Bridging Representation and Visualization in Prosopographic Research: A Case Study

Alessandra Urbinati [1,*], Enrico Burdisso [1], Arthur Thomas Edward Capozzi Lupi [1], Claudio Mattutino [1], Salvatore Vilella [1], Alfonso Semeraro [1], Giancarlo Ruffo [4], Carlo Corti [3], Stefano De Martino [2], Elena Devecchi [2], Erica Scarpa [2], Giulia Torri [3] and Rossana Damiano [1,*]

[1] *University of Turin, Italy, Dipartimento di Informatica*
[2] *University of Turin, Italy, Dipartimento di Studi Storici*
[3] *University of Florence, Italy, Dipartimento di Storia, Archeologia, Geografia, Arte e Spettacolo*
[4] *University of Eastern Piedemont "A. Avogadro", Dipartimento di Scienze ed Innovazione Tecnologica*

Abstract

In the last decade, the research on ancient civilizations has started to rely more and more on data science to extract knowledge on ancient societies from the written sources delivered from the past. In this paper, we combine two well-established frameworks: Linked Data to obtain a rich data structure, and Network Science to explore different research questions regarding the structure and the evolution of ancient societies. We propose a multi-disciplinary pipeline where, starting from a semantically annotated prosopographic archive, a research question is translated into a query on the archive and the obtained dataset is the input to the network model. We applied this pipeline to different archives, a Hittite and a Kassite collection of cuneiform tablets. Finally, network visualization is presented as a powerful tool to highlight both the data structure and the social network analysis results.

Keywords

Knowledge Graphs, Social Network Analysis, Digital Humanities, Network Science

1. Introduction

In the last decade, the research on ancient civilizations has started more on more to rely on data science to investigate social and cultural aspects of these civilizations from the written sources delivered from the past. In particular, a line of research has addressed the use of social network analysis (SNA) for studying the social and political structures of the past by leveraging the mention of entities such as locations and personages in texts. The advancements in network analysis for historical research have paved the way to novel methods for investigating the structure of communities from texts and their relationship with personal and geographical data, with proof of concepts ranging in time and space [1, 2, 3, 4, 5].

In parallel with this trend, the advent of Linked Data has made semantic resources available for archaeological and historical research, with notable examples such as CRM Archaeo [6] and

1st Italian Workshop on Artificial Intelligence for Cultural Heritage (AI4CH22), co-located with the 21st International Conference of the Italian Association for Artificial Intelligence (AIxIA 2022). 28 November 2022, Udine, Italy.
*Corresponding authors.
alessandra.urbinati@unito.it (A. U.); rossana.damiano@unito.it (R. Damiano)

FPO [7]. Today, the use of semantic representation techniques and SNA methods can provide an integrated approach which combines the shared, unambiguous definition of entities and relationships in the historical domain with the tools and methods made available by SNA. Thanks to the visualization of the social networks created from the semantically annotated historical data, the implicit social and political structures of the past can be appraised in a more insightful way by the domain experts.

In this paper, we describe a multi-disciplinary pipeline where, starting from a semantically annotated prosopographic archive, a research question is translated into a query to the archive, and the obtained dataset is the input to a SNA component, which outputs a graph and the corresponding visualization. The pipeline is being developed as part of a research project entitled “Networks of Power: Institutional Hierarchies and State Management in Late Bronze Age Western Asia (NePo)”, funded by the Italian Ministry of Research, which concerns the study of networks operating at local and regional levels in the Mesopotamian-Anatolian region. At the this stage of the project, the complete, yet not fully automated pipeline has been created and tested on the prosopographic data gathered so far from two corpora of cuneiform texts: the Hittite archive including data on 353 personages, and Kassite archive including data on 236 personages.

The paper is structured as follows. In Section 2, we introduce the domain and survey the applications of network analysis to historical research. In Section 3, the representation of prosopograhic data and the creation of the datasets; Section 4 describes in more detail the network analysis tasks. In Section 5, network analysis and visualization is described through a set of case studies. Conclusion and future work end the paper.

