Relationship Prediction in a Knowledge Graph Embedding Model of the Illicit Antiquities Trade

Shawn Graham (ID)*, Donna Yates* (ID)*, Ahmed El-Roby, Chantal Brousseau* (ID)*, Jonah Ellens, and Callum McDermott*

ABSTRACT

The transnational networks of the illicit and illegal antiquities trade are hard to perceive. We suggest representing the trade as a knowledge graph with multiple kinds of relationships that can be transformed by a neural architecture into a "knowledge graph embedding model." The result is that the vectorization of the knowledge represented in the graph can be queried for missing "knowledge" of the trade by virtue of the various entities' proximity in the multidimensional embedding space. In this article, we build a knowledge graph about the antiquities trade using a semantic annotation tool, drawing on the series of articles in the Trafficking Culture Project's online encyclopedia. We then use the AmpliGraph package, a series of tools for supervised machine learning (Costabello et al. 2019) to turn the graph into a knowledge graph embedding model. We query the model to predict new hypotheses and to cluster actors in the trade. The model suggests connections between actors and institutions hitherto unsuspected and not otherwise present in the original knowledge graph. This approach could hold enormous potential for illuminating the hidden corners of the illicit antiquities trade. The same method could be applied to other kinds of archaeological knowledge.

Keywords: knowledge graph, embedding models, illicit antiquities trade, link prediction, networks

Las redes transnacionales del comercio ilícito e ilegal de bienes culturales son difíciles de comprender. Sugerimos representar el comercio como un gráfico de conocimiento con múltiples tipos de relaciones. Esta representación puede transformarse via una arquitectura neuronal en un grafo de conocimiento "embedded model". El resultado es que la vectorización del conocimiento representado en el grafo puede consultarse en busca de las conexiones que faltan en el comercio en debido a la proximidad de las distintas entidades en el espacio de incrustación. En este artículo, construimos un grafo de conocimiento sobre el comercio de bienes culturales utilizando una herramienta de anotación semántica, basándonos en las series de artículos del Trafficking Culture Encyclopedia. A continuación, utilizamos AmpliGraph, una serie de herramientas para el aprendizaje automático supervisado (Costabello et al. 2019) para convertir el gráfico en un grafo de conocimiento "embedded model". Consultamos el modelo para predecir nuevas hipótesis y agrupar a los actores en el comercio. El modelo sugiere conexiones entre actores e instituciones hasta ahora insospechadas y no presentes en el gráfico de conocimiento original. Este enfoque podría tener un enorme potencial para iluminar los esquinas oscuras del comercio ilícito de bienes culturales. El mismo método podría aplicarse a otros tipos de conocimiento arqueológico.

Palabras clave: grafo de conocimiento, embedding models, comercio ilícito d bienes culturales, link predicción, redes

KNOWLEDGE GRAPH EMBEDDING MODELS AND THE ORGANIGRAM

In 1995, a hand-drawn organizational chart depicting the network of dealers, intermediaries, and looters in Italy's illegal antiquities trade was seized by the Carabinieri, Italy's national military police force. This "organigram" depicted two interconnected but broadly independent "cordata"—or "the people roped together"—showing the networked structure of the antiquities trade in Italy at the time (Brodie 2012a). From the late 1960s until their respective convictions for antiquities-related crimes in 2005 and 2011, Giacomo Medici and Gianfranco Becchina headed parallel "cordatas" that supplied the world art market with looted and trafficked Italian antiquities (Watson and Todeschini 2007). It is their supply networks in

particular that were depicted on the original organigram. Research has shown that the antiquities trade (hereafter referred to as "the trade") is built similarly on personal relationships; what gets traded or purchased is often a function of building trust in anticipation of better materials to come (Oosterman et al. 2021).[1]

In this article, we transform what we *do* know about the historical contours of the illicit antiquities trade into an embedding model (see below), a kind of machine-learning representation, which enables us to make predictions about what we do not *yet* know. We draw our data from the "Encyclopedia" at the Trafficking Culture Project website (as it stood in May 2022). The encyclopedia reflects the research interests of the members of the Trafficking Culture Project, so is not an exhaustive "last word" on the subject of the illicit antiquities trade but rather a bounded body of knowledge. An immediate and fair question might be "Why"? And furthermore, who is this approach for? What information does it offer us that we could not obtain by other means? What does this approach "solve"?

Recent high-level discussions (and funding prioritizations) related to attempts to disrupt the illicit trade in antiquities have focused on the development of "digital tools" and other "tech solutions" to this form of crime—for example, the European Union's recently implemented Horizon Europe funding scheme offering multiple millions of euros for development in this field. Prior major funding for related tech-based "solutions" have had ambiguous results, ranging from limited proofs of concept to social platforms that no one uses. One recent European Commission report (Brodie et al. 2019), coauthored by one of the authors of this article, assessed the general situation as "technologies in search of an application" (Brodie et al. 2019:187) and generally disparaged the lack of attention being paid to researcher and practitioner needs before money is spent.

What we discuss in this article speaks directly to an identified researcher and practitioner need in the field of antiquities trafficking research. Experts in this field hold a vast and varied amount of qualitative knowledge about thousands of individual cases of antiquities-related crime, and research into these and new cases follows a series of patterns based on prior experience. Researchers look for continuations of patterns they have already detected or expect, follow established pathways for question posing and evidence gathering, and ultimately create a locally effective but limiting box for themselves. It is incredibly difficult for researchers and investigators to step outside of this box—what digital humanist Matthew Lincoln (2015, developed further in Lincoln 2017) calls the problem of "confabulation"—to set aside what they believe they already know and to develop new but plausible and even important leads to investigate. To put it another way, researchers know they are missing something from their understanding of antiquities trafficking networks, but they do not know what it is, nor do they have the ability look at everything with fresh eyes. This has been not only our own experience in our decades of working in this field but also a sentiment expressed to us by fellow academics as well as investigators within police and public authorities.

Consequently, we offer this piece in that spirit, introducing a new methodology that can *deform* what we already know to offer researchers meaningful suggestions for further investigation—to create useful and information-based nudges in directions that the researcher likely never considered. Knowledge graph embedding

models are research tools that generate compelling possibilities. We do not claim that these suggestions *could not* have been noticed via other means available to the researcher, but we argue that they probably *would not* have been noticed. This approach allows the researcher to look at existing knowledge in a different way, prompting the investigation of alternatives. And, as we will present briefly at the conclusion of this article, the results for us have been immediate and dramatic: we are currently charting new patterns of crime related to antiquities simply from following a single prompt generated by this model.

The Approach

The first step in our approach, conceptually, is to transform what we know into a knowledge graph or semantic network where the nodes, differentiated by their attached properties, are connected by relationships that are similarly differentiated by their attached properties (for an overview of the field and its animating questions, see Garg and Roy 2022; Ji et al. 2022). Knowledge graphs, as a technology, only became widely known with Google's purchase of the Freebase platform to enhance search in 2012. Google's use allows it to suggest likely results based on its knowledge of the world and not just on link structures, as in its original incarnation powered by the PageRank algorithm. Perhaps more familiar to archaeologists is the concept of "linked open data," which can also be thought of as a knowledge graph in which the entities are anchored to online authority files using the infrastructure of the web itself to represent connections. For an archaeological overview of linked open data, see Schmidt et alia (2022).

"Facts" in a knowledge graph are represented as relationships between entities—for example, "Giacomo_Medici SOLD_TO Christian_Boursaud." We build up a series of such statements derived from the Encyclopedia entries. These statements can be represented as a network, or graph (the terms are synonyms). The structural properties of the graph's nodes (the entities, such as people, businesses, locations, and objects) and edges (the differing kinds of relationships between the entities) allow insights about the complex networks that facilitate this type of crime that might otherwise go undetected (Fensel et al. 2020:69–93).

Graph-based approaches to the illicit antiquities trade that employ social network metrics have been used with some success by Tsirogiannis and Tsirogiannis (2016). In their work, they focus on the transaction paths through a simplified representation of a known network to estimate the most probable paths, drawing on Watson and Todeschini (2007). In this way, they are able to assess which of a variety of network algorithms might prove useful on other, incomplete networks. Other successful network structure approaches to the broader field include the work of Fabiani and Marrone (2021) on auctions, and D'Ippolito's (2014) consideration of what structural network metrics might be appropriate to measure.

