Ideally, galleries and auction houses should adhere to precise, structured, and shared open-data ontologies to enable the creation of a real KG of circulating artefacts. This is definitely a path that could be pursued at the political and decision-making level.

A promising avenue of exploration would involve merging other similar datasets, as the one discussed in Fabiani and Marrone (2021), which encompasses data from external auction catalogues and extend over a wider time frame. Lastly the KG would benefit from the implementation of Computer Vision (CV) and embedding of visual data (Eyharabide, Bekkouch & Constantin 2021). By combining textual and visual data, similarity calculations, clustering and identification of the same objects/highly similar objects described differently and/or with misaligned provenance information would be possible with less human inspection. This implementation would

facilitate integration with external image-based databases (such as LEONARDO or INTERPOL Databases) as well.

5 CONCLUSIONS

AIKoGAM stands as a substantial stride in comprehending the antiquities market and combatting the illicit trafficking of cultural artefacts. Its innovation lies in the utilisation of different automated and semi-automated features to establish and analyse a domain-specific KG. While the initial outcomes offer promise, the genuine potential of this tool resides in its capacity to establish a framework for the subsequent implementation of advanced ML algorithms.

Adopting general-purpose NER models for automated information extraction, AIKoGAM has successfully collected data from various sources and built an interconnected KG, even though with some errors and inconsistencies that needed manual inspection. Further improvements in domain-specific models will be fundamental to refine the NER results and to accurately automatise the extraction of relationships.

A key aspect of AIKoGAM is the potential to foster cooperation and interoperability with related tools and initiatives. The depth and breadth of the KG, as well as the new knowledge that can be generated, would benefit from the integration of previous works described in Section 2, as well as other existing or future domain-specific databases. Standardised data formats and ontology mapping would enable seamless integration and cross-referencing of data from a variety of sources and external databases, and this is something that should involve market representatives.

The refined dataset becomes a valuable asset for a wide range of applications, from training machine learning models to constructing KGs, while the scripts contribute proven methodologies for addressing intricacies in naming conventions and linguistic variations.

In conclusion, AIKoGAM's methodology provides a foundation for future work in combating illicit trafficking of cultural property. As fast-paced advancements are made in NER, data integration, CV, and LOD, AIKoGAM has the potential to become a tool for researchers and LEAs, fostering a more transparent and secure cultural heritage landscape. Continued research and collaboration will be key in refining the methodology, and enabling the development of advanced big data models that can analyse, predict, and prevent the illegal trade of cultural heritage.

DATA ACCESSIBILITY STATEMENT

The scripts that support the findings of this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.8184155.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Riccardo Giovanelli orcid.org/0000-0001-8884-8492
Università Ca' Foscari Venezia, Italy; Istituto Italiano di Tecnologia - Center for Cultural Heritage Technology, Italy
Arianna Traviglia orcid.org/0000-0002-4508-1540
Istituto Italiano di Tecnologia – Center for Cultural Heritage Technology, Italy

REFERENCES

Ahmad, I, Akhtar, MU, Noor, S and **Shahnaz, A.** 2020. Missing Link Prediction using Common Neighbor and Centrality based Parameterized Algorithm. *Scientific Reports*, 10(364). DOI: https://doi.org/10.1038/s41598-019-57304-y

Akers, T. 2023. Manhattan District Attorney's Office returns two looted statues to Libya from the collection of disgraced former dealer Robin Symes. *The Art Newspaper*, 17 July [online access at https://www.theartnewspaper.com/2023/07/17/manhattan-district-attorneys-office-returns-looted-statues-robin-symes last accessed 18 July 2023].

Albertson, L. 2020a. Provenance Research Today. Principle, Practice, Problems. In: Tompkins, A (ed.), *From Soil to Showroom: Tracing Illicit Antiquities Across the World*. Lund Humphries, pp. 172–184.

Albertson, L. 2020b. Some background on Erdal Dere and Faisal Khan and Fortuna Fine Arts ltd., charged in the sd/new York. *Association for Research into Crime against Art*, September [online access at https://art-crime.blogspot.com/2020/09/some-background-on-erdal-dere-and.html last accessed 01 October 2023].

Barabási, A-L and **Pósfai, M.** 2016. *Network Science*. Cambridge University Press.

Bastian, M, Heymann, S and **Jacomy, M.** 2009. Gephi: an open source software for exploring and manipulating networks. *Proceedings of the International AAAI Conference on Web and Social Media*, 3(1): 361–362. DOI: https://doi.org/10.1609/icwsm.v3i1.13937

Bogdanos, M and **Iyer, A.** 2021. *In the Matter of A Grand Jury Investigation into a Private New York Antiquities Collector*, 06 December. Available at https://images.law.com/contrib/content/uploads/documents/292/102693/2021-12-06-Steinhardt-Complete-Agreement-w-Exhibits-Filed.pdf [Last accessed 16 June 2023].