2. Ancient Near-Eastern Prosopography

The NePo project aims at a detailed analysis of Late Bronze Age (LBA) court structures, their internal networks, and the economic systems they were controlling, on the basis of selected epigraphic and archaeological sources, with the ultimate goal of obtaining a comparative picture of the royal elites of LBA Western Asia through the analysis of multiple sources from the geographical areas under consideration.

Late Bronze Age documents dealing with the administration and economy of the Near Eastern kingdoms and polities differ as regards their typology, contents, and aims. The corpus of the Kassite cuneiform tablets includes royal inscriptions, internal and international correspondence, and legal and administrative documents. Although Kassite material mostly comes from the city of Nippur, the project case study focuses on a corpus of ca. 800 Kassite-period cuneiform tablets published by [8] and dating to the 14th-13th centuries BC and whose provenance is not certain. These documents currently represent the second largest, internally coherent set of Kassite-period sources after that of Nippur. They are administrative records mainly dealing with the income, storage, and redistribution of agricultural products (mostly cereals, but also sesame, pulses, and cress) and by-products (beer and flour), animal husbandry, and textile production; smaller groups of texts include legal documents and letters.

The Hittite documentation includes very few administrative and economic records [9] but is very informative as far as the governance of the state and the role played by the court and the officials are concerned. Differently from Mesopotamia, in fact, where very large archives of administrative and economic records have been discovered and studied, little is known about administration in Hittite Society. The corpus of Hittite written documents mainly consists of the tablets and sealings found at Hattusa and other Anatolian sites such as Maşat Höyük/-Tapikka, Kuşaklı Höyük /Sarissa, Kayalıpinar/Samuha, and Ortaköy/Sapinuwa. The research takes also into consideration the documents from the Syrian polities subordinated to Hatti, namely, Karkemish, Ugarit, Alalah, and Emar. Hittite written documents and the sealings mention a huge number of personal names, titles and professions that await to be fully studied.

In order to overcome the intrinsic limits posed by the fragmented nature of the evidence, the project is characterised by an interdisciplinary approach that combines the traditional methods of philology, archaeology, prosopography and historical research with methods and tools developed in the field of the digital humanities, such as: a factoid-based approach for developing consistent models that schematically describe the activities of a target group with direct connections with the sources; social network analysis for studying and comparing the network of relations through the ages and geographical places even in presence of datasets of different size [1, 2]; data visualization tools that can enhance the interpretation of data by effectively displaying the results of quantitative and quantitative queries.

In the state of the art, the use of SNA techniques in the study of cuneiform archives relies on two main approaches, which consist, respectively, in using the social network to infer new information about the nodes (e.g., merging nodes representing the same individual [10]), and in inspecting the properties of the network on global scale (e.g., clustering nodes based on their properties and relations [11, 1] to discover and confirm working hypothesis about social structures). The NePo projects aims at overcoming the limitations pointed out for previous approaches: in particular, the size of the available data, and lack of standardization of the semantic labels employed for the annotation. The adoption of a semantic model for describing the data partly overcomes these limitations, since it paves the way to the integration of new data sets and the application of more advanced SNA techniques.

3. Representing Prosopographies

The representation of prosopography in the project relies on the pioneering work carried out by Pasin and Bradley 2015. The Factoid-based Prosopographic Ontology[1] revolves around the notion of factoid, intended as a believed-to-be-true, reported event in some written source, a definition that fits very precisely the data inferred from the corpora investigated by the project. The Factoid model puts into play two basic entities connected with the notion of factoid: the Source, where the factoid is asserted (a Hittite or Kassite cuneiform text), and the Relation it describes (e.g., an administrative or legal relation), further specialized in a Person reference and in a Location reference (respectively, the personages and places involved in the relation). Starting from FPO, classes and properties have been specialized to adapt them to the Hittite and Kassite domains (thus yielding the Hittite FPO, or HFPO, and the Kassite FPO, or KFPO), by adding more specific classes and properties to the ones in the ontology, leaving the the core model unaffected. In particular, Cuneiform texts and Seals have been introduced as specific