However, our approach using a knowledge graph embedding model differs from these kinds of network approaches in that we are not conducting a social network analysis of the graph. We are transforming the graph into a kind of neural network representation of the latent concepts *in* the knowledge *itself* that is captured by the graph (a neural network is a machine learning approach that uses interconnected layers of simulated neurons to process information in order to simulate human cognition). The knowledge graph embedding model approach preserves the semantic

context of the different kinds of relationships in the trade, whereas a network-based approach focuses on the structure of connections.[2]

Once the graph of known relationships is drawn out, the next step is to deploy the full suite of machine learning tools on the subject to create the embedding model. We can train a neural network to "understand" the trade and represent statements about the trade as vectors, mathematical representations, or directions in a multidimensional space—hence, "embeddings." Consequently, statements that are conceptually similar lie in similar regions of this multidimensional space, and the distance or similarity of this positioning can be measured.

This is the same approach used with language models, and which permits machine translation, where equivalent statements in one language have a similarity in multidimensional space as statements in another language: *Je vais à l'école* occupies similar space as "I go to school." Word embedding models can also be used for analogical reasoning, so we can retrieve vectors of words and perform a kind of algebra on them to see, for instance, how language is gendered: in a word embedding model of English, take the vector for "king," remove the vector for "man," add the vector for "woman," and the result *is the same* as the vector for "queen." Word embeddings depend on word positioning in a statement in order to affect the translation into a numerical vector. When an embedding model is derived from a knowledge graph, the same thing is accomplished by taking a node's positioning in relative terms to other neighboring nodes. In our case, we can then examine statements such as "Medici sold_to Hecht" and hypothesize other statements about the trade to see where in the model's vector space such statements fall. The closer to existing clusters of knowledge, the greater the likelihood the statement might be true (see below).

We build the knowledge graph embedding model by scaffolding the nodes and relationships onto a neural network using the AmpliGraph tool (Costabello et al. 2019). Consequently, the concepts and relationships modeled by the graph become vectors as the neural network learns the structure and content of the graph (for the mathematical details, see "Background" in Costabello et al. 2019). The model trains by comparing statements known to be true (the training data) and statements likely to be untrue based on local closed-world assumptions—that is, that which is not known is assumed to be false. The result is that we can measure the distances between different concepts or statements, including relationships not yet seen by the neural network, to predict the likelihood of a relationship being true (with a given confidence). We can give the machine a statement such as "Giacomo_Medici sold_to Marion_True" and measuring the vectorized representation of this statement against the neural model to determine the likelihood of that statement being true.[3]

This multidimensional space can be hard to imagine; techniques exist to project the complexity of the embedding vector model to two or three dimensions. We use the TensorBoard feature of the TensorFlow machine learning Python package from Google to do this. This allows us to visualize the similarity of the nodes' positioning in the original multidimensional vector space and to hypothesize predictions about potential connections.

METHODS

Please see our data availability statement to obtain our data and code. Our data are the 129 case-study-based entries in the Trafficking Culture Encyclopedia at https://traffickingculture.org, as it stood in May 2022. The Trafficking Culture Encyclopedia is a bounded resource, consisting of an approachable number of case studies, many of which were written by one of the coauthors of this article. They represent summaries of antiquities trafficking cases, but as summaries, some details are excluded from them. The authors have collected additional data on these cases outside of the Trafficking Culture Encyclopedia, which allows for model evaluation across two different sources of material. It also allows us to speculate how this model would respond to a larger dataset of material about which we have comparatively less additional knowledge.

To prepare the article files for text extraction and labeling, we begin by scraping article text into separate text files using the conventional HTML parsing package "Beautiful Soup 4" (Richardson 2015) for the Python language. We initially hoped that we could generate the knowledge graph automatically from this scraped data. State of the art approaches at present use large-scale language models to understand a variety of different kinds of relationships, using a kind of transformer-based neural network architecture. In other words, such models understand how to look backward and forward within a text to identify and understand the relationships between nouns. We tried Cabot and Navigli's REBEL model (Cabot and Navigli 2021), and although it extracted many kinds of conventional relationships ("Rome is_located_in Italy"), it missed the players in and the nuances of our subject matter—which is probably a function of how the language model was constructed in the first place—and its training data and did not move us any closer toward reaching our goal.

We turned to the Stanza natural language processing tool from Stanford University's NLP Group (Qi et al. 2020) as a shortcut to automatically tag many of the people, places, objects, and organizations mentioned in the text. Stanza identifies many, but not all, of these "nouns" (and did a better job than the REBEL model in this regard) using a Named Entity Recognition (NER) model trained with the OntoNotes corpus (Weischedel et al. 2013). However, it does not identify the relationships between entities. For that, we imported the tagged documents into the INCEpTION semantic annotation tool (Klie et al. 2018; see Stanza export notebook for our code) for manual annotation of the relationships. INCEpTION provides a browser-based interface for annotation projects (Figure 1). By manually dragging subjects onto objects, we annotated the text from a list of statements that captured the essential relationships: LOOTED, STOLEN_FROM, SOLD_TO, WORKED_WITH, and so on. The team annotated the articles and used INCEpTION's curation tools to reconcile the annotations by multiple team members.[4] The list of relationships or predicates was generated through a close read of the source articles. A first list included every single verb we found. We then reduced the list by coding close synonyms or concepts as the same term.

The resulting data were exported in the WebAnno text format, which we turned into a series of triples, or subject-predicate-object statements (see conversion notebook). These may be found in the file "knowledge-graph.csv." These statements also
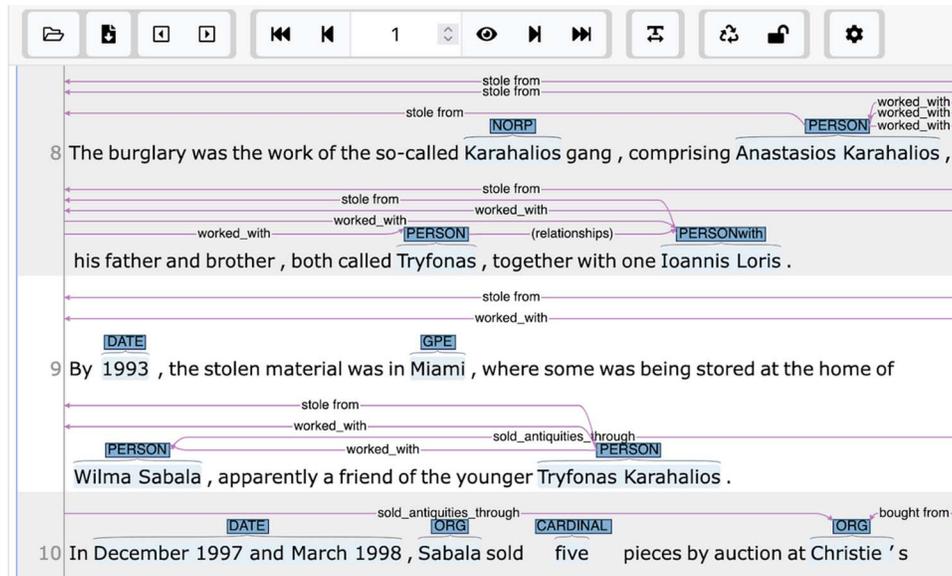
FIGURE 1. The INCEpTION interface, showing an annotation-in-progress. The "nouns" of interest were identified with Stanza, whereas the relationships were drawn by hand by dragging and dropping subjects onto objects.

represent a directed network, and the edges (relationships) can be of multiple types. Conventional network analysis generally assumes that in any particular graph the relationships have to be of the same type—that is, the network is unimodal or 1-mode. Already, we can see one of the advantages of a knowledge graph approach, because it is able to capture and represent a great deal more complexity. Nevertheless, applying conventional network analysis to this material can provide insight about the nature of the graph as a whole, which we discuss below, where we will imagine that the knowledge graph is a 1-mode graph in which the nodes are all actors with agency (even the objects), and the relationships are all reframed simply as "connected_to."