Boyd, D and **Crawford, K.** 2011. Six Provocations for Big Data. In: *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011*.

Brennan, M and **Tsirogiannis, C.** 2022. *Illicit Antiquities Network*. Available at https://illicitantiquities.net/about [Last accessed 19 June 2023].

Brodie, N. 2009. Consensual relations? Academic involvement in the illegal trade in ancient manuscripts. In: Mackenzie, S and Green, P (eds.), *Criminology and archaeology: studies in looted antiquities*. London: Hart Publishing, pp. 41–59. DOI: https://doi.org/10.5040/9781472564931.ch-003

Brodie, N. 2011. Congenial bedfellows? The academy and the antiquities trade. *Journal of contemporary criminal justice*, 27(4): 408–437. DOI: https://doi.org/10.1177/1043986211418885

Brodie, N. 2015. The Internet Market in Antiquities. In: Desmarais, F (ed.), *Countering Illicit Traffic in Cultural Goods: the global challenge of protecting the world's heritage*. The International Council of Museums.

Brodie, N, Kersel, MM, Mackenzie, S, Sabrine, I, Smith, E and **Yates, D.** 2022. Why There is Still an Illicit Trade in Cultural Objects and What We Can Do About It. *Journal of Field Archaeology*, 47(2): 117–130. DOI: https://doi.org/10.1080/00934690.2021.1996979

Brughmans, T, Collar, A and **Coward, F.** (eds.) 2016. *The Connected Past: Challenges to Networks Studies in Archaeology and History*. Oxford: Oxford University Press. DOI: https://doi.org/10.1093/oso/9780198748519.003.0006

Brughmans, T and **Peeples, MA.** 2018. Network Science. In: Varela, SLL (ed.), *The Encyclopedia of Archaeological Sciences*. John Wiley and Sons, Inc. DOI: https://doi.org/10.1002/9781119188230.saseas0402

Brughmans, T and **Peeples, MA.** 2023. *Network Science in Archaeology*. Cambridge University Press. DOI: https://doi.org/10.1017/9781009170659

Campbell, PB. 2013. The Illicit Antiquities Trade as a Transnational Criminal Network: Characterizing and Anticipating Trafficking of Cultural Heritage. *International Journal of Cultural Property*, 20: 113–153. DOI: https://doi.org/10.1017/S0940739113000015

Campfens, E. 2020. Whose Cultural Objects? Introducing Heritage Title for Cross-Border Cultural Property Claims. *Neth Int Law Rev*, 67: 257–295. DOI: https://doi.org/10.1007/s40802-020-00174-3

CECOJI-CNRS-UMR6224 and European Commission, Directorate-General for Migration and Home Affairs. 2019. *Study on Preventing and Fighting Illicit Trafficking in Cultural Goods*. Pubblications Office. DOI: https://doi.org/10.2766/183649

Chandra, RV and **Varanasi, BS.** 2015. *Python requests essentials*. Packt Publishing Ltd.

Chappell, D and **Polk, K.** 2011. Unraveling the "Cordata": Just How Organized Is the International Traffic in Cultural Objects? In: Manacorda, S and Chappel, D (eds.), *Crime in the Art and Antiquities World*. New York: Springer New York, pp. 99–113. DOI: https://doi.org/10.1007/978-1-4419-7946-9_6

Chasing Aphrodite. 2012. The Getty List: 10 Objects At The J. Paul Getty Museum That Turkey Says Were Looted. *Chasing Aphrodite*, 20 April [online access at https://chasingaphrodite.com/2012/04/20/the-getty-list-10-objects-at-the-j-paul-getty-museum-that-turkey-says-were-looted/ last accessed 01 October 2023].

Christie's. 2012. *RELEASE: Christie's Announces First in a Series of Online-Exclusive Sales Optimized for a Global Collecting Audience*, 25 June. Christie's. Available at: https://www.christies.com/about-us/press-archive/details?PressReleaseID=5690 [Last accessed 10 July 2023].

Christie's. 2023a. *A Greek Bronze Zeus Hurling A Thunderbolt*. Christie's. Available at: https://www.christies.com/lot/lot-6436491 [Last accessed 10 July 2023].

Christie's. 2023b. *Antiquities*. Christie's. Available at https://www.christies.com/en/auction/antiquities-29962/ [Last accessed 10 July 2023].

Collar, A, Coward, F, Brughmans, T and **Mills, B.** 2015. Networks in archaeology: phenomena, abstraction, representation. *Journal of Archaeological Method and Theory*, 22: 1–32. DOI: https://doi.org/10.1007/s10816-014-9235-6

Comando Carabinieri Tutela Patrimonio Culturale. 2023a. *Attività Operativa 2022*.

Comando Carabinieri Tutela Patrimonio Culturale. 2023b. *Reperti Archeologic Confiscati*. Available at https://tpcweb.carabinieri.it/SitoPubblico/home/contenuti/beni-confiscati [Last accessed 19 June 2023].