[1]https://github.com/johnBradley501/FPO

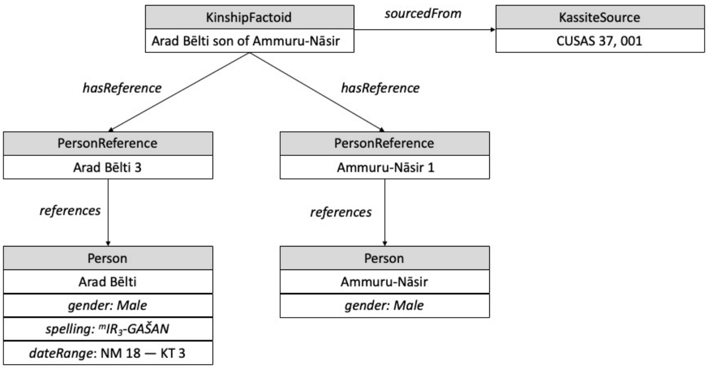


Figure 1: Representation of a kinship relationship in the Kassite dataset.

source types; specific factoid types have been added to account for the information conveyed by the cuneiform texts in the two domains, which concern Administrative, Kinship, Professional and Legal relationships between the personages in the Hittite and Kassite world; finally, new authoritative lists have been created to fit the representation of sources and personages in the Hittite and Kassite worlds, where dates refer to sovereigns, and people are described according to culture-specific terms for age and gender. Figure 1 illustrates an example factoid, namely the kinship relationship between two personages sourced from a cuneiform text from Kassite Babylonia.

The collection of Hittite person records is in the project carried out by using an installation of the Omeka S platform, which allows collecting, publishing and sharing data with Linked Open Data. The core of Omeka S is a relational database, but records can also be ingested and accessed by third parties via an API in JSON-LD format (Fig. 2). The format of a record consists of a template, which defines the record's metadata scheme by combining the terms from the RDF vocabularies uploaded to the platform; records can be interlinked when needed, so as to represent straightforwardly the relations over classes encoded in the ontology. The representation of factoids, personages and sources in the Hittite and Kassite domain, in fact, relies on a set of vocabularies representing the main entities of the domain: besides the Factoid-based Prosopographic Ontololgy (FPO) for the prosopographic domain (extended for the case of Hittite and Kassite prosopography as sketched above), the Bibliographic Ontology vocabulary (BIBO) for the relations with bibliographic sources, and the Dublin Core Metadata Element Set (DC-MES) for the provenance and description of data.

4. Social Network Analysis

Network models conceptualize interactions between entities and allow to gain both a local and a global view of the system. We define a network as $G = (V, E)$, where V is a set of nodes, and E is the set of links that encode the relationship structure between nodes. This definition can be enriched to improve the adherence of the model to reality. In particular, a network

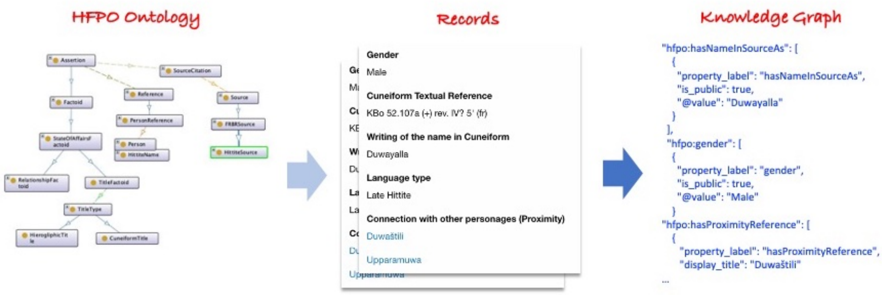


Figure 2: The role of Omeka-S (center) in bridging the ingestion of data into Linked Data (right), according to the Factoid-based model (left)

can be defined as directed, if the edges retain the information of direction; weighted, where the function $\varpi : V \times V \rightarrow \mathbb{N}$ assigns for each pair of nodes $i, j \in V$ the weight of edge (i, j); temporal, having specific representation for each timestamp T. Moreover, a network model can be dynamic [12], and can grow over time [13]. Another well established class is the bipartite network whose nodes are divided into two sets, and only connections between two nodes in different sets are allowed [14]. Since links often exhibit heterogeneous features, new structures were theorized, with layers [15] or multiedges in addition to nodes and links [12]. In the more general multilayer framework, a node i in layer a can be connected to any node j in layer b. Layers will represent aspects or features that characterize the nodes or the links that belong to that layer. As a consequence, the set of links can be partitioned into "intra-layer links", that is links that connect nodes set in the same layer, and "inter-layer" which are those that connect nodes set in different layers [16]. This flexibility allows applying the framework on multiple tasks.

