heritage

MDPI

Article

From Provenance Statements to Antiquities Trafficking Networks: A Privacy-Aware Workflow Using Repatriation and OSINT Data

Michela Herbert [1], Katherine Davidson [2] and Pier Matteo Barone [1,*]

[1] Cultural Heritage Program, The American University of Rome, Via Pietro Roselli 4, 00153 Rome, Italy; m.herbert737@aur.edu
[2] Anthropology Program, Carleton University, 1125 Colonel By Dr, Ottawa, ON K1S 5B6, Canada; katherinesdavidson@cmail.carleton.ca
* Correspondence: p.barone@aur.edu

Abstract

It is difficult to capture the realities of the illicit antiquities market because of the lack of accessible, unsiloed data from underground trade networks. Despite existing literature on social network analyses and machine-learning experiments with antiquities data, there is a gap in simple open-source methodologies accessible to the non-academic public. By using a provenance-based analysis, we present a case study of the Italian antiquities trafficking networks that more fully captures their complexity. This study culls provenance data from repatriated antiquities gathered in the Museum of Looted Antiquities' dataset to create a network visualization for analysis. Using open-source provenance and repatriation data from 1950 to July 2025, we built a dataset of 233 repatriation events with 15.858 objects to produce a network that reveals central actors, roles, and locations while staying within ethical privacy limits. This study captures large portions of the trafficking network by using accessible data and produces a reproducible, ethically framed workflow.

Keywords: antiquities trafficking; network analysis; repatriation; provenance; open source; heritage crime

Check for updates

Academic Editor: George Pavlidis

Received: 14 February 2026
Revised: 22 March 2026
Accepted: 24 March 2026
Published: 25 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

There is a substantial body of literature on the Italian antiquities trafficking network [1,2]. However, the prevailing understanding of the network fails to fully capture its complexity, as analyzing underground trade poses challenges. A provenance-based network analysis built on open-source data from repatriated antiquities can help fill the gaps and flesh out this network, among others, in a novel methodology.

Possibly the most well-known visualization of the Italian antiquities is the 'organigram', a document allegedly written in the early 1990s by smuggler Pasquale Camera [3]. The organigram lays out Camera's conceptualization of the network in a hierarchical diagram, which was explored by Watson & Todeschini [1]. While the organigram covered a section of the broader network, it missed key actors and relationships. Yet this image formed a foundational understanding of the illicit antiquities trade, leading even Italian prosecutors to describe the network as a "hierarchical antiquities organization" [4] (p. 118). However, more recent scholarship has convincingly argued that the hierarchical view of the trade is inaccurate, changing contemporary understanding of the Italian trafficking network [4,5]. While there is still a lot of uncertainty and unavailable data due to the illicit

nature of the trade [5,6], Campbell [4] argues that its complexity is well-suited to network analysis (p. 114).

Social network analysis (SNA) enables a systematic, nuanced understanding of the complex social relationships and convoluted trade in antiquities trafficking networks [4,5,7]. Fabiani & Marrone [7] argue that "SNA methods can also make use of many types of data" (p. 12). This flexibility in data collection is helpful for the illicit trade in antiquities, as data can be hard to come by [5]. Tsiriogiannis & Tsiriogiannis [8] advocate for using the movement of goods to create a network, which Fabiani & Marrone [7] did by using auction catalogue text to visualize how objects changed hands. We build from these ideas, adding to the current body of scholarship, by positing network analysis in a novel way to capture the complexities of the Italian trafficking trade.

As Gerstenblith [9] argues, the loss of cultural heritage through the looting of antiquities harms society. Accessible archives on antiquities, provenance, and repatriation are of the utmost importance for this analysis, as they serve as both data and tools for social justice in the face of this harm. Tsirogiannis [3] has written about the importance of digitized archives in his research on the illicit antiquities market, which is contextualized by Moss et al. [10], who argue that digitizing archives transforms them into "data to be made sense of" (p. 120). This data, when accessible to the public online and not held behind borders, institutions, credentials, or other "contexts of power", has the potential to restore [11] (p. 332). Anderson [12] writes about how, like archaeology, archives can "preserve the context wherein objects and records are found and created to aid in their interpretation" (p. 761). Any archival evidence we have of trafficked and/or repatriated objects is powerful, as looted antiquities are robbed of archaeological context. Tracking their next steps after antiquities leave the ground can not only help researchers find their provenience but also identify the individuals responsible for their theft and alert the market to bad-faith players. This could help slow the trade in looted antiquities, giving justice to a country's patrimony. Thus, these archives, if they are accessible, can help "produce and reproduce justice [. . .] in [. . .] how they shape the past and engage the present" [11] (p. 321). Both the Central Registry's [13] Art Looting Intelligence Unit list and Tsirogiannis & Brennan's [14] Illicit Antiquities Network website are examples of accessible archival data in the illicit art trade. The former identifies 'red flag' actors in Nazi-looted art, and the latter makes archival provenance data and networks behind cases of antiquities trafficking publicly visible. Making archival information accessible can bring records to light that have concealed trafficking.

One accessible archive that helps combat data siloing and is built from open-source data is the Museum of Looted Antiquities (MOLA) [15]. The cases for this analysis have all been pulled from MOLA's dataset of looted and repatriated antiquities and join other studies investigating how open-source information can be used as a tool for cultural heritage analysis [16,17]. Leeson et al. [18] describe the existing cultural sector as rife with siloed data, creating "fragmented and incomplete resources" (p. 1). Therefore, datasets like MOLA's, which compile data on global repatriations, are highly valuable for "researchers, law enforcement, and cultural heritage professionals" to overcome these gaps and gather open-source intelligence (OSINT) in one place [18] (p. 3). There is a growing body of literature on the value of OSINT for cultural heritage, especially in times of conflict [19,20]. Gwardzińska-Chowaniec [19] argues that OSINT data collection "should constitute the foundation of an institutionalized system for the protection and management of cultural heritage", and that its growing popularity speaks to its applicability in "contemporary informational challenges" (p. 242). By using open-source data compiled in a digital archive to conduct SNA, this study contributes to the contemporary literature on digital solutions for network analyses aimed at regulating the antiquities market [21–23]. We propose

a workflow that leverages open-source provenance data from repatriated antiquities to capture substantial parts of the trafficking network, using accessible data and staying within privacy standards.

2. Materials and Methods

The following outlines the main steps in creating a privacy-aware methodology to capture trafficking networks using open-source data.

2.1. Data and Sources

All data for this analysis were sourced from the Museum of Looted Antiquities, which "[relies] upon open-source data: information from court records, media reports, academic studies and other sources available online in a variety of languages, all of which are cited as sources" [24]. These sources were collected by MOLA founder Jason Felch and a volunteer-based team of contributors and peer reviewers, and through web scraping conducted by Katherine Davidson, then MOLA project manager and data specialist. The web scraping was aimed at updating MOLA data to fill lacunae in data collection, notably between 2016–2021 and up to July 2025. This process used ScrapeGraphAI (v3.2.1), an LLM-assisted Python library that allows users to employ prompts to direct a web search and export the results as a spreadsheet [25]. The authors searched more than a dozen reliable journalistic sources for reporting on repatriation cases and compared the results with extant MOLA data. This resulted in several hundred additional cases being written by project staff, which were peer-reviewed (per MOLA's peer-review process) before being added to the dataset. Data was also obtained via Freedom of Information Law requests from the Manhattan D.A. and information from the Carabinieri TPC and other government entities. Because this data is limited to publicly disclosed repatriation cases, it does not claim to represent the entire illicit market. However, provenance information from repatriated items in an accessible dataset is a crucial source of data in an otherwise underground market.