CORDIS. 2022a. *Applying New solutions for Cultural Heritage protection by Innovative, Scientific, social and economic Engagement*. European Commission. Available at https://cordis.europa.eu/project/id/101094824 [Last accessed 19 June 2023].

CORDIS. 2022b. *Artwork Unique RecognitiOn and tRacking through chemicAl encoded data, miniaturized devices and blockchain alliance*. European Commission. Available at https://cordis.europa.eu/project/id/101094245 [Last accessed 19 June 2023].

CORDIS. 2022c. *ENDORSING SAFEGUARDING, PROTECTION, AND PROVENANCE MANAGEMENT OF CULTURAL HERITAGE*. European Commission. Available at https://cordis.europa.eu/project/id/101094237 [Last accessed 19 June 2023].

Costabello, L, Bernardi, A, Janik, A, Pai, S, Van, CL, McGrath, R, McCarthy, N and **Tabacof, P.** 2019. *AmpliGraph: a Library for Representation Learning on Knowledge Graphs*. Zenodo. DOI: https://doi.org/10.5281/zenodo.2595043

Cuno, JB. (ed.) 2012. *Whose culture? the promise of museums and the debate over antiquities*. Princeton: Princeton University Press.

D'Ippolito, M. 2014. New Methods of Mapping. The Application of Social Network Analysis to the Study of the Illegal Trade in Antiquities. In: Kennedy, GW, Agarwal, N and Yang, SJ (eds.), *Social Computing, Behavioral-Cultural Modeling and Prediction. SBP 2014. Lecture Notes in Computer Science*. Springer. DOI: https://doi.org/10.1007/978-3-319-05579-4_31

Davis, T. 2020. The Return of the Plundered: Case Studies in the Provenance of Illicit Antiquities. In: Tompkins, A (ed.) *Provenance research today: principle, practice, problems*. Lund Humphries: pp. 185–193.

Dijkstra, EW. 1959. A note on two problems in connexion with graphs. *Numerische mathematik*, 1: 269–271. DOI: https://doi.org/10.1007/BF01386390

Dörpinghaus, J, Klante, S, Christian, M, Meigen, C and **Düing, C.** 2022. From social networks to knowledge graphs: A plea for interdisciplinary approaches. *Social Sciences & Humanities Open*, 6(1): 100337. DOI: https://doi.org/10.1016/j.ssaho.2022.100337

eBay Inc. Staff. 2014. *Ebay and Sotheby's join forces on new live auction experience*, 14 July. Available at https://www.ebayinc.com/stories/news/ebay-and-sothebys-join-forces-new-live-auction-experiences/ [Last accessed 10 June 2023].

Ehrlinger, L and **Wöß, W.** 2016. Towards a Definition of Knowledge Graphs. *SEMANTiCS (Posters, Demos, SuCCESS)*.

EC–DGEAC, Brodie, N, Batura, O, Hoog, G, Slot, B, Wanrooij, N and **Yates, D.** 2019. *Illicit trade in cultural goods in Europe: characteristics, criminal justice responses and an analysis of the applicability of technologies in the combat against the trade: final report*. Publications Office. DOI: https://doi.org/10.2766/183649

EUROPOL. 2023. International art trafficking sting leads to 60 arrests and over 11 000 objects recovered. *EUROPOL*, 04 May [online access at https://www.europol.europa.eu/media-press/newsroom/news/international-art-trafficking-sting-leads-to-60-arrests-and-over-11-000-objects-recovered last accessed 19 June 2023].

Eyharabide, V, Bekkouch, I, Imad, EI and **Constantin, ND.** 2021. Knowledge Graph Embedding-Based Domain Adaption for Musical Instrument Recognition. *Computers*, 10(8). DOI: https://doi.org/10.3390/computers10080094

Fabiani, MD and **Marrone, JV.** 2021. Transiting Through the Antiquities Market. In: Oosterman, N and Yates, D (eds.), *Crime and Art*. Studies *in Art, Heritage, Law and the Market*. Springer International Publishing, pp. 11–28. DOI: https://doi.org/10.1007/978-3-030-84856-9_2

Fensel, D, Şimşek, U, Angele, K, Huaman, E, Kärle, E, Panasiuk, O, Toma, I, Umbrich, J and **Wahler, A.** 2020. *Knowledge Graphs. Methodology, Tools and Selected Use Cases*. Springer. DOI: https://doi.org/10.1007/978-3-030-37439-6

Fletcher, G. 2020. The Ethics of Provenance Research and the Art Market. In: Tompkins, A (ed.), *Provenance Research Today. Principle, Practice, Problems*. Lund Humphries, pp. 124–134.

Frammolino, R and **Flech, J.** 2005. Murky World of Antiquities Trade. *Los Angeles Times*, 28 December [online access at https://www.latimes.com/archives/la-xpm-2005-dec-28-me-dealers28-story.html last accessed 01 October 2023].