Returning to the full knowledge graph, we employ the AmpliGraph Python library of machine learning to knowledge graph embedding modules in order to transform the graph into a vectorized multidimensional representation of the statements it contains. There are a number of potential embedding model architectures available through AmpliGraph according to a variety of potential parameters. To find the best results, we sweep through the various combinations of parameters, building and comparing the results. A computational notebook that demonstrates how to do this is available in our repository. For the comparison, we used AmpliGraph's function for finding the best "mean reciprocal rank" (MRR) score (see the AmpliGraph documentation for the mathematical definition: https://docs. ampligraph.org/en/1.4.0). The literature on training such models suggested to us that the ComplEx architecture would return the best results (Rossi et al. 2021; Ruffinelli et al. 2020), so we restricted our sweep to settings using that architecture. Our precise model settings are in our code notebook file; we found that using 400 dimensions achieved the best results in this architecture.

To get a sense of the quality of our model (its ability to predict true statements that we know are true but that the model has not yet seen), we split our knowledge graph statements so that 80%

were used for training and 20% were held back for evaluating the model. The procedure for evaluating the model generates "negative" triples (false statements) by taking our test statements and "corrupting" the subject or the object. It filters these statements for any positive statements (known in the training and test sets) inadvertently created during that process. It then ranks the statements in the test set against the negatives to test each statement's likelihood of being true. With our first pass at turning the statements into an embedding model, the evaluation scored a "true" statement as true less than one-third of the time. We improved this score by reexamining our knowledge graph and deducing reciprocal relationships in the graph. For instance, if

   "person_A sold_to person_B"

was in the graph, we created a reciprocal relationship, adding

   "person_B purchased_from person_A"

to the dataset. We proceeded to adjust the statements to clarify the relationships involved, removing ambiguity and adding appropriate reciprocal relationships. We then considered that the domain of our knowledge graph was about actors (humans, organizations) and particular objects in the trade. Consequently, we pruned statements such as "Etruscans area_of_activity Italy" and other similar statements that, although true, did not necessarily enhance the knowledge representation. Many of these statements, if we represented them as a network visualization, would have consisted of dyads floating away from the core "knowledge" captured in the graph.

To evaluate the effectiveness of our model on unseen data, we applied 10-fold cross-validation by shuffling the statements randomly and dividing them into 10 chunks of equal size. We iterated over each of these chunks as the test set (20%) and used the rest of the chunks (80%) as the training set. We report the average scores

of the 10 runs (each run consisted of 1,000 epochs or cycles through the training data); the mean reciprocal ranks, or MRR score, gives us a sense of how often the model evaluates a known true triple or statement as likely being true. The "hits at *n*" score indicates how many times on average a true statement was evaluated within the top 10, three, or first ranks (there are as many ranks as there are statements).

- Average MRR: 0.86
- Average hits@10: 0.89
- Average hits@3: 0.87
- Average hits@1: 0.83

Over the 10 runs, the MRR ranged from 0.81 to 0.90. The hits@10 score ranged from 0.85 to 0.94. The hits@3 score ranged from 0.82 to 0.92, and hits@1 ranged from 0.78 to 0.87. Therefore, for our knowledge graph embedding model, we might say that it can identify a known "true" statement as probably true around eight times out of 10.

After annotation and reconciliation, the knowledge graph contained 1,204 statements about 478 entities using 81 unique verbs (relationships/predicates) derived from the 129 encyclopedia articles that describe the illicit and illegal antiquities trade. We then proceeded to explore this knowledge graph and compare its predictions with what we already know about the trade, fitting a model to the complete dataset (all 1,204 statements) while being cognizant of its limitations.

NETWORK VISUALIZATION AS A CHECK ON THE PROCESS

Although we will not perform a "conventional" network analysis, it can be helpful to get an overview of the knowledge graph by thinking of it as a regular network where all entities are imagined as "actors" and all relationships are imagined as "connected_to." In other words, we reduce what is technically a multimodal graph from a conventional network analysis perspective to a simple unimodal graph to obtain a coarse vision of its overall structure.

A visualization of these statements as a network gives us a sense of the nature of the knowledge graph (Figure 2). This visualization imagines every entity as being of the same kind of thing, an actor in this particular universe, and the connections between them simply that—a mere connection. This allows us to see at a glance that there is a complex core of ideas, actors, and connections at the heart of the Trafficking Culture Encylcopedia's representation of the antiquities trade, with some isolated concepts in its periphery. This reflects what we know about how the encyclopedia was constructed. The visualization is generated using the network visualization software Gephi (Bastian et al. 2009), and the colors are from the "modularity" routine that identifies clusters of nodes based on the self-similarity of their connections. The trails of connected nodes remind us indeed of "cordata," as "people roped together," while there is an outer orbit of concepts and ideas floating freely or in small clumps (the inset image).

In Figure 2, we see the centrality of the figure of Giacomo Medici as represented in the encyclopedia articles from Trafficking

Culture. Other important nodes tying this all together include the Sotheby's, Christie's, and Bonhams auction houses; dealers such as Gianfranco Becchina; and museums such as the Getty Museum and the Metropolitan Museum of Art. Indeed, this visualization serves as a kind of check in that it represents what we already know about the trade in general, confirms our expectations about our data, and also illustrates the European- and North American–centric nature of a lot of the knowledge graph as represented in this source. *In the gaps between this central core and the periphery lie all of the things we do not yet know about the trade.* This is where the use of machine learning and knowledge embeddings to perform "link prediction" comes into play. We use the tools of "link prediction" from AmpliGraph on the embedding model to work out hypotheses about these blanks on our map.

RESULTS

The knowledge statements, remember, are descriptions of relationships; the existence of a relationship not previously seen by the model is the problem of predicting the likelihood of a semantic connection of some kind, given what the model already knows. The model represents our statements and their interconnections as a mathematical vector in a multidimensional space. Predicting these connections, therefore, becomes a question of crafting statements that feature the subject and object. When such statements lie as close as possible to known statements within that space, we have a measurement of the likelihood that the statement is true. Consequently, "link prediction" in the context of a knowledge graph embedding model is not the same thing as "path prediction" as investigated by Tsirogiannis and Tsirogiannis (2016); it is less about structure and more about testing the likelihood of various hypotheses.

What links should we test? The statements must feature entities and relationships already in the training data (for methods on out-of-vocabulary predictions, see Demir and Ngonga Ngomo 2021). For instance, if we wanted to assess the likelihood of the statements below, we ask the model to predict the probability of the linkage. None of these exact statements are in the knowledge graph we derived from the Trafficking Culture Encyclopedia, and we are not implying here that they are or are not true. The code block looks like this:

```
["Giacomo Medici," "employed, " "Marion True"],
["Giacomo Medici, " "sold_antiquities_to," "Marion True"],
["Marion True," "bought_from," "Giacomo Medici"],
["Roger Cornelius Russell Yorke," "bought_from," "Robin Symes"],
["Fritz Bürki," "sold_antiquities_to," "Leon Levy"],
["Gianfranco Becchina," "partnered," "Hicham Aboutaam"],
["Robert Hecht," "sold_antiquities_to," "Barbara Fleischman"]
```