All repatriations to Italy from 1950 to July 2025 were exported from MOLA into a spreadsheet, with each line representing a repatriation event. In total, we have data on 233 repatriation events involving 15,858 objects. For each event, the transfer of ownership was tracked, from the removal of the object to its return, with as much detail as was publicly accessible. All citations for the provenance of each case can be accessed in MOLA's public items or by requesting information from the MOLA team for yet-unpublished items. Because provenance reporting varies significantly across jurisdictions, institutions, and market sectors, the dataset should not be read as a uniform body of evidence. Rather, it is a structured aggregation of claims, disclosures, and archival traces that are documented in different ways. For that reason, provenance information was treated comparatively and critically, with preference given, where possible, to claims corroborated across multiple public sources such as court records, repatriation announcements, investigative reporting, museum disclosures, and archival documentation. Where provenance chains were incomplete, contradictory, or unverifiable, those gaps were preserved in the dataset rather than normalized away, so that uncertainty remained visible within the analytical workflow.

2.2. Edge Chain Construction

In constructing the chain of ownership, cases with no data were excluded (e.g., the Manhattan D.A. releasing a press release claiming that X number of antiquities were returned, with no information about what they are, where they were seized, or which traffickers they were associated with). However, the few cases that only had end-source data (such as the name of the gallery they were seized from) were included. This is an

honest reflection of the difficulty of capturing the market and still reveals the dealers, institutions, or collectors who acquired looted antiquities.

The finalized chain of ownership was manually converted into an edge chain, with each change of hands occupying one row, from source to target. These transactions were repeated based on the number of items they represent to weight the analysis accordingly. This means that if person A sold 40 antiquities to person B, the A source to B target is repeated in the spreadsheet 40 times. These weighted transactions reveal how many antiquities these participants touched. This can help contextualize massive seizures and properly demonstrate the scale of antiquities trafficking. However, weighting by item can also hide players who operated in the market for shorter periods or trafficked fewer, higher-value antiquities.

The edge chain was then uploaded to the open-source tool OpenRefine to trim the leading and trailing whitespace, fix name misspellings, and check duplicate listings. The refined edge chain was then anonymized according to the following convention:

- Listings that are **lettered** (i.e., Dealer—Basel A) are tied to specific people or galleries.
- Listings that have a **location but no letter** (i.e., Collector—London) mean that only their location is known.
- Listings with "**unknown**" (i.e., Unknown Trafficker) lack any identifying information.
- **Roles** are chosen (Dealer, Collector, Curator, etc.) based on individuals' most accurate market activities.
- **Galleries** are associated with specific dealers (if they have a primary dealer) to bring attention to the individuals participating in this trade, while **institutions**, such as auction houses and museums, are public and remain as-is.

We chose to keep the anonymization location-based to show geographic hubs of activity. The location was narrowed down to the city, if known, except for three cases that were so specific they would have revealed the individuals' identities. In these cases, we resorted to the county name or the nearest larger city. We referenced provenance records to research each anonymized individual's location, which allowed us to double-check the data. The new, anonymized edge chain was returned to OpenRefine to clean the data and check the naming conventions.

In practical terms, inclusion in the edge chain did not imply that every reported transfer was equally reliable. Instead, the chain records documented associations present in public provenance records, while acknowledging that some links are stronger than others. When sources conflicted, the more conservative interpretation was preferred, and when a transfer could not be attributed with sufficient clarity, it was either generalized through an "unknown" category or left incomplete rather than artificially resolved. This approach prioritizes transparency over false precision and is intended to reduce the risk of reproducing unsupported provenance claims as settled fact.

2.3. Network Construction

The anonymized, refined data were uploaded to Gephi (0.10.1), another open-source program for exploring and manipulating network visualizations [26]. We chose to input the edge chain in Gephi as both a directed and an undirected network. Because we are following provenance data, the directed graph allows us to track the paths that the items took and visualize the order of the actors' transactions [27]. However, working with the undirected network was also useful to create a simple "structure of acquaintances in the network" [8] (p. 142). After the edge chain, we uploaded a list of attributes to merge with it in Gephi. This attribute table assigned roles and locations to each node, allowing us to better visualize the network later.

Our directed network was run through the Yifan Hu Proportional algorithm, a force-based approach that pulls related nodes into the center and pushes unrelated nodes to the graph's outskirts. Yifan Hu Proportional spaces nodes appropriately, which we optimized by setting the Optimal Distance to 50 to keep the nodes (people, groups, and institutions) close for easy visualization [28]. Our undirected network was run through Fruchterman–Reingold, a standard algorithm for network visualization. This method provides a "generalized view" that's easy for viewers to interpret [28] (p. 94). After being run through their initial algorithms, both networks were then run through Noverlap to avoid node overlap and Label Adjust to avoid label overlap [26]. These algorithms assist in the data visualization and readability. As Hu [29] argues, networks only yield useful visualizations when they are "aesthetic" (p. 1).

The undirected graph ranked node size, node color, and text size by degree, to visualize "the amount of 'activity' a node experiences" [5] (p. 254). This helps us demonstrate the central, red-flag actors within this network. The degree range was limited to 2–46, where 46 being the upper limit and two cutting off minor actors. By focusing this undirected graph on degree centrality, we chose to focus on nodes' importance in the network [28].

Our directed graph was visualized in four ways—two focusing on in-degree and two on out-degree. Chervin [28] explains that in-degree centrality "tells us how likely other nodes are to seek out a single node" (p. 194). For a trafficking network, this reveals who the receivers of looted and stolen antiquities are. Node and label sizes were ranked by in-degree centrality, and the network was filtered to 2–23. The nodes' colors were partitioned by attributes, so our in-degree visualizations were run once based on the nodes' roles in the network [dealer, collector, museum, gallery, etc.] and once based on their general location [North America, Europe, Asia, etc.]. The out-degree graphs underwent a similar process. Out-degree centrality reveals "connections flowing from a [. . .] node to a range of other network members" [28] (p. 187). These graphs show who is sending material out, as indicated by the node and label sizes. We filtered the out-degree graphs to 2–46 and partitioned node colors by the same attributes—role and location. The workflow for this methodology is outlined below in Figure 1.

2.4. Ethics and Privacy Considerations

This project predominantly uses publicly available archival and provenance data from repatriation announcements, academic and news investigations, museum and auction records, and other open-source information. This report only publishes the names of people who have previously been made public by law enforcement, courts, academic publications, and investigative journalism. Additionally, the names of people published in gray literature, such as trafficking blogs, are present only if they are deceased. Every other actor has been anonymized for publication; this is consistent with ethics protocols in our academic institutions [30,31]. The working datasets have been shared only with supervisors and stored on password-protected drives and personal laptops. The data collected is all secondary information, and no direct contact with human subjects occurred.

This methodology is descriptive and intended to demonstrate how to develop a more accurate understanding of red flag actors within trafficking networks, helping researchers and institutions identify the alleged major players to avoid when dealing with suspect art. In publishing the results below, the authors do not allege criminal activity among the named network participants. Though some have justly been found guilty or penalized for their knowing participation in the illicit antiquities trade, legal standards for dealing in stolen goods often rely upon the participants' knowledge of wrongdoing [32,33]. This data comes from repatriated antiquities; therefore, there was seemingly sufficient evidence of their wrongful removal from the archaeological context or country of origin, or of illegal transfers

of ownership. However, just because there is evidence that something was looted or stolen, it does not mean that all participants knew and are therefore guilty. Because so many provenance statements are incomplete or vague for illicit and licit antiquities alike, "it is hard, if not impossible, to separate antiquities with a clean bill of health from those that have been illicitly-traded" [9,34] (p. 428). It is clear that there are many actors in the antiquities trade who use this obscurity to intentionally obfuscate their trade of looted archaeological material; there are actors who are happy to look the other way when a suspect provenance appears; and there are actors who believe they have kept their hands clean [9,33,35,36]. It is likely that our network has captured this range among its participants. Therefore, in the words of Graham et al. [23], these networks "could suggest [nodes] may have worked together at some point", but they are not proof of illegal action (p. 10). The methodology can support due diligence, investigations, restitution, and intelligence-led inquiries, but on its own, it is insufficient to establish criminal liability.