Freeman, LC. 1978. Centrality in social networks conceptual clarification. *Social Networks*, 1(3): 215–239. DOI: https://doi.org/10.1016/0378-8733(78)90021-7

Gerstenblith, PG. 2020. Provenience and Provenance Intersecting with International Law in the Market for Antiquities. *North Carolina Journal of International Law*, 45(2): 457–496. SSRN: https://ssrn.com/abstract=3573208

Gill, DWJ. 2019. Context matters: Nicolas Koutoulakis, the antiquities market and due diligence. *Journal of art crime*, 22: 71–78.

Gill, DWJ and **Tsirogiannis, C.** 2016. Polaroids from the Medici Dossier: Continued Sightings on the Market. In: Charney, N (ed.), *Art Crime: Terrorists, Tomb Raiders, Forgers and Thieves*. London: Palgrave Macmillan UK, pp. 229–239. DOI: https://doi.org/10.1007/978-1-137-40757-3_17

Giovanelli, R. 2018. Provenance non verificabili nel mercato di antichità romane: case study sui 300 oggetti di più alto valore in vendita in eBay US. *Archeomafie*, X: 115–135.

Giovanelli, R. 2021a. 150 oggetti trafficati in 30 anni di attività da Almagià potranno tornare in Italia. *The Journal of Culural Heritage Crime*, 21 December [online access at https://www.journalchc.com/2021/12/21/150-oggetti-trafficati-in-30-anni-di-attivita-da-almagia-potranno-tornare-in-italia/ last accessed 19 June 2023].

Giovanelli, R. 2021b. Michael Steinhardt: nessuna condanna per il magnate americano che acquistava antichità di dubbia provenienza. *The Journal of Cultural Heritage Crime*, 14 December [online access at https://www.journalchc.com/2021/12/14/michael-steinhardt-nessuna-condanna-per-il-magnate-americano-che-acquistava-antichita-di-dubbia-provenienza/ last accessed 19 June 2023].

Giovanelli, R. 2023. Operazione Pandora VII. 60 arresti e 11.000 opere d'arte rubate, recuperate in 14 paesi. *The Journal of Cultural Heritage Crime*, 04 May [online access at https://www.journalchc.com/2023/05/04/operazione-pandora-vii-60-arresti-e-11000-opere-darte-rubate-recuperate-in-14-paesi/ last accessed 19 June 2023].

Giovanelli, R and **Traviglia, A.** 2023. *riccardogvn/AIKoGAM: An AI-driven Knowledge Graph of the Antiquities market: toward automatised methods to identify illicit trafficking networks (v.1.1.2-alpha)*. Zenodo. DOI: https://doi.org/10.5281/zenodo.8241780

Graham, S, Yates, D and **El-Roby, A.** 2023. Investigating antiquities trafficking with generative pre-trained transformer (gpt)-3 enabled knowledge graphs: a case study [version 1; peer review: awaiting peer review]}. *Open Res Europe*, 3(100). DOI: https://doi.org/10.12688/openreseurope.16003.1

Graham, S, Yates, D, El-Roby, A, Brousseau, C, Ellens, J and **McDermott, C.** 2023. Relationship Prediction in a Knowledge Graph Embedding Model of the Illicit Antiquities Trade. *Advances in Archaeological Practice*, 11(2): 126–138. DOI: https://doi.org/10.1017/aap.2023.1

Granovetter, M. 1983. The strength of weak ties: A network theory revisited. *Sociological Theory*, 1: 201–233. DOI: https://doi.org/10.2307/202051

Granovetter, M. 1985. Economic Action and Social Structure: The Problem of Embeddedness. *The American Journal of Sociology*, 91(3): 481–510. DOI: https://doi.org/10.1086/228311

Goh, KI, Khang, EOB and **Kim, D.** 2003. Betweennes centrality correlation in social networks. *Physical Review E*, 67(1). DOI: https://doi.org/10.1103/PhysRevE.67.017101

Gujral, G and **Shivarama, J.** 2023. Knowledge Graphs. Connecting Information over the Semantic Web. In: Patel,

A and Debnath, NC (eds.), *Data Science with Semantic Technologies. Deployment and Exploration*. CRC Press, pp. 13–28. DOI: https://doi.org/10.1201/9781003310792-2

Gur, I, Nachum, O, Miao, Y, Safdari, M, Huang, A, Chowdhery, A, Narang, S, Fiedel, N and **Faust, A.** 2023. Understanding html with large language models. *ArXiv*, 2210.03945. DOI: https://doi.org/10.18653/v1/2023.findings-emnlp.185

Heafield, K, Kshirsagar, R and **Barona, S,** 2015. Language identification and modeling in specialized hardware. *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing.* Beijing, China: Association for Computational Linguistics, pp. 384–389. DOI: https://doi.org/10.3115/v1/P15-2063