Information is rooted in the "connectedness structure" of the network and to exploit it researchers must decide which resemblances, relationships between entities, or nodes encoding. This relates to the kind of aspect of the data to be included as well as their relative importance. Different computational methods aim to preserve those properties the researchers want to study and replicate.

Content exploration. Entity co-occurrence in text-analysis has been extensively explored when dealing with ancient text, for example, Bornhofen et al. [17] employ a corpus of digitized resources about European integration since 1945 to generate a dynamic multilayer network that represents different kinds of named entities appearing and co-appearing in the collections. Then they build a visualization tool that allows interest-driven navigation in the corpus and the discovery of the interconnections of its entities across time [17]. Another fundamental research question regards annotation comparison. Wieneke et al. [18] have developed a tool to establish false negatives in the depiction of the same entities in the context of annotation comparison.

Society structure. Social networks describe the connection of individuals. To better understand the hidden structure of these types of networks fundamental tasks are identifying the most central nodes, and understanding the organization of nodes into groups. Centrality measures and resulting node raking target the scope to establish the main player/actor in the

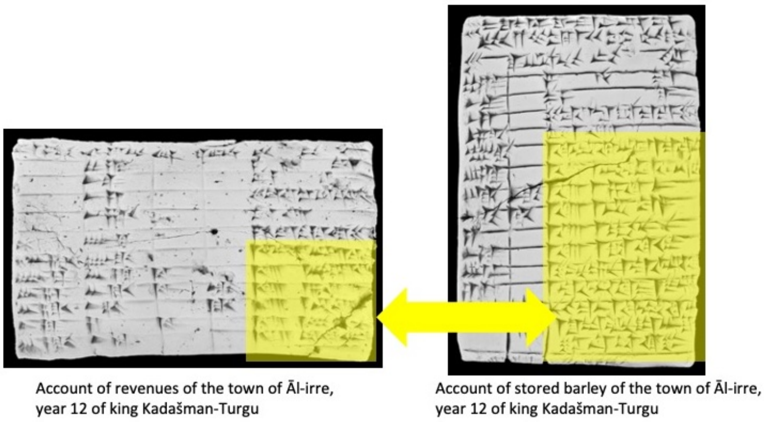


Figure 3: Excerpts of cuneiform texts referencing the same location (the town of Āl-irre)

network [19]. While clustering or community detection tackles the problem to establish partition of the network [20]. Studying social roles could also deeply impact the understanding of how society and in particular ancient ones worked, and analyzing the interplay of political and economic relationships can shed light on hierarchy and power dynamics. Breigher at al. [21] famously set the basis for many other works coupling the marriage and the economic trading between fifteenth-century Florence families.

Spatial network. Network model has been used to analyze hierarchical predominance in cultural practices across different regions and the evolution of cultural trends. Mizoguchi et al. [22] establish links between ten regional entities whenever the author found archaeologically recognizable similarities in pottery styles and mortuary traditions. The work by Schich et al. [23] aims to understand which processes shape and drive the geopolitical aspect of cultural history by a birth-to-death network, where nodes are countries and links represent the migration of notable individuals over time from birth to death places. Spatial and social entities could interact and discovering interesting motifs in the network could lead to new insights into the life of ancient empires [24].