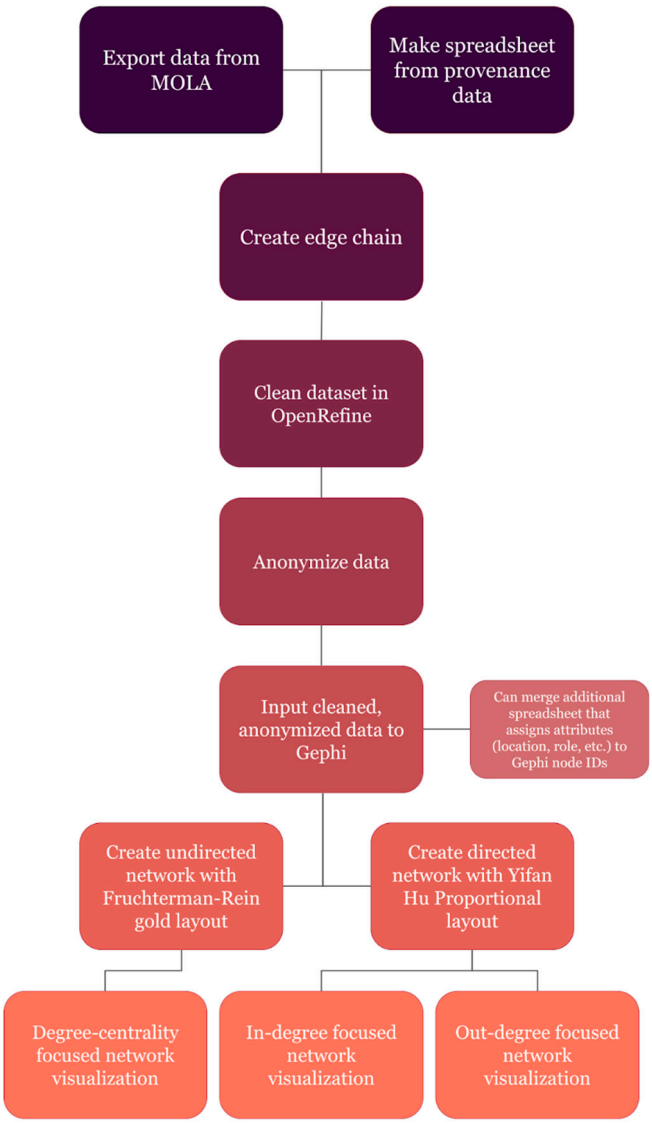


Figure 1. Simple provenance statement to trafficking network workflow.

3. Results

Before filtering, the undirected network has 220 nodes and 512 edges; after filtering, it has 173 nodes (78.64% visible) and 465 edges (90.82% visible). In the directed network, there are 220 nodes with 535 edges pre-filtering. For the filtered in-degree analysis, there

are 74 nodes (33.64% visible) with 276 edges (51.59% visible), and for the filtered out-degree analysis, there are 53 nodes (24.09% visible) and 221 edges (41.31% visible). Running basic statistics in Gephi reveals that the average degree in the undirected network is 4.655, indicating that participants typically have between 4 and 5 associates—the network's baseline connectivity. The average degree in the directed network is 2.432—the network's general level of directional flow. The statistics for the directed degree, in-degree, and out-degree measures are in Table 1.

Table 1. Basic measurements for the directed trafficking network, ranked by the top 10 participants for each statistic.

Degree Ranking		In-Degree Ranking		Out-Degree Ranking	
Degree	**ID**	**In-Degree**	**ID**	**Out-Degree**	**ID**
46	Giacomo Medici	23	Metropolitan Museum of Art	46	Tombaroli
46	Tombaroli	22	J. Paul Getty Museum	41	Giacomo Medici
43	Sotheby's	20	Sotheby's	32	Gianfranco Becchina
41	Gianfranco Becchina	20	Christie's	31	Unknown Trafficker
38	Robin Symes	15	Jerome Eisenberg	25	Robin Symes
34	Unknown Trafficker	14	Unknown Collector	24	Unknown Thief
32	Robert Hecht	13	Robin Symes	23	Sotheby's
27	Christie's	12	Museum of Fine Arts, Boston	22	Robert Hecht
26	Unknown Collector	11	Michael Steinhardt	13	Edoardo Almagia
25	J. Paul Getty Museum	10	Robert Hecht	12	Unknown Collector

For a trafficking network, a directed analysis is most useful, as the direction and flow of antiquities through the market are key to the analysis. Graph density measures "how tightly interconnected a network is", from 1—a completely connected network—to 0—a totally disconnected network [28] (p. 19). The directed network has a density of 0.011, meaning only 1% of nodes are connected. This reveals that there are narrow routes through which antiquities travel. Most actors are not trading with many others.

There are 7 weakly connected components (WCCs) and 144 strongly connected components (SCCs) in the directed network. The WCCs indicate that there are 7 separate nodal groups with no relationships among them [28]. These subgroups are like islands within the antiquities trade, revealing fragmentation that might be due to different dealer paths (Symes v Almagià), unknown origins (tombaroli vs. Unknown Thief), or disconnected one-off tourist thefts. The strongly connected components are 144 subgraphs "in which every pair of vertices can reach each other" [37] (p. 1). This is further proof of the network's fragmentation, as it suggests limited reciprocity and a high number of directional, one-way connections. This is not a highly surprising result, as the market generally moves from looter or thief to trafficker to dealer, and so on. When paired with the low density, this understanding of the SCCs suggests that the network is composed of many smaller clusters that trade within themselves but are disconnected from one another, rather than a tight-knit single operation

3.1. Undirected Network

The undirected network (Figure 2) highlights the nodes with the highest degree centrality, revealing major players in the Italian antiquities trafficking network. Because we structured the edge chain by duplicating each source-to-target transaction for each antiquity transferred, the more influential actors—that is, those who allegedly handled the most antiquities—are weighted and therefore highlighted in the network visualizations. Figures such as Giacomo Medici, Gianfranco Becchina, Robin Symes, and Robert Hecht are known for their influence within the network, as this analysis demonstrates. However, this can also help us identify which participants are central and therefore well-connected to major players. These actors might not have the same business longevity or mass-collecting power as the key traffickers and dealers, and have therefore escaped public scrutiny, but they are nonetheless central participants to watch. The nodes on the outskirts, which are blue according to this idea of a red flag list, are minor actors with less (or no) connection to the central actors.