Hixenbaugh, R. 2019. The Current State of the Antiquities Trade: An Art Dealer's Perspective. *International Journal of Cultural Property*, 26(3): 227–238. DOI: https://doi.org/10.1017/S0940739119000183

Hogan, A, Blomqvist, E, Cochez, M, D'amato, C, Melo, GD, Gutierrez, C, Kirrane, S, Gayo, JEL, Navigli, R, Neumaier, S, Ngomo, A-CN, Polleres, A, Rashid, SM, Rula, A, Schmelzeisen, L, Sequeda, J, Staab, S and **Zimmermann, A.** 2021. Knowledge graphs. *ACM computing surveys*, 54(4): 1–37. DOI: https://doi.org/10.1145/3447772

Huang, S. 2022. Relationship Extraction with GPT-3, *Medium*, 7 April [online access at https://medium.com/geekculture/relationship-extraction-with-gpt-3-bb019dcf41e5 last accessed 19 June 2023].

Isman, F. 2009. *I predatori dell'arte perduta. Il saccheggio dell'archeologia in Italia*. Skira.

Jaccard, P. 1912. The Distribution of the Flora in the Alpine Zone. *New Phytologist*, 11(2): 37–50. DOI: https://doi.org/10.1111/j.1469-8137.1912.tb05611.x

Jackson, MO. 2010. *Social and Economic Networks*. Princeton University Press. DOI: https://doi.org/10.2307/j.ctvcm4gh1

Kanyou, C, Kouokam, E and **Emvudu, Y.** 2022. Structural network analysis: Correlation between centrality measures. *CARI 2022*. Yaoundé, Dschang, Cameroon. Available at https://hal.science/hal-03714191 [Last accessed 16 December 2023].

Klie, J-C, Bugert, M, Boullosa, B, Eckart de Castilho, R and **Gurevych, I.** 2018. The INCEpTION platform: machine-assisted and knowledge-oriented interactive annotation. *Proceedings of the 27th international conference on computational linguistics: system demonstrations*. Santa Fe, New Mexico, pp. 5–9. Available at https://www.aclweb.org/anthology/C18-2002 [Last accessed 16 December 2023].

Krátký, M, Bača, R, Bednář, D, Walder, J, Dvorský, J and **Chovanec, P.** 2011. Index-based n-gram extraction from large document collections. *2011 sixth international conference on digital information management*, Melbourne, Australia: IEEE, pp. 73–78. DOI: https://doi.org/10.1109/ICDIM.2011.6093324

La Corte Suprema di Cassazione, Sezione II penale. 2011. *Sentenza N. 2870/204 R.G. N.8028/2010, 7 dicembre 2011, sui ricorsi proposti da Jacques Albert Henry e Medici Giacomo.*

Lawler, A. 2005. Looted tablets pose scholar's dilemma. *Science*, 309(5736): 869. DOI: https://doi.org/10.1126/science.309.5736.869

Levine, JA. 2009. The Importance of Provenance Documentation in the Market for Ancient Art and Artifacts: The Future of the Market May Depend on Documenting the Past. *DePaul J. Art, Tech. and Intell. Prop. L.,* 19(2). Available at https://via.library.depaul.edu/jatip/vol19/iss2/2 [Last accessed 19 June 2023].

Liew, CY, Labadin, J, Kok, WC and **Eze, MO.** 2023. A methodology framework for bipartite network modeling. *Applied Network Science*, 8(6): 2023. DOI: https://doi.org/10.1007/s41109-023-00533-y

Lü, L and **Zhou, T.** 2011. Link prediction in complex networks: A survey. *Physica A: Statistical Mechanics and its Applications*, 390(6): 1150–1170. DOI: https://doi.org/10.1016/j.physa.2010.11.027

Mackenzie, S. 2011. The Market as Criminal and Criminals in the Market: Reducing Opportunities for Organised Crime in the International Antiquities Market. In: Manacorda, S and Chappel, D (eds.), *Crime in the Art and Antiquities World*. New York: Springer New York, pp. 69–85. DOI: https://doi.org/10.1007/978-1-4419-7946-9_4

Mackenzie, S, Brodie, N, Yates, D and **Tsirogiannis, C.** (eds.) 2019. *Trafficking Culture. New Directions in Researching the Global Market in Illicit Antiquities: New Directions in Researching the Global Market in Illicit Antiquities*. Routledge. DOI: https://doi.org/10.4324/9781315532219

Mackenzie, S and **Davis, T.** 2014. Temple looting in Cambodia: Anatomy of a statue trafficking network. *British Journal of Criminology*, 45: 722–740. DOI: https://doi.org/10.26686/wgtn.12972851

Mackenzie, S and **Yates, D.** 2020. Researching the Structure of the Illicit Antiquities Trade. In: Tompkins, A (ed.) *Provenance Research Today. Principle, Practice, Problems*. Lund Humphries, pp. 162–171.