For context, Giacomo Medici is an Italian antiquities dealer convicted of antiquities-related crimes in 2005. Marion True was a curator at the J. Paul Getty Museum until 2005, who was charged with antiquities-related crimes but not convicted. Robin Symes is a British antiquities dealer, who was convicted of antiquities-related crimes in 2005. Roger Cornelius Russell Yorke is a Canadian art dealer, who was convicted of antiquities-related crimes in 1992.
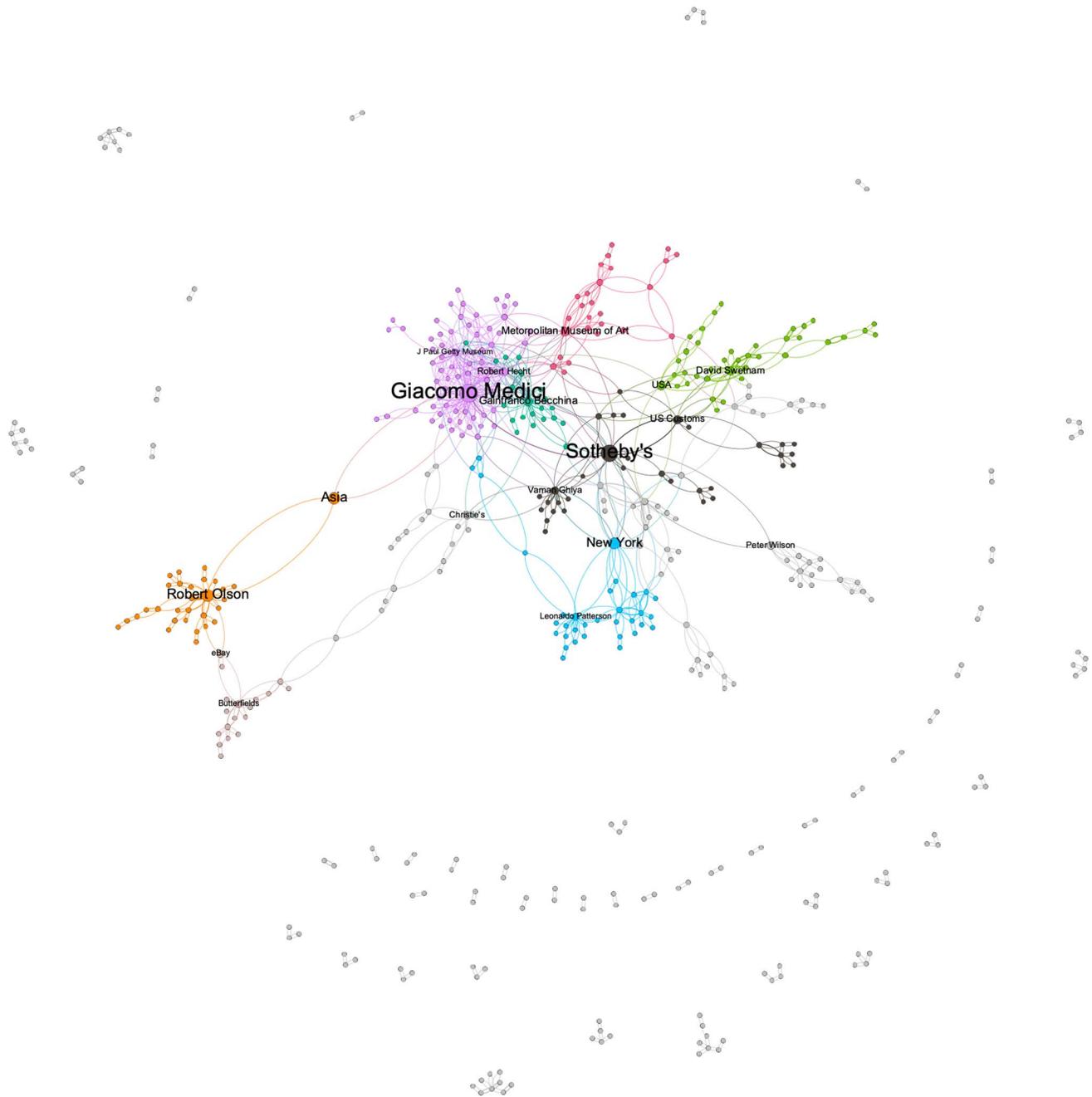
FIGURE 2. A network representation of the knowledge graph created through the annotation of Trafficking Culture Encyclopedia articles. Node size, and the associated label, is scaled to reflect a node's importance as measured by "betweenness centrality": the more times a node lies on the shortest path between any two other nodes, the larger that node is depicted. The smaller nodes, then, represent those that are not "important" on this measure and are therefore purposefully deemphasized for the user. Note the unconnected periphery of isolated clumps.

Fritz Bürki is a Swiss art conservator, who often acted as a front for Robert Hecht. Leon Levy was a New York–based antiquities collector. Gianfranco Becchina is an Italian antiquities dealer convicted of antiquities-related crimes in 2011. Hicham Aboutaam is a cofounder of the dealership Phoenix Ancient Art and was convicted of antiquities-related crimes in 2004. Robert Hecht was an

antiquities dealer and the American end of the trafficking chains beginning with Medici and Becchina. Barbara Fleischman is an American antiquities collector.

In the code block, the statements are passed through the model and returned with a rank (i.e., "1," the first rank, is predicted to be

TABLE 1. Rank, Score, and Probability of Statements Tested via Knowledge Graph Embedding Model.

Statement	Rank	Score	Probability
Giacomo Medici sold_antiquities_to Marion True	1	14.41	0.99
Marion True bought_from Giacomo Medici	2	3.95	0.98
Fritz Bürk sold_antiquities_to Leon Levy	5	2.69	0.94
Robert Hecht sold_antiquities_to Barbara Fleischman	19	1.82	0.86
Gianfranco Becchina partnered_with Hicham Aboutaam	207	0.21	0.55
Roger Cornelius Russell Yorke bought from Robin Symes	890	−1.40	0.19
Giacomo Medici employed Marion True	932	−1.95	0.12

Note: These particular statements are used to demonstrate the output of the various possible measurements of the model using AmpliGraph.

most likely true), a score (where the greater the positive value, the more likely the statement), and a probability between 0 and 1. The results for our example statements above are in Table 1. We can consider these statements to be hypotheses that one might float to guide further research.

The model returns the following ranks, scores, and probabilities (Table 1). We will discuss these scores below in the discussion section.

As indicated, a limitation of the model is that we cannot ask it to predict the likelihood of statements where the subject, object, or predicate are individually not already present in its knowledge. For instance, if there is a statement about "OTTAWA" elsewhere in the model, then we could ask it to assess the likelihood of "Giacomo Medici WORKED_IN Ottawa." But if there is *no* existing knowledge about Ottawa in the model, then the evaluation will return an error. AmpliGraph comes with a number of functions to facilitate discovery of new knowledge in the embedding model from the existing entities. These function in a way similar to how the model as a whole was evaluated when we first trained it. These functions generate new statements from the entities and predicates in the graph and evaluate their likelihood by way of ranking them against corrupted sets. Corrupted sets are true statements in which the subject or object gets swapped out. The statements get filtered against the training data to make sure we do not create true statements, and the resulting statement is then assumed—under closed world assumptions—to be a known false statement (in logic, the closed world assumption is the idea that any statement that is not known to be true is assumed to be false). True statements that fall closely in the embedding space to known false statements therefore rank lower. Top-ranked statements are taken as having the highest probability of being true. In this way, we use the knowledge graph embedding model as a way to produce new leads—new ideas to pursue.

For the discovery of new statements/hypotheses that we might not have generated ourselves, we retrain the model on the full knowledge captured in the original graph. We create candidate statements and then evaluate their probability. We can write these statements by hand and then pass them through the model, or we can use the strategies encoded in the function for statement creation. The function assumes that for well-connected parts of the graph most facts are known, so it uses measurements such as the degree of an entity (the count of its

relationships) to create and evaluate statements for entities from the poorer-known regions, and it measures where these statements fall in that multidimensional space.

We generated 20,000 statements five separate times, using five separate strategies of "entity frequency," "graph degree," "clustering coefficient," "cluster triangles," and "cluster squares" and the predicate "bought_from." The top most likely statements by the various strategies are compiled in Table 2. Note that none of these statements exist in the original Trafficking Culture Encyclopedia knowledge graph.

In interpreting these scores, one should want to take into account the rank, score, and probability altogether. Therefore, we might decide to keep the statements in the first few ranks and with the higher probabilities as hypotheses worth exploring.

We generated candidate statements again using the same five strategies run 20,000 times each, with the predicate "partnered." The most likely statements are compiled in Table 3.

Visualizing the Knowledge Embedding Space

We can also visualize the entire knowledge graph embedding model as a two-dimensional space where entities are clustered more closely together depending on our entire knowledge of the domain in question. When the model was first specified, we set the number of dimensions at 400; the reduction and then the visualization to two dimensions is accomplished using the Uniform Manifold Approximation and Projection (UMAP) algorithm for 500 epochs and visualized with the TensorBoard extension for the TensorFlow Python package (see the code notebook). We set it to use the 15 nearest neighbors to approximate the overall shape of the space.

The career, connections, and activities of Giacomo Medici are well known. We find him in the visualization, and we see that another dealer of interest—Leonardo Patterson—is in the same general proximity. In other words, the model correctly identifies that Leonardo Patterson is a figure somewhat similar to Medici in the broader antiquities trade. We know, however, that Patterson's activities were within the ambit of antiquities from Central and South America. Patterson and Medici are, globally, in the same bottom-right quadrant of the overall knowledge graph embedding model (zooming into the model causes a dynamic expansion of the points in TensorBoard).