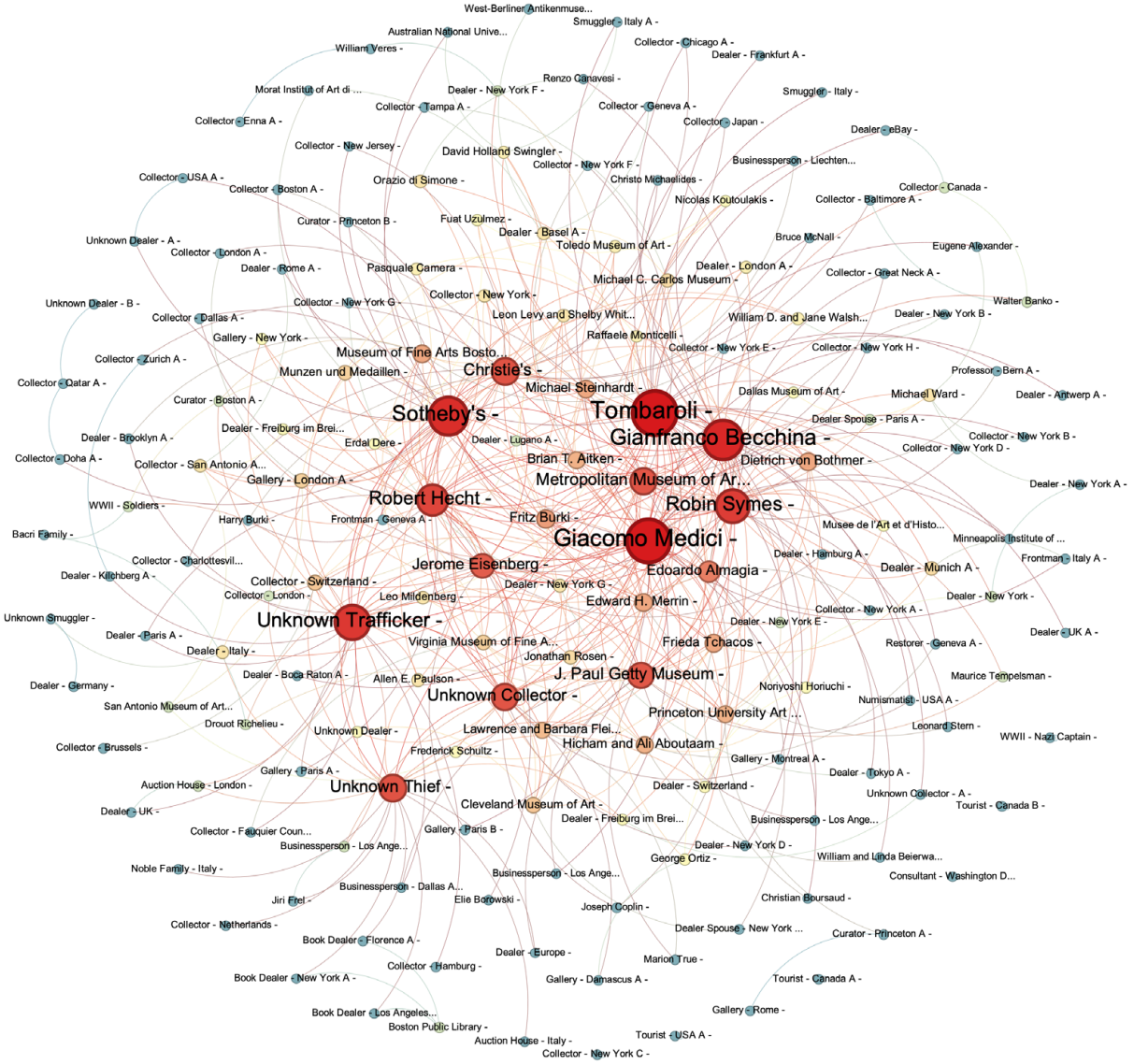


Figure 2. Undirected network in the Fruchterman–Reingold layout.

3.2. Directed Network

The directed networks below are separated by attribute to allow for easier comparison. The visualizations are color-coded by location and role, as seen below in Figure 3.



Figure 3. Key for node coloring by role and general location.

3.2.1. Directed Network by Location

Figures 4 and 5 are both directed networks partitioned into colors based on general location—the continents where the actors live and operate. Figure 4 shows in-degree centrality, with the largest nodes being the actors receiving the most connections from other nodes. Since we are analyzing the flow of goods, as represented by their provenance information, these are the nodes that are receiving the most goods. Figure 5 uses the same coloring to visualize out-degree, the nodes that have the most outward connections with other nodes. For the sake of our analysis, these represent the nodes that send or pass along the most antiquities. Comparing them, we can see that Figure 4's largest nodes are blue, while Figure 5's are red. This reveals how the goods are most often distributed outward from Europe, by actors such as tombaroli, Italian tomb looters, and dealers working in Switzerland and the UK, into North America, to be received by American museums and dealers.

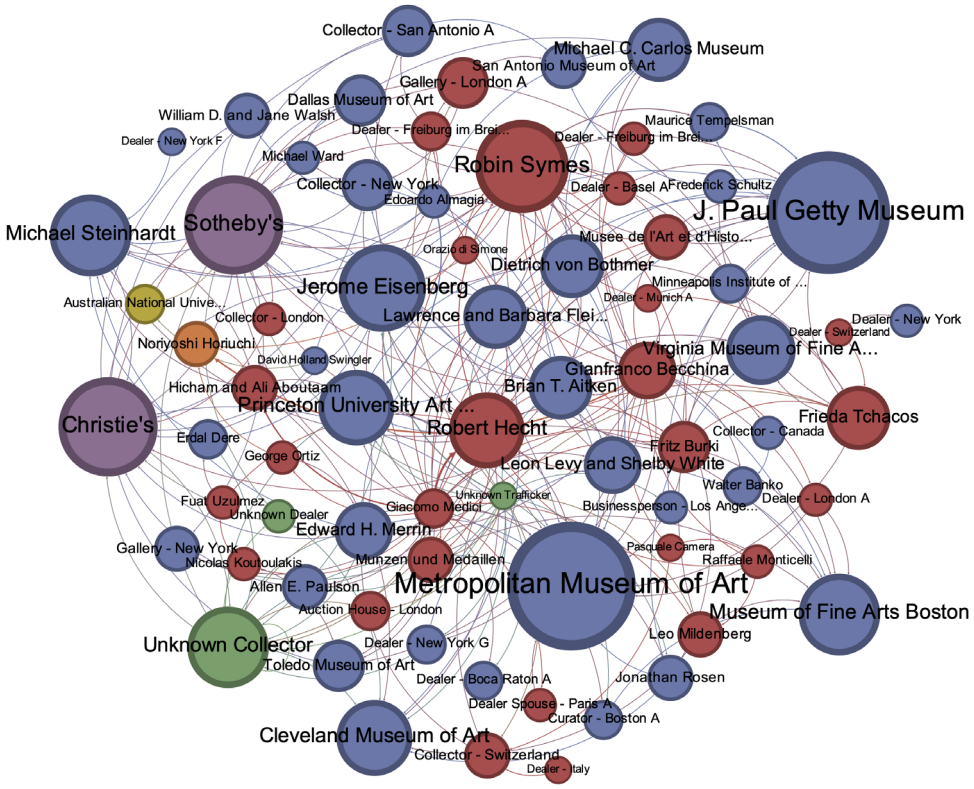


Figure 4. Directed graph modeling in-degree centrality, grouped by general location.

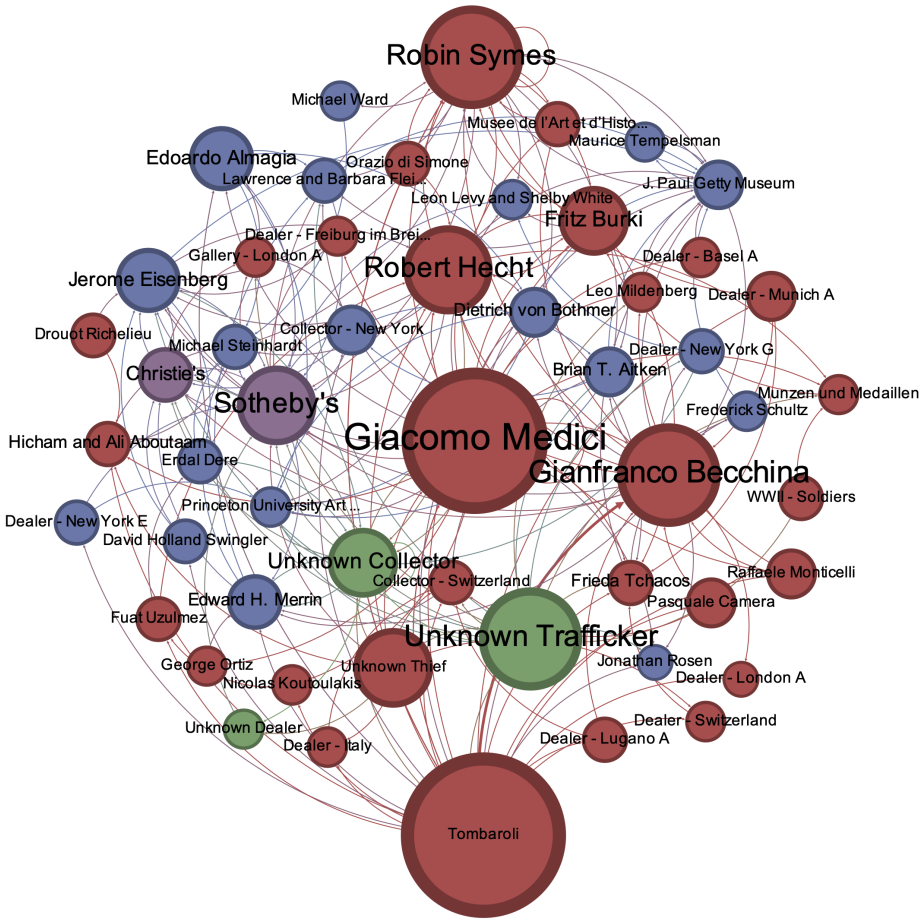


Figure 5. Directed graph depicting out-degree centrality, grouped by general location.