5. Bridging KG and SNA

The pipeline starts with the ingestion of the data into the Omeka S platform and ends with a visualization of a network generated from the data to explore them or answer a specific research question. The pipeline, which involves the roles of domain expert (*historian* and *archaeologist*), *knowledge representation engineer* and *data scientist*, consists of the following steps:

1. **Record creation**: this step, carried out by a domain expert (namely, an *archaeologist* or a *historian*), starts with the ingestion of the data extracted from the cuneiform texts encoded in the clay tablets. This is where the interpretation of data first occurs, since experts are called to identifies the entities mentioned in the cuneiform text, often encoded with different spellings, and the relations between them. An example is provided by Fig-

ure 3, which represents two cuneiform texts mentioning the same location in different contexts. This step does not require any familiarity with the Linked Data formats, since the ingestion is carried out via a set of interlinked web forms. Before the data gathering, in fact, the interconnected record types representing the entities involved in the interpretation of data (personages, locations, sources, etc.) have been created by the *knowledge engineer*, in cooperation with the *domain experts*: for example, with reference to Fig. 1, the Kinship Factoid record type (template in Omeka S terminology), the Person record type and the Source record type are involved.

2. **Triplification**. In this phase, carried out by the *knowledge engineer*, the archive (or a portion o it) is extracted in JSON-LD format from the platform via the Omeka S built-in REST APIs[2] and the resulting knowledge graph is stored into an Apache Fuseki[3] triple store. In practice, each record in the platform is mapped onto the corresponding entity (e.g., factoid type, sources, person) in the ontology, and the entities are connected to each other according to the ontology model; the record fields provide the triples describing the entity. Notice that, differently from a relational scheme, this mapping is straightforward, since records are directly encoded in the vocabulary provided by the prosopographic ontology. For example, the kinship relation between two personages, a kinship factoid (instance of the `KinshipFactoid` class), relates the two personages (instances of the `Person` class) and the cuneiform text excerpt where their relationship is represented (an instance of the `Source class`); additional triples – such as the specific kinship relation type, the age and gender of the personages, or the reference location in the source text – are added depending on the details available in the record.
3. **Data extraction**. Although in principle Omeka S allows formulating queries in a semantic format by using the vocabularies employed to describe the data, in practice this strategy falls short to translate the most complex research questions into queries on the archive. To bypass this limitation, the research questions formulated by the *domain experts* (historian and archaeologist) are translated by the *knowledge engineer* into SPARQL queries and executed on the knowledge graph. In order to reduce the interpretation gap between the domain experts and the datasets, a form-based interface for preparing the queries has been created. allows non-experts to query the graph by selecting the desired characteristics of the personages and obtaining the corresponding SPARQL query (in the figure, the personages related with the town of Nippur). Similar interfaces have been created for the other relevant entities of the domains, such as the sources and the relations between the personages.
4. **Visualization**. Finally, a network analysis algorithm is applied by the *data scientist* to the data extracted from the dataset through the SPARQL query, and the obtained graph is visualized and submitted to the *domain expert* for interpretation, with the goal of confirming or rejecting the assumptions behind the research question.

In the following, we describe three different research questions that have been explored by employing the pipeline described above.

[2]https://omeka.org/s/docs/developer/api/rest_api/
[3]https://jena.apache.org/documentation/fuseki2/

R1. *Investigating the co-occurrence between persons and locations* sourced from the Kassite document collection. The raw data, as shown in Figure 3, has been ingested into records and stored in the knowledge graph. Through the query 5, we filter the information needed. For this task, we defined a bipartite network $G = (U, V, E)$, where U is the set of Kassite persons, V the set of locations and E the set of edges (i, j) that exist between nodes $i \in U$ and $j \in V$ if the person i appear in some activity in location j. The network is shown in Figure 4a. Further analysis involves both the possible projections of the bipartite network. A projection is a compressed version of the bipartite network that contains nodes of only either of the two sets, nodes are connected only when they have at least one common neighboring in the other set. Figures 4b and 4c show the projections obtained with simple weighting, that is where edges are weighted by the number of times a common association between two nodes of the same set with the same node of the other set is repeated. The projection networks can be leveraged to gain a deeper view of the twofold system.