Marchi, E and **Miguel, O.** 1974. On the Structure of the Teaching-learning Interactive Process. *International Journal of Game Theory*, 3: 83–99. DOI: https://doi.org/10.1007/BF01766394

Marsden, PV. 2005. Recent Developments in Network Measurement. In: Carrington, PJ, Scott, J and Wasserman, S (eds.), *Structural Analysis in the Social Sciences*. Cambridge: Cambridge University Press, pp. 8–30. DOI: https://doi.org/10.1017/CBO9780511811395.002

Meirelles, I, Schich, M and **Malina, R.** 2014. Arts and Humanities, Complex Network Analysis of. In: Alhajii, R and Rokne, J (eds.), *Encyclopedia of Social Network Analysis and Mining.* Springer New York, pp. 56–60. DOI: https://doi.org/10.1007/978-1-4614-6170-8

Montani, I, Honnibal, M, Honnibal, M, Landeghem, SV, Boyd, A, Peters, H, Samsonov, M, Geovedi, J, Regan, J, Orosz, G, McCann, PO, Kristiansen, SL, Altinok, D, Roman, Fiedler, L, Howard, G, Phatthiyaphaibun, W, Bot, E, Bozek, S, Amery, M, Tamura, Y, Böing, B, Tippa, PK, Vogelsang, LU, Balakrishnan, R, Mazaev, V, Dubbin, G, Fukumaru,

J, Møllerhøj, JD and **Patel, A.** 2020. Explosion/spacy: v3.0.0rc: transformerbased pipelines, new training system, project templates, custom models, improved component api, type hints & lots more (v.v3.0.0rc1). *Zenodo*. DOI: https://doi.org/10.5281/zenodo.4091419

Neo4j. 2023a. *Neo4j cypher manual*. Neo4j. Available at https://neo4j.com/docs/cypher-manual/current/introduction/ [Last accessed 19 June 2023].

Neo4j. 2023b. *Neo4j desktop*. Neo4j. Available at https://neo4j.com/docs/ [Last accessed 19 June 2023].

Newman, MEJ. 2003. The Structure and Function of Complex Networks. *SIAM Rev*, 45(2): 167–256. DOI: https://doi.org/10.1137/S003614450342480

Nguyen, Q, Nguyen, N-K-K, Cassi, D and **Bellingeri, M.** 2021. New Betweenness Centrality Node Attack Strategies for Real-World Complex Weighted Networks. *Complexity*, 2021(1677445). DOI: https://doi.org/10.1155/2021/1677445

Oosterman, N. 2019. *Regional Overviews of the Policing of Art Crime in the European Union*. In: Hufnaget, S and Chappell, D (eds.), The Palgrave Handbook on Art Crime. London: Palgrave Macmillan, pp. 213–235. DOI: https://doi.org/10.1057/978-1-137-54405-6_10

OpenAI. 2023. *Gpt3.5*. OpenAi. Available at https://platform.openai.com/docs/models/gpt-3-5 [Last accessed 19 June 2023].

Palombo, C and **Yates, D.** 2023. Digital transit ports for the illicit trade in antiquities: the case of the 'Afghan Genizah'. *Digital Scholarship in the Humanities,* 38(1): 257–276. DOI: https://doi.org/10.1093/llc/fqac032

Patias, P and **Georgiadis, C.** 2023. Fighting Illicit Trafficking of Cultural Goods: The ENIGMA Project. *Remote Sensing*, 15(10). DOI: https://doi.org/10.3390/rs15102579

Pellegrini, M. 2023. Ritorno al futuro. *The Journal of Cultural Heritage Crime*, 01 January, [online access at https://www.journalchc.com/2023/02/01/ritorno-al-futuro/ last accessed 19 June 2023].

Pellegrini, M and **Rizzo, D.** 2021. The Italian Archaeological Heritage Abroad Between Agreements, Debates and Indifference. In: Traviglia, A, Milano, L, Tonghini, C and Giovanelli, R (eds.), *Stolen Heritage. Multidisciplinary Perspectives on Illicit Trafficking of Cultural Heritage in the EU and the MENA Region*. Venice: Edizioni Ca' Foscari. DOI: https://doi.org/10.30687/978-88-6969-517-9/005

Pellegrino, MA, Scarano, V and **Spagnuolo, C.** 2022. Move cultural heritage knowledge graphs in everyone's pocket. *Semantic Web*, 14(2): 323–359. DOI: https://doi.org/10.3233/SW-223117

Perry, BL, Pescosolido, BA and **Borgatti, SP.** 2018. *Egocentric Network Analysis: Foundations, Methods, and Models*. Cambridge University Press. DOI: https://doi.org/10.1017/9781316443255

Phoenix Ancient Art. 2023a. *Graeco Roman Bronze Statuette of Dionysus*. Phoenix Ancient Art. Available at https://phoenixancientart.com/work-of-art/graeco-roman-bronze-statuette-of-dyonisus/ [Last accessed 10 July 2023].