3.2.2. Directed Network by Role

The next set of directed networks is partitioned by the actors' roles. Their roles were chosen based on their primary activity in the art market. Like Figures 4 and 5, Figures 6 and 7 use the same in-degree and out-degree visualizations, but with different color partitions. In Figure 6, the nodes receiving the most goods, those with the highest in-degree centrality, are museums, dealers, collectors, and auction houses, while Figure 7's largest nodes, with the highest out-degree centrality, are the tombaroli, dealers, and traffickers. This role visualization helps support the idea that the network is not hierarchical. Not only does the graph's unstructured shape make that clear, but if the network were a strict, organized hierarchy, we would not see any museums in the out-degree graph (Figure 7), which we do (in light blue). Instead, there are institutions, like the J. Paul Getty Museum, and collectors who are both receiving and sending goods. They are not end-stage receivers but active trading participants in the market. However, this reflects documented provenance flows, not legal liability. We also see dealers filling different roles in the market, such as Medici being a primary distributor, but Symes, Hecht, and Jerome Eisenberg being primary receivers. This information, gathered from the visualization with role partitioning, helps us understand the network's actors and supports the claim that it is non-hierarchical.

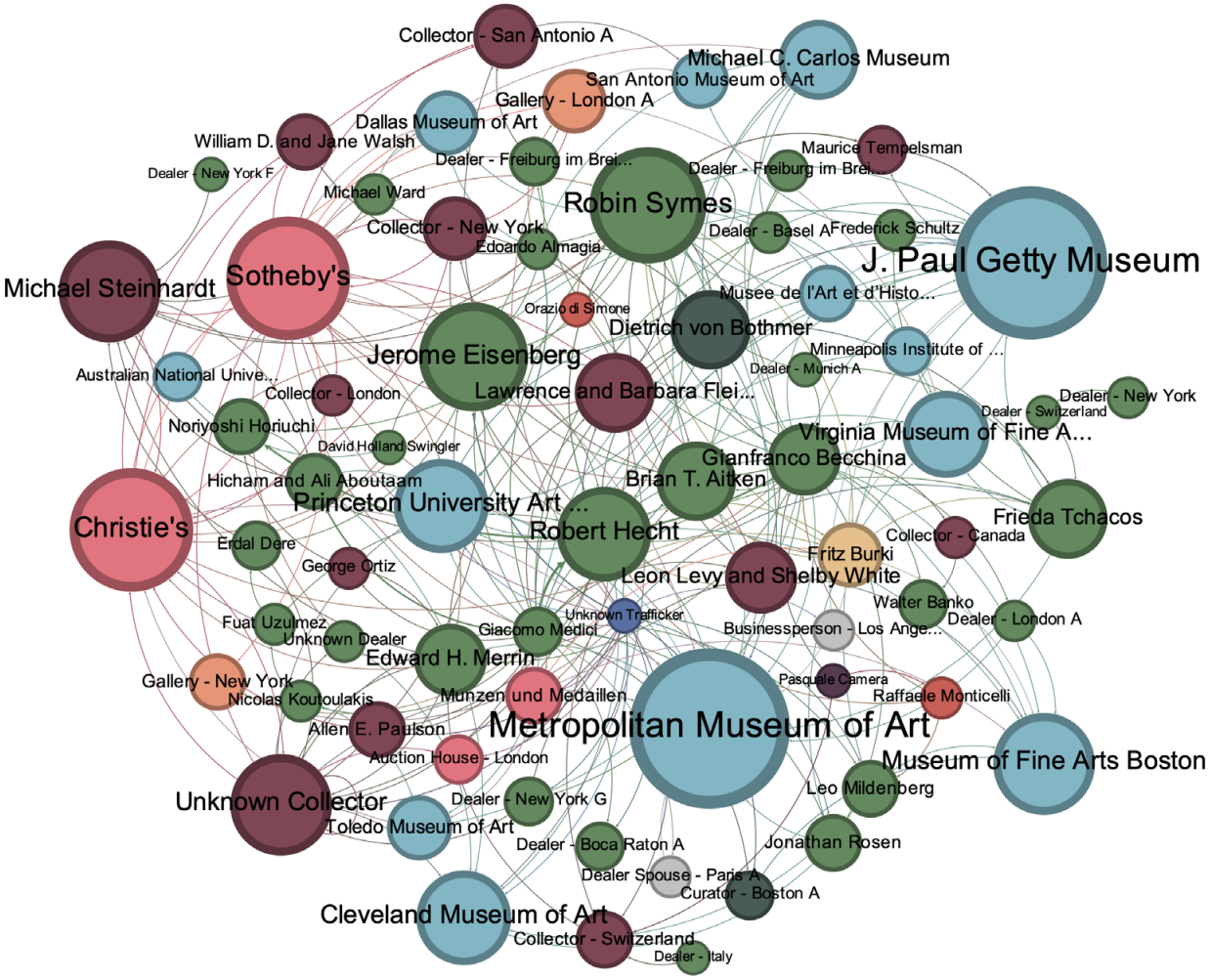


Figure 6. Directed graph showing in-degree centrality, grouped by market role.

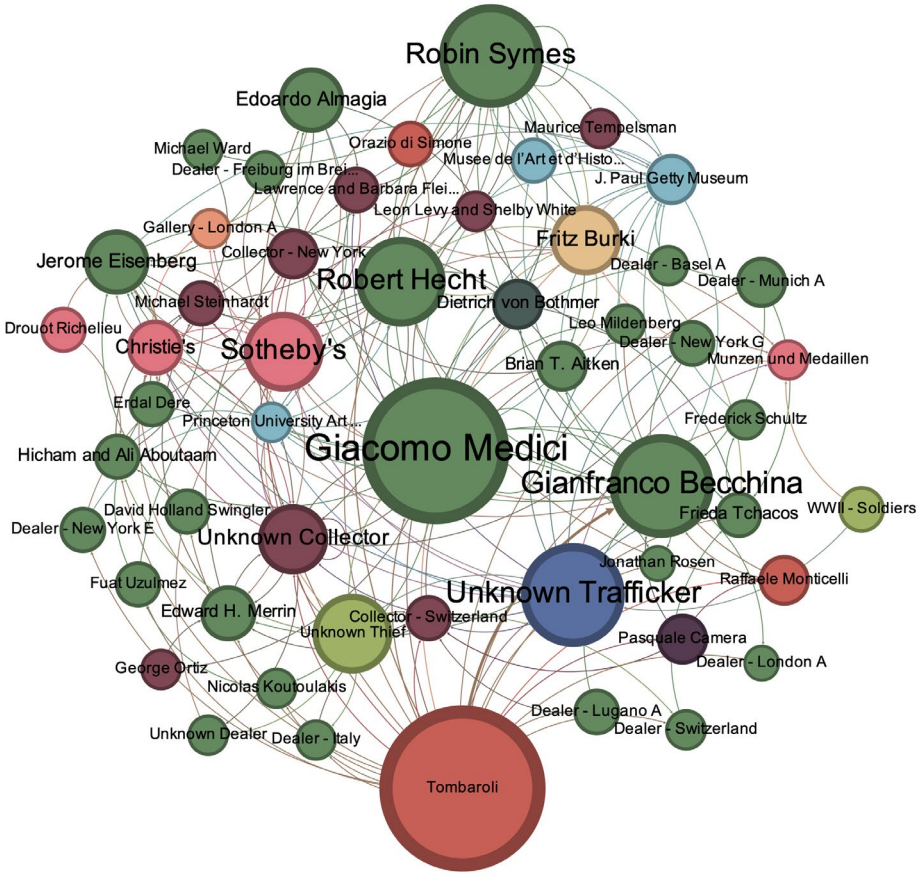


Figure 7. Directed graph depicting out-degree centrality, grouped by market role.