TABLE 2. Candidate Statements "Bought_from" with Rank, Score, and Probability, Given Knowledge Graph Embedding Model.

Statement	Rank	Score	Probability
J Paul Getty Museum *bought_from* Samuel Schweitzer	1	12.87	0.99
J Paul Getty Museum *bought_from* Frieda Tchacos	1	12.63	0.99
Charles Craig *bought_from* David Swetnam	1	12.60	0.99
Dietrich von Bothmer *bought_from* Gianfranco Becchina	1	11.60	0.99
Robert Hecht *bought_from* Robin Symes	3	4.52	0.99
Harry Brown *bought_from* Johnnie Brown Fell	1	4.26	0.99
Marion True *bought_from* Giacomo Medici	3	3.74	0.98
Giacomo Medici *bought_from* Nikolas Koutoulakis	2	3.54	0.97
Hydra Gallery *bought from* Antonio 'Nino' Savoca	1	3.34	0.97
Benjamin Bishop Johnson *bought_from* Fred Drew	1	2.90	0.95
Leon Levy *bought_from* Fritz Bürki	24	1.49	0.82
Joel Malter *bought_from* Marquis of Tavistock	38	1.04	0.74
Pereda *bought_from* J Paul Getty Museum	126	0.65	0.66
Roger Cornelius Russell Yorke *bought_from* Harry Brown	274	0.33	0.58
Vaman Ghiya *bought_from* David Bernstein	379	0.18	0.55
Leonardo Patterson *bought_from* Clive Hollinshead	829	−0.76	0.32

Note: These should be regarded as "hypotheses" for further exploration.

TABLE 3. Candidate Statements "Partnered" with Rank, Score, and Probability, Given Knowledge Graph Embedding Model.

Statement	Rank	Score	Probability
Anton Tkalec *partnered* Mansur Mokhtarzade	1	9.27	0.99
Robert Hecht *partnered* Robin Symes	1	6.92	0.99
Michael Kelly *partnered* Miguel de Osma Berckemeyer	1	4.68	0.99
United States Customs *partnered* Royal Canadian Mounted Police	1	3.61	0.97
Mario Bruno *partnered* Giacomo Medici	3	3.08	0.96
Roger Cornelius Russell Yorke *partnered* Charles Craig	2	2.36	0.91
Charles Craig *partnered* Roger Cornelius Russell Yorke	6	1.82	0.86

Note: These should be regarded as "hypotheses" for further exploration.

We take the cosine distance and find the other entities closest to "Leonardo Patterson" are these entities listed in Table 4 (illustrated in Figure 3b; some points overlap, so they are not labeled).

We are not arguing that these other entities are "the same" as "Leonardo Patterson." The representation of statements about these entities, when translated into an embedding knowledge, are this distance away from each other, which suggests—in a fuzzy way—that there are aspects about them (which we cannot determine from this visualization) that create a kind of clustering. But the distances here do not seem that close.

Consider instead the space closest to "Giacomo Medici." The closest entities for "Giacomo Medici" are rather closer to the "Giacomo Medici" point than those closest entities for "Leonardo Patterson" (Table 5; space illustrated in Figure 3c):

Given that we know that these individuals were indeed associated with one another, these distances might be a useful threshold for prompting further investigation on a researcher's part. In this case,

with regard to "Leonardo Patterson," one might wish to look into whether there are indeed any relationships between "Leonardo Patterson" and the "Brooklyn Museum," for instance, as the closest entity to Patterson in the vector space of the model.

DISCUSSION

Consider the example statements we crafted for Table 1. The model considers it extremely likely that Giacomo Medici sold antiquities to Marion True and, of course, the inverse—that Marion True bought antiquities from Giacomo Medici. Giacomo Medici is an Italian antiquities dealer known to occupy an important place within illicit antiquities networks emanating out of Italy until his conviction in 2005 (Watson and Todeschini 2007). Marion True was a curator at the J. Paul Getty Museum from 1986 until 2005 when she was charged, but not ultimately convicted in Italy, of antiquities-trafficking-related offences (Felch and Frammolino 2011). Although the Trafficking Culture Encyclopedia does not explicitly say that Medici sold antiquities to

TABLE 4. Cosine Distance from "Leonardo Patterson" as Projected in the UMAP Visualization Using the Default Settings in TensorBoard.

Entity	Cosine Distance from "Leonardo Patterson"
Brooklyn Museum	0.512
André Emmerich	0.619
Marjorie Neikrug	0.659
Glenn Rittenour	0.755
Edward Merrin	0.891
Phoenix Ancient Art	0.910
Fritz Bürki	0.920
Frederick Schultz	0.922
Abraham Trust	0.927
Art Gallery of New South Wales	0.929

True, he did, and, as the encyclopedia entry for True states, "True was charged in Italy with receiving stolen antiquities and conspiring with dealers Robert Hecht and Giacomo Medici to receive stolen antiquities, and she was ordered to stand trial in Rome" (Brodie 2012b).

Turning to the two least likely examples, the model predicts that it is extremely unlikely that Medici employed True. As previously stated, True was employed by the Getty Museum, and Medici was an active antiquities trafficker. There are few conceivable scenarios where their relationship would involve True's employment by Medici, and there is no evidence that it ever did. The model also considers it unlikely that Roger Cornelius Russell Yorke bought from Robin Symes (Table 1). Symes is a British former antiquities dealer who primarily traded in Greek and Italian antiquities and who was heavily involved in Medici's network (Watson and Todeschini 2007). Yorke is a Canadian collector and dealer in Andean textiles who, in 1993, became the first person convicted under Canada's Cultural Property Export and Import Act of 1977, which was related to the illicit trafficking of Bolivian objects (Paterson 1993; Paterson and Siehr 1997). The market networks between Andean textiles and Classical antiquities are not known to have much crossover, and we have no knowledge of Yorke ever purchasing the type of antiquities that Symes would sell. Again, the model conforms to our knowledge.

Perhaps more challenging are the statements that are less likely but are still deemed probable by the model. Take, for example, the statement that Robert Hecht sold antiquities to Barbara Fleishman, which was assigned 86% probability (Table 1). Robert Hecht was a dealer in Greek and Italian antiquities who was indicted alongside Marion True for involvement in the greater network that also involved Medici and Symes. Barbara Fleishman, alongside her late husband Lawrence, is a collector of often unprovenanced Classical antiquities, many of which were acquired by the Getty Museum. Fleishman and Hecht clearly had an interest in the same material and ran in the same circles at the same time. Although the authors do not have direct knowledge that Hecht did, indeed, sell to Fleishman directly, we do know of numerous objects that connect the two (e.g., a looted fresco fragment from Pompeii [Alberge 2022]).

Further provenance research may confirm this predicted connection.

Turning to Table 2 and the predictions that the model makes using its own generated statements, we see some interesting ideas but, perhaps, some space for improvement. Many of the high-ranking statements are demonstrably true. More interesting is where the model went wrong. For example, take the statement that the "J Paul Getty Museum bought_from Samuel Schweitzer." The Schweitzer Collection is actually considered to be a false provenance, a fake ownership history provided to looted antiquities. The Getty may have been told that the objects they were buying were from the Schweitzer Collection, but they were not. The model, it seems, is tricked in the same way as the Getty Museum, but the museum should have known better. Also curious is just how unlikely "Leonardo Patterson bought_from Clive Hollinshead" is deemed by the model. Both of these men were involved in the trafficking of illicit Maya antiquities into the United States in the 1970s and 1980s, both men have convictions in the United States for this activity, and both men were within the network of people who knew about the illicit movement of Machaquilá Stela 2 from Guatemala (Yates 2020). Although the authors have no direct evidence that Patterson ever bought from Hollinshead, it does not seem entirely unlikely.

In considering Table 3, where we ask the model to generate likely partnerships, once again, most of the results are objectively true. However, the model predicts a possible partner relationship between Roger Cornelius Russell Yorke (mentioned above) and Charles Craig, who was a Santa Barbara–based retired bank executive involved in the receiving of looted antiquities from the site of Sipán, Peru (Yates 2012). Although our initial thought was that this pairing was unlikely, on further consideration, it is a possibility worth investigating. Both men were involved in the trafficking of admittedly different types of antiquities from the neighboring countries of Peru and Bolivia during the same time period. It is a connection that is not impossible, and one that we are likely to never have considered without the model's suggestion.