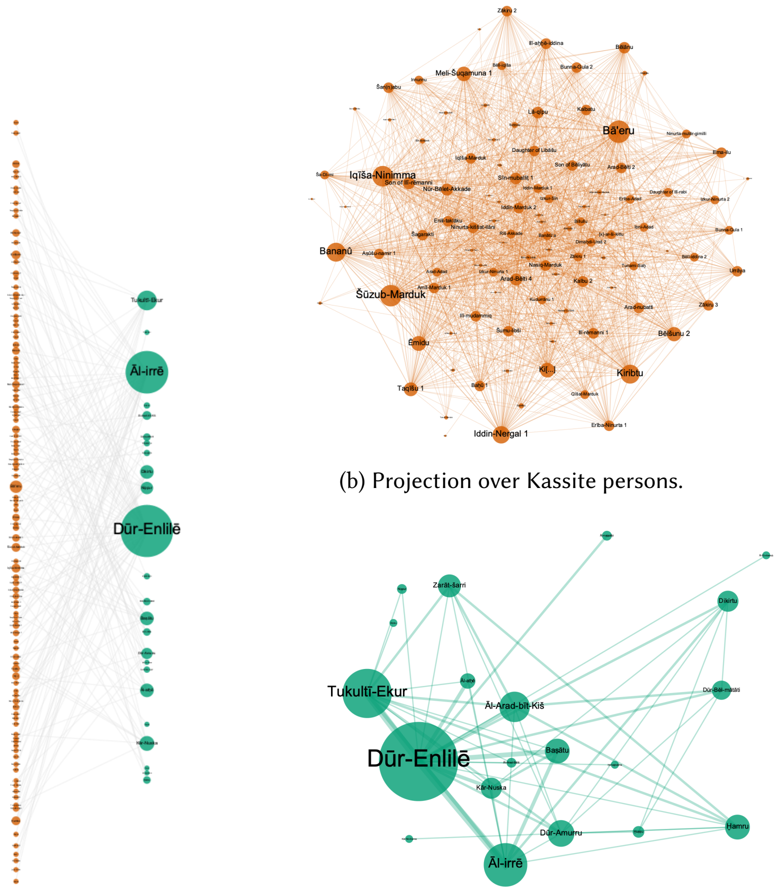
```
SELECT ?id  ?name ?location
WHERE {
   ?person rdf:type kfpo:KassitePerson .
   ?person  dcterms:title ?name.
   ?person  omeka:id ?id.
   OPTIONAL{?person kfpo:hasLocationName ?location}
}
```

R2. *Investigating the geographical trade* – in particular, how a specific administrative role distributes over the cities of the Hittite Empire. We filter the data through the query 5: we retain all the persons in the late Hittite period that appear as witnesses in some transactions. We define the network to be undirected, weighted with colored node, $G = (V, \varpi)$, where V is the set of Hittite persons and $\varpi : V \times V \rightarrow \mathbb{N}$ is the function defining for each pair of nodes $i, j \in V$ the weight of edge (i, j) that measures the times the nodes appear as witnesses in the same trade. A network is said to be *colored* if we associate colors, or labels, to each of its vertices and/or edges. We assign as labels the city to which the person belongs. Discovering motifs in the resulting network would mean discovering interesting trade routes that were used in the Empire.

```
SELECT ?name ?link ?to
WHERE {
   ?link dcterms:title ?name;
   hfpo:hasWitnessReference ?to.
   ?link hfpo:sourceLanguage "Late Hittite".
}
```

R3. *Understanding how different aspects of society relate to each other.* To investigate the role duality in the Kassite Empire, namely how kinship relations impact administrative relations, after querying the data (a partial query is shown in Listing 5) we define a multilayer network, where each node i can belong to a different layer α. In this context, we define two layers, one that encodes all the kinship relationships and one for the administrative relationship. Figure 6 shows the flattened version of the described network. Multilayer networks combine multiple dynamics, that can be embedded both inside each layer and between them. To fully exploit this type of structure, expertise in ancient cultures and societies must guide the final model definition.

```
SELECT ?name ?type ?entities
WHERE {
   ?kin rdf:type kfpo:AdministrativeRelationshipFactoid .
   ?kin dcterms:title ?name.
   ?kin kfpo:administrativeRelationType ?type.
   ?kin kfpo:hasAdministrativeReference ?entities
}
```



(b) Projection over Kassite persons.