4. Discussion

This methodology demonstrates how open-source provenance information from repatriated goods can be used to construct a partial view of the illicit antiquities market. Using this approach offers opportunities for both the private and civil sectors to reduce the trade in looted antiquities.

As seen in Figure 1, making a simple, undirected graph ranked by degree centrality reveals the network's key actors. Interested parties can use repatriated antiquities' provenance data, from archives such as MOLA's or their own institution, to conduct this research and identify red-flag actors in the illicit trade. Avoiding provenance statements that include some of these network participants could help demonstrate the strength of future acquisitions for cultural market participants, such as dealers, museums, or collectors. Additionally, this information can reveal connections among actors and establish patterns of trade with convicted or alleged looters, traffickers, and dealers, providing critical evidence for investigations into suspect actors or supporting repatriation cases.

Furthermore, creating directional graphs partitioned by location helps display the broad movement of the antiquities trafficking trade and the locations of the primary distributors and acquirers of antiquities. This information can inform the need for border control measures or memoranda of understanding on heritage protection between countries. This potential is heightened because of the network's low density. It is not tightly connected; everyone is not connected to everyone. Therefore, there are a few key paths through which antiquities move. This means that those channels are particularly primed for influence. If they are disrupted, the network is disrupted.

Similarly, knowing the roles the major actors occupy in the market can help indicate which sectors could benefit from more oversight. For example, are auction houses or

antiquities dealers more important to regulate? A directional, role-partitioned graph can inform these decisions. Additionally, it is useful to compare actors' influence in the market between in-degree and out-degree. If market participants who are typically end-point receivers, such as collectors and museums, also have a sufficient presence in out-degree graphs, that can raise red flags. These can be early-stage participants who acquire from notorious dealers or traffickers, move the piece through their collection, and return it to the market—a possible way to create a clean provenance for a looted antiquity. Finding nodes with these roles in out-degree analyses can mark them as potential collaborators.

We believe that understanding red flag actors in trafficking networks, what roles they occupy, and where they operate from can be used to reduce the number of illicit antiquities on the market through relevant market, policy, enforcement, and legislative actions. This methodology, by itself, is not a mechanism for establishing criminal liability. Rather, its evidentiary value is best understood as corroborative and contextual: it can help identify recurring intermediaries, reconstruct documented chains of association, highlight inconsistencies in provenance narratives, and support risk-based inquiry in due diligence, investigation, and restitution contexts. In legal or restitution settings, such visualizations may therefore assist in organizing complex factual records and identifying patterns worthy of further scrutiny, but they should be read alongside documentary, testimonial, and jurisdiction-specific legal evidence rather than as standalone proof.

There is an opportunity for future research to apply this methodology to LLMs. An LLM can reduce the time spent manually crafting edge chains and chains of ownership and can be used as a tool for processes such as web scraping, as done for the MOLA dataset. There is a swell of current research on using machine learning for the "development of tools for automating provenance analysis and detecting risk in cultural heritage datasets" [16] (p. 19). Graham [38] experimented with using GPT3 to write an ontology for Giacomo Medici's network; Ferro et al. [16] used Natural Language Processing tools to enrich artifact provenance statements; and Graham et al. [23] worked with GPT-3 to create a knowledge graph on the antiquities trade. Our methodology closely aligns with these projects in both scope and aim. We chose to keep the methodology manual for this study to make it accessible for a range of researchers and market participants, but it could be used for LLM-facilitated workflows.

The applied value of this research is best suited for established trafficking networks. While contemporary OSINT analysis has been essential for protecting heritage in conflict zones, this specific methodology draws on repatriated antiquities, and a necessary passage of time is required that might not apply to transboundary trafficking from current conflict areas [18,19]. We created the network from provenance statements for repatriated antiquities. However, repatriating antiquities is not a priority during ongoing conflict, and therefore, drawing contemporaneous data from this methodology is limited. That said, it can be useful to create the network visualizations with other open-source provenance data to help uncover traffickers or dealers who acquire antiquities from conflict areas. It can also be used to analyze the movement of antiquities from past conflicts, like the trafficking of Iraq's looted cultural property, to keep an eye on red flag actors who may take advantage of current conflicts. In this sense, the methodology may be particularly useful for transboundary trafficking from conflict-affected territories when applied retrospectively or in a hybrid form. Although repatriation-based datasets are often delayed, the workflow can still help connect later restitution cases to earlier open-source reporting, thereby clarifying how objects, actors, and routes intersect across borders over time.

While the Italian illicit antiquities trade was chosen as an example, this methodology can be applied to other antiquities markets, given that there is adequate data on repatriations. Campbell [4] calls for continued network analysis in the antiquities market, as "the

network paradigm offers a nuanced view of cultural trafficking" (p. 138). The goal of this research is to make the methodology accessible to others so they can continue the necessary research on antiquities trafficking.

Limitations

This network was constructed with only the data that we could access. It does not claim to represent the entire Italian antiquities trafficking network. Because the network inherently functions underground, much of the data is hidden. We attempted to work around this by using only data from repatriated antiquities—items that are public and proven to have been illicitly excavated, exported, or traded. However, even with repatriated items, significant information is still missing. Some cases are still private to law enforcement agencies, some are held up in Freedom of Information Law requests, and some have never been published. Furthermore, some institutions have a more generous approach to repatriations, while others, specifically under UK statutory restrictions, have a strict no-return policy [39]. This can skew the data and impact our understanding of the network. Working with these limitations, we can view this network as a work in progress that will continually grow and be refined as new repatriations are shared.

Tsiriogiannis & Tsiriogiannis [8] explain that there are holes in these networks because there are "transactions that police investigations were not able to trace" or "specific people involved in these transactions [who] are unknown" (p. 140). We worked around these missing links as best we could. For example, if specific tombaroli were known, they were represented in their own nodes, but for the hundreds of cases where their names were not known or revealed, they joined the general "Tombaroli" node to represent their role and movement in the network. We took the same approach for "Unknown Collector" and "Unknown Dealer", among others. However, this means that there are nodes that are almost certainly double-represented. Meaning that Giacomo Medici has his own node but is likely also counted in the "Unknown Trafficker" or "Unknown Dealer" nodes. If the provenance information was clearly missing links or the actor's name was never published, we attempted to fill in the data by creating these general unknown links in the chain. While this means the data might not perfectly mirror the real-world network, it accurately captures how convoluted and private this network is.

Furthermore, some actors are likely underrepresented in the network. For example, there are New York collectors whose pieces were seized, but the exact numbers or their ownership histories were never reported. Therefore, we omitted them from our data. Some actors are underrepresented because they are closely tied to someone else. Robin Symes, for example, has a strong presence in the network, but his business and life partner, Christos Michaelides, is much smaller because it is nearly impossible to separate who was involved with which particular item. In these cases, we chose to prioritize the better-known name, except if the other was explicitly named in a provenance statement. Additionally, because the edge chain was weighted, the directed and undirected networks produce different degree rankings. We chose to prioritize directed rankings because directionality reveals structural information about the network that the undirected network's focus on reciprocity cannot capture.