All told, the most interesting possible associations generated by the model seem to fall in the 80% range. Those in the approximately 90% range are so obvious as to be well known to everyone involved in this line of research. Those in the much lower percentage range are mostly, but not entirely, objectively very unlikely. However, there is an interesting middle here of proposed connections that rest outside of our existing knowledge but within what we consider possible, yet we were unlikely to propose their possibility independently.

The reduction of the model to two dimensions so that we can see (and measure) distances in the similarity space is another approach to generating hypotheses. In this case, based on the well-attested nexus of relationships around Giacomo Medici (the cosine distances in the UMAP visualization of the space, Figure 3c), we take those distances as a kind of rule of thumb to look at another individual, Leonardo Patterson (mentioned above). Patterson is a Costa Rican national with a long history of antiquities crime convictions in multiple countries, alongside other forms of dubious behavior related to so-called precolumbian antiquities (Elias 1984; Yates 2016). Most recently, in 2015, Patterson was convicted in a German court for crimes related to both fake and real Olmec antiquities (Mashberg 2015). Patterson's participation in the illegal
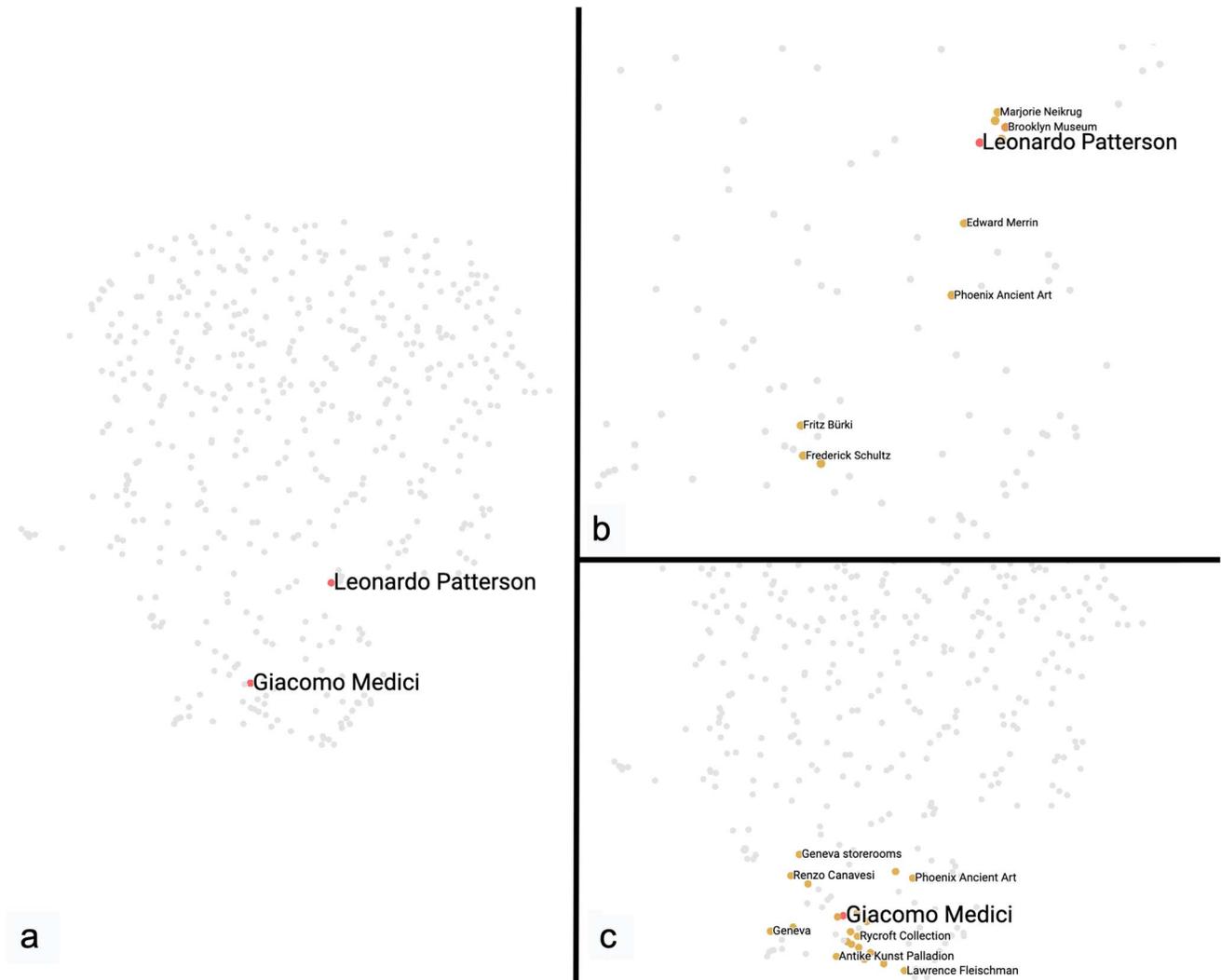
FIGURE 3. Visualization of the knowledge graph embedding model projected to two dimensions via UMAP approixmation showing (a) 15 nearest neighbors, indicating the "Leonardo Patterson" and "Giacomo Medici" points; (b) zoom into the area around the "Leonardo Patterson" point; (c) zoom into the area around the "Giacomo Medici" point.

trade in antiquities is well known and well documented. As can be seen in Figure 3b, Patterson is spaced relatively close to another precolumbian antiquities dealer, André Emmerich, although the two are not directly linked in the Trafficking Culture Encyclopedia. However, we know that the two men had significant links: Emmerich's gallery records, housed at the Smithsonian Archives of American Art, contain no less than nine folders of correspondence with and documents about Patterson—including such titillating contents as a post-it note stating that the FBI was looking for Patterson—and documentation related to the fake Olmec sculpture that is connected to Patterson's German convictions.[5] That speaks well of the model but does not yet tell us something we did not already know. The model might be useful for guiding network research, something we sought to test.

This visualization of the model as points in a two-dimensional space generates a hypothesis that Patterson somehow is "similar" or close to the Brooklyn Museum, implying some sort of

connection although the two are not directly linked in the Trafficking Culture Encyclopedia. Patterson was known to be based out of New York City during the late 1960s and into the 1970s, so in close proximity to the museum. The Brooklyn Museum was engaged in the trade in precolumbian material at the same time Patterson was in New York, culminating in its repatriation of fragments of a stela that had been stolen from the Guatemalan site of Piedras Negras (*Current Anthropology* 1973). That said, we had no prior knowledge of a link between Patterson and the Brooklyn Museum, and we had never thought to investigate such a connection.

A search of the Brooklyn Museum website shows that the model guided us toward something interesting. As it turns out, in 1969, Patterson donated at least two precolumbian antiquities to the Brooklyn Museum: a ceramic whistle shaped like a dog (accession number 69.170.1) and a small seated figurine (accession number 69.170.2), both of which are still in the museum collection. Neither

TABLE 5. Cosine Distance from "Giacomo Medici" as Projected in the UMAP Visualization Using the Default Settings in TensorBoard.

Entity	Cosine Distance from "Giacomo Medici"
Robert Hecht	0.565
Gianfranco Becchina	0.613
Antike Kunst Palladion	0.650
Rycroft Collection	0.653
Pasquale Camera	0.662
Fritz Bürki	0.664
Maurice Tempelsman	0.668
Dikran Sarrafian	0.682
Samuel Schweitzer	0.707
Robin Symes	0.717

item is presented as having any provenance information, and both were accessioned at the same time that the museum was dealing with the abovementioned looted stela fragments. The fact that they were *donated* by Patterson rather than *sold* raises a number of intriguing questions that we are currently following up on with additional research. This connection alone has enriched our understanding of the New York–based networks involved in pre-columbian antiquities trafficking. The potential for this model to provide fruitful possibilities for researchers to elaborate on is clear.

CONCLUSION

Considering the Patterson–Brooklyn Museum example begs the question, Could other methods have drawn our attention to this connection? Obviously, yes. The information that Patterson donated to the Brooklyn Museum is available online *if one knew to look*. However, and we stress, until this model suggested a connection between these two entities, we had no reason to suspect a connection at all. It is a question we never would have asked.