Phoenix Ancient Art. 2023b. *Works of Art*. Phoenix Ancient Art. Available at https://phoenixancientart.com/works-of-art/ [Last accessed 10 July 2023].

Ramasco, JJ and **Morris, SA.** 2006. Social inertia in collaboration networks. *Phisical reviewe*, 73. DOI: https://doi.org/10.1103/PhysRevE.73.016122

Richardson, L. 2019. *Beautiful Soup Documentation*, 24 December. Available at https://readthedocs.org/projects/beautiful-soup-4/downloads/pdf/latest/ [Last accessed 10 June 2023].

RITHMS. 2023. *Research, Intelligence and technology for heritage and market security*. European Commission. Available at https://cordis.europa.eu/project/id/101073932 [Last accessed 19 June 2023].

Rogers, F, Vere-Hodge, T and **Dimmek, A.** 2021. New York art dealers charged with fabricating provenance information and impersonating deceased collectors. *Art At Law*, 18 January. Available at https://www.artatlaw.com/new-york-art-dealers-charged-with-fabricating-provenance-information-and-impersonating-deceased-collectors [Last accessed 01 October 2023].

Roodt, C. 2016. Stolen Art Databases and Due Diligence. In: *6th International Conference of Experts on the Return of Cultural Property, Gyeongju-si, Republic of Korea, 17–19 Oct 2016*, pp. 215–225.

Rush, L and **Millington, LB.** 2015. *The Carabinieri Command for the Protection of Cultural Property: Saving the World's Heritage*. Heritage Matters. Boydell & Brewer. DOI: https://doi.org/10.1017/9781782045373

Ryan, YC and **Ahnert, SE.** 2021. The Measure of the Archive: The Robustness of Network Analysis in Early Modern Correspondence. *Journal of Cultural Analytics*, 6(3): 57–88. DOI: https://doi.org/10.22148/001c.25943

SeatGeek. 2023. TheFuzz. *GitHub*. Available at https://github.com/seatgeek/thefuzz [Last accessed 20 June 2023].

Singhal, A. 2012. *Introducing the Knowledge Graph: things, not strings*, 16 May. Available at https://blog.google/products/search/introducing-knowledge-graph-things-not/ [Last accessed 19 June 2023].

Smith, J. 2018. Toward "Big Data" in Museum Provenance. In: Schiuma, G and Carlucci, D (eds.), *Big Data in the Arts and Humanities. Theory and Practice*. CRC Press, pp. 41–50.

Sotheby's. 2019. *The untold story of Stephen Junkunc III, one of the great Chinese art collectors*, 20 August. Available at https://www.sothebys.com/en/articles/the-untold-story-of-stephen-junkunc-iii-one-of-the-great-chinese-art-collectors [Last accessed 19 June 2023].

The Value. 2018. Sotheby's Withdrew Chinese Buddha Statue Which Could Be Lost Relic From Longmen Grottoes. *The Value*, 16 September [online access at https://en.thevalue.com/articles/stephen-junkunc-iii-sothebys-buddha-head-longmen-caves last accessed 19 June 2023].

Trafficking Culture. 2023. *Encyclopedia*. Available at: https://traffickingculture.org/encyclopedia/ [Last Accessed 19 June 2023].

Tribunale di Roma. 2004. *Sentenza N. 40402/00 R.G. notizie di reato N.3553/01 R.G. G.I.P, 13 dicembre 2004, nei confronti di Medici Giacomo e Jacques Albert Henry.*

Trouillon, T, Welbl, J, Riedel, S, Gaussier, E and **Bouchard, G,** 2016. Complex embeddings for simple link prediction. In: Balcan, MF and Wienberger KQ (eds.), *Proceedings of the 33rd international conference on machine learning.* New York, USA: PMLR, pp. 2071–2080. Available at https:// proceedings.mlr.press/v48/trouillon16.html [Last accessed 01 June 2023].

Tsirogiannis, C. 2015a. 'Due Diligence'? Christie's antiquities auction, London, October 2015. *Journal of Art Crime*, 14: 27–37.

Tsirogiannis, C. 2015b. Mapping the supply: usual suspects and identified antiquities in 'reputable' auction houses in 2013. *Cuadernos de Prehistoria y Arqueología de la Universidad de Granada*, 25: 107–144.