Lastly, provenance itself can present vague and tenuous data. The ideal provenance goes back to the object's creation, but very few, if any, examples like this exist with antiquities [40]. What provenance information is known is not always publicly shared, is rarely complete, and is often inadequate. As Gerstenblith [40] explains, many provenance statements are made up of "stylized facts" that fail to show any objectively verifiable proof (p. 302). In addition to being unsupportable, provenance can be easily falsified [40]. Therefore, some of the provenance statements originally provided by network participants and

subsequently published may be false. This is another inherent risk in working with looted antiquities, but it should be acknowledged. By filtering the graphs to exclude some of the looser connections, these erroneous names (if they exist in our data) are hopefully cut from the final graph visualizations. The risk of including falsified or incomplete provenance data cannot be avoided entirely. However, we hope that the accurate data displayed in the network visualizations can establish connections between market participants, helping fill gaps or contextualize other provenances outside this data set.

A further limitation is that provenance quality is not only uneven but structurally shaped by legal, institutional, and commercial practices. Museums, dealers, auction houses, journalists, and law-enforcement agencies document ownership histories for different purposes and to different standards, meaning that apparently similar provenance statements may carry different evidentiary weight. Accordingly, the present methodology is best understood as a framework for organizing, comparing, and stress-testing published provenance claims rather than certifying them as true. Its value lies partly in exposing recurrent names, routes, and associations that merit deeper scrutiny, and partly in making visible where the record remains too thin, stylized, or contradictory to sustain stronger conclusions.

This is also where OSINT is both valuable and vulnerable. Open-source information can reveal patterns that would otherwise remain siloed, especially when different fragments are brought together across jurisdictions and institutional contexts. At the same time, OSINT inherits the blind spots, delays, and biases of what is made public, what is translated, what is digitized, and what is strategically disclosed. For that reason, this study treats OSINT-derived provenance not as self-validating evidence but as a research layer that becomes more robust when triangulated with independent documentary and legal sources.

5. Conclusions

This study presents a replicable, privacy-aware workflow for transforming dispersed provenance and repatriation information into interpretable visualizations of trafficking networks, using accessible, open-source tools and structured data practices. By treating repatriation cases as chains of custody and modeling those chains as networks, the approach shows how publicly available records—when cleaned, standardized, and analyzed consistently—can reveal recurring pathways, central actors, and role-specific dynamics that are difficult to identify through narrative description alone. The emphasis on open-source tooling and an explicitly ethical anonymization strategy is intended to make network-based insights more attainable not only to academic researchers but also to heritage professionals and market stakeholders who may not have access to specialized infrastructures.

Methodologically, the main contribution lies in converting repatriation events into a weighted edge chain that reflects transactional scale (e.g., repeating edges according to the number of objects transferred) and applying a transparent cleaning and anonymization protocol before visualization and analysis. In practice, the workflow proceeds through: (i) exporting relevant cases, (ii) reconstructing ownership transfers as a sequential chain, (iii) transforming transfers into a weighted edge list, (iv) cleaning and de-duplicating entity names and attributes, (v) anonymizing non-public individuals while retaining analytically meaningful features such as role and generalized location, and (vi) building complementary undirected and directed network views (including degree, in-degree, and out-degree) that can be partitioned by role and geography. The overall sequence is designed to be reproducible without advanced programming while still producing interpretable, defensible results from incomplete and unevenly reported data.

Substantively, the case study indicates that even partial, repatriation-driven information can yield meaningful structural signals. Low density and the presence of multiple weakly connected components suggest fragmented networks with narrow channels of

movement rather than a uniformly interconnected system. In turn, this fragmentation can imply "chokepoints," where a limited number of routes and intermediaries account for a disproportionate share of documented flows. The contrast between in-degree and out-degree patterns further illustrates how "receivers" and "distributors" can differ by sector and geography, and how certain actor categories may appear in both receiving and sending positions within the documented record. Crucially, these network patterns should be interpreted as indicators of association in published provenance flows, not as evidence of wrongdoing or criminal liability. Seen in this way, the methodology contributes less by proving individual culpability than by rendering opaque market structures more legible, comparable, and open to critical legal and provenance analysis.

The applied implications are threefold. First, for museums, auction houses, dealers, and collectors, network visualizations derived from repatriation and provenance records can serve as a practical supplement to due diligence by flagging recurring intermediaries and relationships that warrant deeper scrutiny when evaluating acquisitions or consignments. Second, for policy and enforcement stakeholders, geography- and role-partitioned directed graphs can help clarify where documented flows concentrate and where targeted resources—such as compliance audits, specialized investigative attention, or cross-border coordination—may have an outsized impact. Third, for researchers and civil society organizations, the workflow provides an extendable template that can be applied to other markets or categories of cultural goods where comparable reporting exists. Across these uses, the most responsible objective is preventative and risk-oriented: improving screening, transparency, and harm reduction rather than producing accusatory claims.

Ethics and privacy are integral to the design of this workflow rather than an optional add-on. The study relies on publicly available disclosures and limits the publication of identifiable names according to clearly stated criteria, anonymizing other actors while preserving analytical utility through role- and location-based labeling. Secure data-handling practices and explicit interpretive cautions are essential to ensure that network outputs support responsible heritage protection without creating undue reputational or personal harm.

Looking forward, several directions could strengthen both analytical power and responsible use. Future work could: (i) incorporate temporal slicing to test how network structures shift across decades and enforcement regimes; (ii) compare alternative weighting strategies (e.g., object counts versus case counts) and conduct sensitivity analyses to reduce the masking of short-duration but high-impact activity; (iii) integrate uncertainty or "data completeness" indicators so users can better distinguish well-documented pathways from sparse links; and (iv) explore carefully validated, human-in-the-loop automation for extracting and structuring provenance chains while minimizing error propagation. More broadly, scaling this approach across regions and datasets could support comparative research on market structure and regulatory outcomes—while maintaining the necessary balance between transparency, privacy, and methodological caution.

Author Contributions: Conceptualization, M.H. and K.D.; methodology, M.H. and K.D.; formal analysis, M.H.; investigation, M.H.; data curation, M.H. and K.D.; writing—original draft preparation, M.H. and P.M.B.; writing—review and editing, P.M.B. and K.D.; visualization, M.H.; supervision, P.M.B.; project administration, M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset is available on request from the authors due to privacy reasons.

Acknowledgments: The authors would like to thank Jason Felch (MOLA) for support in accessing and using MOLA data and his support of the research. The authors are grateful to Laura Prota (The American University of Rome) for her technical support with Gephi and network visualizations.

Conflicts of Interest: Michela Herbert and Katherine Davidson have been involved with the Museum of Looted Antiquities as volunteers and contractors.