In the months since we were first prompted by the model, we have opened a completely new research line into Patterson's emerging pattern of museum donations to a number of museums across the world. In contrast to a known museum donation / tax evasion scheme involving Patterson in Australia (see Yates 2016), we are now seeing a tantalizing and previously undocumented pattern of donation of low-value unprovenanced antiquities to multiple institutions. Furthermore, emerging evidence coming from within museum records and court documents seems to connect at least some of these minor museum donations to broader antiquities fraud schemes perpetrated by Patterson. Our running theory, which we are continuing to investigate, is that Patterson sought to launder his own reputation through placing objects within major museums. When he then attempted to convince a buyer to pay a significant amount of money for a fake Maya mural, as he did in 1984, he could point to the fact that his objects were in the collections of the British Museum, the National Gallery of Australia, the National Museum of the American Indian, or the Brooklyn Museum as an indicator of respectability and esteem. We will be presenting this information in future publications. We are now communicating with museums that house objects donated by Patterson, and several of these institutions, disturbed by what they

and we have found, have been prompted to conduct internal reviews of the pieces in question. It is unlikely that we would be uncovering this emerging crime pattern, and it is unlikely that anyone would have looked at these minor old donations, without the model offering us the prompt.

In the digital humanities, it is often easy to say, after the fact, "Oh, we already knew that!" Lincoln (2015, 2017) has identified this problem as "confabulation": after-the-fact rationalizations of the findings of computational approaches. The simple fact remains that, despite our clear prior research interests in Patterson and criminality in the market for antiquities, we had no reason to look for a connection between Patterson and the Brooklyn Museum. Now that we have, we discover a thread connecting Patterson to a much larger pattern of illicit financing and influence laundering that we now get to unravel.

Although we have been concerned here with the trade in antiquities, there is no reason why this same approach could not be applied to other domains of archaeological or historical knowledge. Any place where a network analysis approach might be valid could perhaps be investigated through transformation into a knowledge graph embeddings approach. The use of graph databases in archaeology is gathering some steam (Schmidt et al. 2022). Graph databases can be queried (depending on the approach) with query languages such as Cypher or SPARQL, which focus on traversing the graph in order to surface results. However, it might be that graph embedding models could surface interesting patterns or insights based on patterns in the multidimensional space. In de Haan et alia (2021), the authors create a knowledge graph from an open-access repository of research results (the Cooperation Databank) to generate a graph connecting scientific observations with the published results, and then they use a knowledge graph embedding model (via AmpliGraph) to generate hypotheses about the domain likely to be true. Similar approaches are used in bioinformatics for new drug prediction or disease response (Zhu et al. 2022). Perhaps a similar workflow, using data from Open Context, tDAR, or the Archaeological Data Service could serve as a model here (a pipeline for working with knowledge graphs that uses as an example Dutch linked open-data protocols for archaeological materials is discussed in Wilcke et al. 2019). Simple statements of knowledge can lead to entirely new perspectives.

The simple statements that capture knowledge of the illegal or illicit trade in antiquities as a series of relationships, combined with machine learning, enable us to represent a domain of knowledge in such a way that we can generate predictions. These predictions can then be used to focus research energies. We intend to use these statements, and the embedding model we derive from them, in a further study to create an automated relationship extraction pipeline (at present, the bottleneck is in the annotation and automatic extraction of relationships from unstructured text). We could then use the pipeline on other germane texts such as newspaper articles, the Panama Papers, judicial documents, and open museum collections (for an allied approach in terms of cultural heritage more generally, see Dutia and Stack 2021). Hardy's ongoing explorations of metal-detecting websites and other hidden-in-plain-sight fora (Hardy 2021) might also be amenable. By building a pipeline to derive the relationships from unstructured text automatically rather than relying on hand annotations, we will be able to create an expanded knowledge graph at scale that will help us bridge from these core,

well-known case studies to illuminate the shadier and more hidden aspects of the trade. We will be able to represent this knowledge as a knowledge embedding model and predict more of the hidden structure.

As an often illegal, often illicit, always murky trade, the commerce in antiquities and other cultural heritage materials is only visible to us in those moments when a prosecution is completed, or when elements surface in auction catalogs or other public records. It is filled with gaps and shadows. By taking what we do know and adding to the graph continually, we can begin to see a structure even when we do not know the precise relationship between entities. We can state hypotheses and have some sense of the likelihood of them being true. We caution that this approach does not prove any of these hypotheses, but with careful queries, we can use it to help direct our attention toward elements that might bear further investigation.

Data Availability Statement

All of our computational notebooks and the knowledge graph CSV file are available at https://doi.org/10.5281/zenodo.7506971 and may be run using Jupyter on a personal computer, or online via Google's Colab service; for use on a personal computer, a GPU is recommended.

Competing Interests

The authors declare none.

NOTES

1. For the trade facilitated by online social media platforms, see Al-Azm and Paul (2019). In this article, although the method could usefully be employed for the online dimension of the trade, we are at present dealing with its historical contours.
2. In other words, in social network analysis, it is necessary to jettison the multimodal knowledge of the different kinds of relationships because most network analysis metrics require data be projected to one mode, or one kind of relationship, only. See Graham et alia (2022:205).
3. This is "link prediction." Note that this is not link prediction as might be understood in conventional network analysis as, for instance, in the closure of triangles, where if A connects to B and A connects to C, then B likely connects to C.

4. In a follow-up article, we will detail a new method we have since developed to identify entities and relationships from unstructured text automatically. This significantly reduces the labor involved in developing a knowledge graph and promises to scale easily to thousands of texts.
5. See https://www.aaa.si.edu/collections/andremmerich-gallery-records-6275.

REFERENCES CITED

Al-Azm, Amr, and Katie Paul. 2019. Facebook's Black Market in Antiquities. *Athar Project*, n.d. http://atharproject.org/report2019/, accessed July 13, 2022.

Alberge, Daya. 2022. Fresco Fragment from Pompeii Reopens Row over "Looted" Artefacts. *Guardian*, March 20. https://www.theguardian.com/science/2022/mar/20/getty-museum-fresco-fragment-pompeii-row-looted-artefacts, accessed July 13, 2022.

Bastian, Mathieu, Sébastien Heymann, and Mathieu Jacomy. 2009. The Open Graph Viz Platform. *Gephi*. https://gephi.org/, accessed July 18, 2022.

Brodie, Neil. 2012a. Organigram. *Trafficking Culture Encyclopedia*, August 21. https://traffickingculture.org/encyclopedia/case-studies/organigram/, accessed February 7, 2023.

Brodie, Neil. 2012b. Marion True. *Trafficking Culture Encyclopedia*, September 9. https://traffickingculture.org/encyclopedia/case-studies/marion-true/, accessed July 13, 2022.

Brodie, Neil, Donna Yates, Brigitte Slot, Olga Batura, Niels van Wanrooij, and Gabriëlle op 't Hoog. 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.* Directorate-General for Education, Youth, Sport and Culture, European Commission, Brussels. https://data.europa.eu/doi/10.2766/183649, accessed March 8, 2023.

Cabot, Pere-Lluís Huguet, and Roberto Navigli. 2021. REBEL: Relation Extraction by End-to-End Language Generation. In *Findings of the Association for Computational Linguistics*, pp. 2370–2381. Association for Computational Linguistics, Punta Cana, Dominican Republic. https://doi.org/10.18653/v1/2021.findings-emnlp.204.

Costabello, Luca, Sumit Pai, Chan Le Van, Rory McGrath, Nick McCarthy, and Pedro Tabacof. 2019. AmpliGraph: A Library for Representation Learning on Knowledge Graphs. *Zenodo*. https://doi.org/10.5281/zenodo.2595043.

Current Anthropology. 1973. Brooklyn Museum Presents Stela Fragments to Guatemala. *Current Anthropology* 14(5):579. https://doi.org/10.1086/201388.

de Haan, Rosaline, Ilaria Tiddi, and Wouter Beek. 2021. Discovering Research Hypotheses in Social Science Using Knowledge Graph Embeddings. In *The Semantic Web*, edited by Ruben Verborgh, Katja Hose, Heiko Paulheim, Pierre-Antoine Champin, Maria Maleshkova, Oscar Corcho, Petar Ristoski, and Mehwish Alam, pp. 477–494. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-77385-4_28.