Tsirogiannis, C. 2016. Something is Confidential in the State of Christie's. In: Charney, N (ed.), *Art Crime: Terrorists, Tomb Raiders, Forgers and Thieves*. London: Palgrave Macmillan UK, pp. 220–228. DOI: https://doi.org/10.1007/978-1-137-40757-3_17

Tsirogiannis, C. 2021. The antiquities market we deserve: 'Royal-Athena Galleries' (1942–2020). *Acta ad archaeologiam et artium historiam pertinentia*, 32: 147–175. DOI: https://doi.org/10.5617/acta.9024

Tsirogiannis, C and **Tsirogiannis, C.** 2016. Uncovering the Hidden Routes: Algorithms for Identifying Paths and Missing Links in Trade Networks. In: Brughmans, T, Collar, A and Coward, F (eds.), *The Connected Past: Challenges to Network Studies in Archaeology and History*. Oxford, pp. 103–120. DOI: https://doi.org/10.1093/oso/9780198748519.003.0012

U.S. Department of Commerce and National Institute of Standards and Technology. 2015. *Secure Hash Standard (SHS)*. FIPRS PUB 184–0. Federal Information Processing Standards Publication. DOI: https://doi.org/ 10.6028/NIST.FIPS.180-4

UNESCO. 2021. Follow-up to the Recommendations of the Working Group established by the Eighth Session of the Subsidiary Committee on the Public Statements of the Director-General. In: *Meeting of States Parties to the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, 6th, Paris, 2021*. UNESCO. Available at https:// unesdoc.unesco.org/ark:/48223/pf0000377431_eng [Last accessed 16 June 2023].

UNESCO. 2022a. Presentation of the procedures to be followed for the restitution of cultural property. In: *Subsidiary Committee of the Meeting of States Parties to the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, 10th, Paris, 2022*. UNESCO. Available at https:// unesdoc.unesco.org/ark:/48223/pf0000381462 [Last accessed 16 June 2023].

UNESCO. 2022b. Report on the results of the consultations with the art market on the revisions to the International Code of Ethics for Dealers in Cultural Property. In: *Subsidiary Committee of the Meeting of States Parties to the Convention on the Means of Prohibiting and Preventing the Illicit Import, Export and Transfer of Ownership of Cultural Property, 10th, Paris, 2022*. UNESCO. Available at https:// unesdoc.unesco.org/ark:/48223/pf0000381462 [Last accessed 16 June 2023].

Wang, Y, Zhao, Y, Schutte, D, Bian, J and **Zhang, R.** 2021. Deep learning models in detection of dietary supplement adverse event signals from Twitter. *Jamia open*, 4(4). DOI: https://doi.org/10.1093/jamiaopen/ooab081

Watson, P and **Todeschini, C.** 2007. *The Medici Conspiracy: The Illicit Journey of Looted Antiquities from Italy's Tomb Raiders to the World's Greates Museums*. BBS Public Affairs.

Wilkinson, MD, Dumontier, M, Aalbersberg, IJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, Silva Santos, LB. da, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, Gonzalez-Beltran, A, Gray, AJ, Groth, P, Goble, C, Grethe, JS, Heringa, J, Hoen, P, Hooft, R, Kuhn, T, Kok, R, Kok, J, Lusher, SJ, Martone, ME, Mons, A, Packer, AL, Persson, B, Serra, PR, Roos, M, Schaik, R. van, Sansone, SA, Schultes, E, Sengstag, T, Slater, T, Strawn, G, Swertz, MA, Thompson, M, Lei, J. van der, Mulligen, E. van, Velterop, J, Waagmeester, A, Wittenburg, P, Wolstencroft, K, Zhao, J and **Mons, B.** 2016. The fair guiding principles for scientific data management and stewardship. *Sci data*, 160018(3). DOI: https://doi.org/10.1038/sdata.2016.18

Williams, P and **Godson, R.** 2002. Anticipating organized and transnational crime. *Crime, Law & Social Change*, 37: 311–355. DOI: https://doi.org/10.1023/A:1016095317864

Yates, D. 2015. Value and doubt: the persuasive power of 'authenticity' in the antiquities market. *Parse*, 2: 71–84.

Yates, D. 2022. What auction catalogues analysis cannot tell us about the market. In: Fabiani, MD, Burmon, KM and Hufnagel, S (eds.), *Global Perspectives on Cultural Property Crime*. Routledge, pp. 43–60. DOI: https://doi.org/10.4324/9780367823801-6

Yogev, T and **Grund, T.** 2012. Network Dynamics and Market Structure: The Case of Art Fairs. *Sociological Focus*, 45(1): 23–40. DOI: https://doi.org/10.1080/00380237.2012.630846

Yu, S, Xu, C, Bai, X, Kuncheerathodi, R, Firmin, S and **Xia, F.** 2022. *Deep Learning Meets Knowledge Graphs: A Comprehensive Survey*. Research Square Platform. DOI: https://doi.org/10.21203/rs.3.rs-2021923/v1

Zhang, Z, Yu, B, Liu, T, Wang, Y and **Guo, L.** 2023. Learning structural co-occurrences for structured web data extraction in low-resource settings. *Proceedings of the acm web conference 2023*. Austin, TX, USA: Association for Computing Machinery, pp. 1683–1692. DOI: https://doi.org/10.1145/3543507.3583387

]u[ᘀ