References

1. Watson, P.; Todeschini, C. *The Medici Conspiracy: The Illicit Journey of Looted Antiquities from Italy's Tomb Raiders to the World's Greatest Museums*; PublicAffairs: New York, NY, USA, 2007.
2. Felch, J.; Frammolino, R. *Chasing Aphrodite: The Hunt for Looted Antiquities at the World's Richest Museum*; Houghton Mifflin Harcourt: Boston, MA, USA, 2011.
3. Tsirogiannis, C. Digitized Archives of Illicit Antiquities: Academic Research, Dissemination, and Impact. In *Shaping Archaeological Archives: Dialogues Between Fieldwork, Museum Collections, and Private Archives*; Raja, R., Ed.; Beprols: Turnhout, Belgium, 2023; pp. 409–418.
4. Campbell, P.B. The illicit antiquities trade as a transnational criminal network: Characterizing and anticipating trafficking of cultural heritage. *Int. J. Cult. Prop.* **2013**, *20*, 113–153. [CrossRef]
5. D'Ippolito, M. New methods of mapping: The application of social network analysis to the study of the illegal trade in antiquities. In *Social Computing, Behavioral-Cultural Modeling and Prediction: 7th International Conference, SBP 2014*; Kennedy, G.W., Agarwal, N., Yang, S.J., Eds.; Springer: Cham, Switzerland, 2014; pp. 253–260. [CrossRef]
6. Brodie, N.; Kersel, M.M.; Mackenzie, S.; Sabrine, I.; Smith, E.; Yates, D. Why there is still an illicit trade in cultural objects and what we can do about it. *J. Field Archaeol.* **2021**, *47*, 117–130. [CrossRef]
7. Fabiani, M.D.; Marrone, J.V. Transiting through the antiquities market. In *Crime and Art*; Oosterman, N., Yates, D., Eds.; Springer: Cham, Switzerland, 2021; Volume 1, pp. 11–28. [CrossRef]
8. Tsirogiannis, C.; Tsirogiannis, C. Uncovering the hidden routes: Algorithms for identifying paths and missing links in trade networks. In *The Connected Past: Challenges to Network Studies in Archaeology and History*; Brughmans, T., Collar, A., Coward, F., Eds.; Oxford University Press: Oxford, UK, 2016; Volume 1, pp. 139–160.
9. Gerstenblith, P. Controlling the International Market in Antiquities: Reducing the Harm, Preserving the Past. *Chic. J. Int. Law* **2007**, *8*, 10.
10. Moss, M.; Thomas, D.; Gollins, T. The reconfiguration of the archive as data to be mined. *Archivaria* **2018**, *86*, 118–151.
11. Duff, W.M.; Flinn, A.; Suurtamm, K.E.; Wallace, D.A. Social justice impact of archives: A preliminary investigation. *Arch. Sci.* **2013**, *13*, 317–348. [CrossRef]
12. Anderson, B.G. Kindred contexts: Archives, archaeology, and the concept of provenance. *Arch. Sci.* **2024**, *24*, 761–781. [CrossRef]
13. Central Registry. Post-War Reports: Art Looting Intelligence Unit (ALIU) Reports 1945–1946 and ALIU Red Flag Names List and Index. Available online: https://www.lootedart.com/MVI3RM469661 (accessed on 28 October 2025).
14. Tsirogiannis, C.; Brennan, M. Network. Illicit Antiquities Network. Available online: https://illicitantiquities.net/map (accessed on 28 October 2025).
15. Museum of Looted Antiquities. Available online: https://mola.omeka.net/ (accessed on 3 December 2025).
16. Ferro, S.; Giovanelli, R.; Leeson, M.; De Bernardin, M.; Traviglia, A. A novel NLP-driven approach for enriching artefact descriptions, provenance, and entities in cultural heritage. *Neural Comput. Appl.* **2025**, *37*, 21275–21296. [CrossRef]
17. Ghimpu, A.; Pareniuc, A. Use of information from open sources as research measures in investigation of illicit traffic of cultural goods. *Legea Viața 2025*, 390–400.
18. Leeson, M.; Giovanelli, R.; Ferro, S.; De Bernardin, M.; Traviglia, A. Overcoming data siloes in cultural heritage crime research: A consolidated OSINT derived dataset on art, antiquities, and the trade in cultural goods. *Arch. Sci.* **2025**, *25*, 16. [CrossRef]
19. Gwardzińska-Chowaniec, Ż. The Role of Open-Source Intelligence in Protecting Cultural Heritage During Armed Conflict: The Case of Ukraine. *Santander Art Cult. Law Rev.* **2025**, *11*, 223–244. [CrossRef]
20. CURIA Lab. Russian Destruction and Expropriation of Ukrainian Cultural Heritage. 2025. Available online: https://curialab.org/conflict-observatory/ (accessed on 22 March 2026).
21. Giovanelli, R.; De Bernardin, M.; Traviglia, A. Disentangle the net: How social network analysis can help understand and fight illicit trafficking in antiquities. In Proceedings of the Computer Applications in Archaeology Conference, Amsterdam, The Netherlands, 3–6 April 2023.
22. Giovanelli, R.; Traviglia, A. AIKoGAM: An AI-driven knowledge graph of the antiquities market: Toward automatised methods to identify illicit trafficking networks. *J. Comput. Appl. Archaeol.* **2024**, *7*, 92–114. [CrossRef]
23. Graham, S.; Yates, D.; El-Roby, A. Investigating antiquities trafficking with generative pre-trained transformer (GPT)-3 enabled knowledge graphs: A case study. *Open Res. Eur.* **2023**, *3*, 100. [CrossRef] [PubMed]
24. Felch, J. Welcome to the Museum of Looted Antiquities—Methodology. 2024. Available online: https://mola.omeka.net/exhibits/show/welcome/methodology (accessed on 4 February 2026).
25. Perini, M.; Padoan, L.; Vinciguerra, M. Scrapegraph-Ai, Version v0.0.10; Python: 2024. 2024. Available online: https://github.com/VinciGit00/Scrapegraph-ai (accessed on 4 February 2026).

26. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An open source software for exploring and manipulating networks. In Proceedings of the International AAAI Conference on Web and Social Media, San Jose, CA, USA, 17–20 May 2009; pp. 361–362. [CrossRef]
27. Cherven, K. *Network Graph Analysis and Visualization with Gephi*; Packt Publishing: Birmingham, UK, 2013.
28. Cherven, K. *Mastering Gephi Network Visualization*; Packt Publishing: Birmingham, UK, 2015.
29. Hu, Y. Efficient and high quality force-directed graph drawing. *Math. J.* **2006**, *10*, 37–71.
30. The American University of Rome. *MA in Cultural Heritage: Sustainability and Community Research Policy Guidelines on Research Ethics*; The American University of Rome: Rome, Italy, 2025.
31. Canadian Institutes of Health Research; Natural Sciences and Engineering Research Council of Canada; Social Sciences and Humanities Research Council of Canada. *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*; Government of Canada, Interagency Advisory Panel on Research Ethics: Ottawa, ON, Canada, 2022.
32. Snyder, C. Out of Context: Examining the Role of Context in Active Enforcement Foreign Patrimony Law Disputes. *Columbia Law Rev.* **2019**, *119*, 205–248. [CrossRef]
33. Mackenzie, S.R.M. *Going, Going, Gone: Regulating the Market in Illicit Antiquities*; Institute of Art & Law: Leicester, UK, 2005.
34. Brodie, N. Provenance and Price: Autoregulation of the Antiquities Market? *Eur. J. Crim. Policy Res.* **2014**, *20*, 427–444. [CrossRef]
35. Mackenzie, S.; Green, P. *Criminalising the Market in Illicit Antiquities: An Evaluation of the Dealing in Cultural Objects (Offences) Act 2003*; SSRN Scholarly Paper No. 1004267; Social Science Research Network: Rochester, NY, USA, 2007. [CrossRef]
36. Mackenzie, S.R.M. Dig a Bit Deeper: Law, Regulation and the Illicit Antiquities Market. *Br. J. Criminol.* **2005**, *45*, 249–268. [CrossRef]
37. He, H.; Xu, T.; Chen, J.; Cui, Y.; Song, J. A granulation strategy-based algorithm for computing strongly connected components in parallel. *Mathematics* **2024**, *12*, 1723. [CrossRef]
38. Graham, S. Mermaid Diagram to Ontology via GPT-3 for the Illicit Antiquities Trade. Available online: https://carleton.ca/xlab/2023/mermaid-diagram-to-ontology-via-gpt3-for-the-illicit-antiquities-trade/ (accessed on 28 October 2025).
39. Adams, G.K. Law Change Enabling Restitution on Moral Grounds Comes into Force This Month. Available online: https://www.museumsassociation.org/museums-journal/news/2025/11/law-change-enabling-restitution-on-moral-grounds-comes-into-force-this-month/ (accessed on 19 January 2026).
40. Gerstenblith, P. Provenances: Real, Fake, and Questionable. *Int. J. Cult. Prop.* **2019**, *26*, 285–304. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.