Demir, Caglar, and Axel-Cyrille Ngonga Ngomo. 2021. Out-of-Vocabulary Entities in Link Prediction. *arXiv*. https://doi.org/10.48550/arXiv.2105.12524.

D'Ippolito, Michelle. 2014. New Methods of Mapping. In *Social Computing, Behavioral-Cultural Modeling and Prediction*, edited by William G. Kennedy, Nitin Agarwal, and Shanchieh Jay Yang, pp. 253–260. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-05579-4_31.

Dutia, Kalyan, and John Stack. 2021. Heritage Connector: A Machine Learning Framework for Building Linked Open Data from Museum Collections. *Applied AI Letters* 2(2). https://doi.org/10.1002/ail2.23.

Elias, David. 1984. FBI Arrests Pre-Columbian Art Dealer on Fraud Charges. *The Age*, June 5.

Fabiani, Michelle, and James Marrone. 2021. Transiting through the Antiquities Market. In *Crime and Art: Sociological and Criminological Perspectives of Crimes in the Art World*, edited by Naomi Oosterman and Donna Yates, pp. 11–28. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-84856-9_2.

Felch, Jason, and Ralph Frammolino. 2011. *Chasing Aphrodite: The Hunt for Looted Antiquities at the World's Richest Museum.* Houghton Mifflin Harcourt, Boston.

Fensel, Dieter, Umutcan Simsek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, and Alexander Wahler. 2020. How to Use a Knowledge Graph. In *Knowledge Graphs:*

Methodology, Tools and Selected Use Cases, edited by Dieter Fensel, Umutcan Simsek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, and Alexander Wahler, pp. 69–93. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-030-37439-6_3.

Garg, Satvik, and Dwaipayan Roy. 2022. A Birds Eye View on Knowledge Graph Embeddings, Software Libraries, Applications and Challenges. *arXiv*. https://doi.org/10.48550/arXiv.2205.09088.

Graham, Shawn, Ian Milligan, Scott B. Weingart, and Kim Martin. 2022. *Exploring Big Historical Data: The Historian's Macroscope*. 2nd ed. World Scientific, Singapore.

Hardy, Samuel A. 2021. It Is Not against the Law, if No-One Can See You: Online Social Organisation of Artefact-Hunting in Former Yugoslavia. *Journal of Computer Applications in Archaeology* 4(1):169–187. http://doi.org/10.5334/jcaa.76.

Ji, Shaoxiong, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. 2022. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems* 33(2):494–514. https://doi.org/10.1109/TNNLS.2021.3070843.

Klie, Jan-Christoph, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The INCEpTION Platform: Machine-Assisted and Knowledge-Oriented Interactive Annotation. In *Proceedings of System Demonstrations of the 27th International Conference on Computational Linguistics*, pp. 5–9. Association for Computational Linguistics, Santa Fe, New Mexico. Electronic document, https://aclanthology.org/C18-2002, accessed March 8, 2023.

Lincoln, Matthew. 2015. Confabulation in the Humanities. *Matthew Lincoln, PhD Cultural Heritage Data & Info Architecture* (blog), March 15. Electronic document, https://matthewlincoln.net/2015/03/21/confabulation-in-the-humanities.html, accessed November 4, 2022.

Lincoln, Matthew. 2017. Continuity and Disruption in European Networks of Print Production, 1550–1750 *Artl@s Bulletin* 6(3):Article 2. https://docs.lib.purdue.edu/artlas/vol6/iss3/2/, accessed March 8, 2023.

Mashberg, Tom. 2015. Antiquities Dealer Leonardo Patterson Faces New Criminal Charges. *New York Times*, December 8. https://web.archive.org/web/20220616132416/https://www.nytimes.com/2015/12/09/arts/design/antiquities-dealer-leonardo-patterson-faces-new-criminal-charges.html, accessed July 26, 2022.

Oosterman, Naomi, Simon Mackenzie, and Donna Yates. 2021. Regulating the Wild West: Symbolic Security Bubbles and White Collar Crime in the Art Market. *Journal of White Collar and Corporate Crime* 3(1):7–15. https://doi.org/10.1177/2631309X211035724.

Paterson, Robert K. 1993. Bolivian Textiles in Canada. *International Journal of Cultural Property* 2(2):359–370. https://doi.org/10.1017/S0940739193000372.

Paterson, Robert K., and Kurt Siehr. 1997. Conviction in Canadian Smuggling Case—a Pyrrhic Victory? *International Journal of Cultural Property* 6(2):401–416. https://doi.org/10.1017/S0940739197000465.

Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *arXiv*. https://doi.org/10.48550/arXiv.2003.07082.

Richardson, Leonard. 2015. *Beautiful Soup 4*. Electronic document, https://beautiful-soup-4.readthedocs.io/, accessed July 13, 2022.

Rossi, Andria, Denilson Barbosa, Donatella Firmani, Antonio Matinata, and Polo Merialdo. 2021. Knowledge Graph Embedding for Link Prediction.

ACM Transactions on Knowledge Discovery from Data 15(2):1–49. https://doi.org/10.1145/3424672.

Ruffinelli, Daniel, Rainer Gemulla, and Samuel Broscheit. 2020. You CAN Teach an Old Dog New Tricks! On Training Knowledge Graph Embeddings. Paper presented at International Conference on Learning Representations (ICLR) 2020, Online. Electronic document, https://paperswithcode.com/paper/you-can-teach-an-old-dog-new-tricks-on, accessed March 8, 2023.

Schmidt, Sophie C., Florian Thiery, and Martina Trognitz. 2022. Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata. *Digital* 2(3):333–364. https://doi.org/10.3390/digital2030019.

Tsirogiannis, Constantinos, and Christos Tsirogiannis. 2016. 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*, edited by Tom Brughmans, Anna Collar, and Fiona Coward, pp. 103–120. Oxford University Press, Oxford.

Watson, Peter, and Cecilia Todeschini. 2007. *The Medici Conspiracy: The Illicit Journey of Looted Antiquities—From Italy's Tomb Raiders to the World's Greatest Museums*. Public Affairs, New York.

Weischedel, Ralph, Martha Palmer, Mitchell Marcus, Eduard Hovy, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, et al. 2013. *OntoNotes Release 5.0*. Linguistic Data Consortium, Philadelphia. https://catalog.ldc.upenn.edu/LDC2013T19, accessed July 27, 2022.

Wilcke, Xander, Victor de Boer, Maurice de Kleijn, Frank van Harmelen, and Henk Scholten. 2019. User-Centric Pattern Mining on Knowledge Graphs: An Archaeological Case Study. *Journal of Web Semantics* 59. https://doi.org/10.1016/j.websem.2018.12.004.

Yates, Donna. 2012. Sipan. *Trafficking Culture Encyclopedia*, August 17. https://traffickingculture.org/encyclopedia/case-studies/swetnam-drew-kelly-smuggling-of-objects-from-sipan/, accessed July 13, 2022.

Yates, Donna. 2016. Museums, Collectors, and Value Manipulation: Tax Fraud through Donation of Antiquities. *Journal of Financial Crime* 23(1):173–186. https://doi.org/10.1108/JFC-11-2014-0051.

Yates, Donna. 2020. Machaquilá Stela 2 . *Trafficking Culture Encyclopedia*, August 3. https://traffickingculture.org/encyclopedia/case-studies/machaquila-stela-2/, accessed July 13, 2022.

Zhu, Chaoyu, Zhihao Yang, Xiaoqiong Xia, Nan Li, Fan Zhong, and Lei Liu. 2022. Multimodal Reasoning Based on Knowledge Graph Embedding for Specific Diseases. *Bioinformatics* 38(8):2235–2245. https://doi.org/10.1093/bioinformatics/btac085.

AUTHOR INFORMATION

Shawn Graham, Chantal Brousseau, Jonah Ellens, and **Callum McDermott** ■ Department of History, Carleton University, Ottawa, Ontario, Canada

Donna Yates ■ Faculty of Law, Maastricht University, Maastricht, Netherlands (d.yates@maastrichtuniversity.nl, corresponding author)

Ahmed El-Roby ■ School of Computer Science, Carleton University, Ottawa, Ontario, Canada