(a) Bipartite network.

(c) Projection over Kassite Empire locations.

Figure 4: Kassite co-occurrence person-location bipartite network and projections. (b) represents the network projection over the persons ($|U| = 108$, $|E| = 1475$). (c) represents the network projection over the locations ($|U| = 20$, $|E| = 51$). Node dimensions scale over weighted degree.

6. Discussion and Conclusion

In this paper, we described a pipeline developed to support the study of ancient societies with visualizations developed from a semantic representation of prosopographic data. Starting from two related, yet distinct domains, we implemented the steps in the pipeline to test its feasibility and methodological validity for the archaeologists and historians involved in the project. The methodology and the obtained visualizations, designed by a multi-disciplinary team involving knowledge representation experts, digital humanists, historians and visualization experts, have been positively assessed by the domain experts, who have been able to confirm the research hypotheses behind the creation of the networks.

As future work, we plan to fully automatize the pipeline, so that the extraction of data, the construction of the network and its visualization can be executed in the same environment.

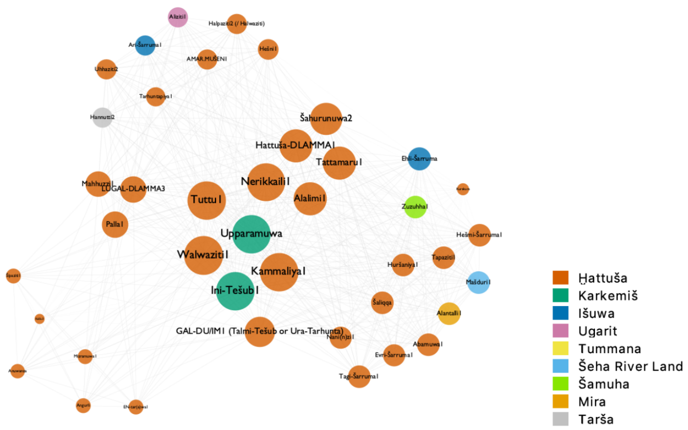


Figure 5: Late-period Hittite witness relationships. Nodes ($|U| = 40$) are Hittite persons; dimension scales over the weighted degree; color schema is location. Edges ($|E| = 449$) stand for two persons who appeared as witnesses in the same transaction.

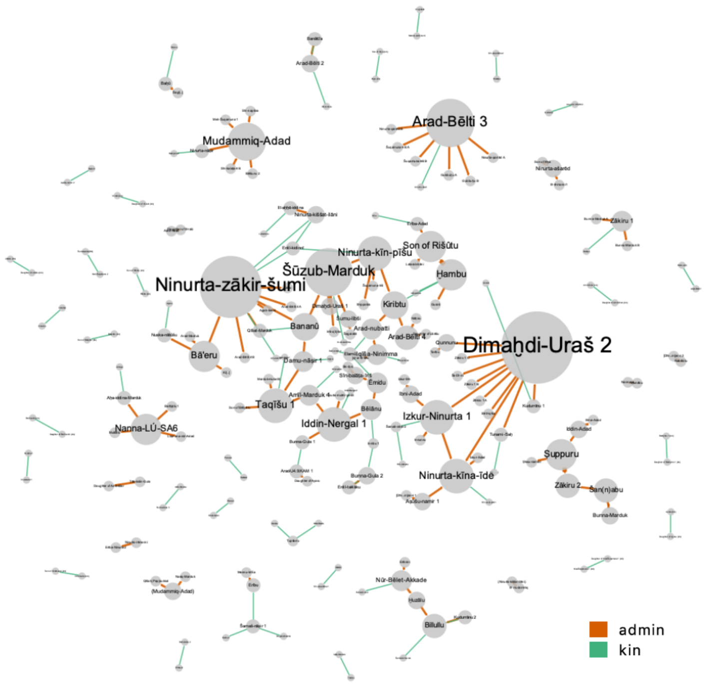


Figure 6: Kassite kinship and administrative relationships network. Nodes are Kassite persons ($|U| = 208$), their dimension scale over weighted degree. Edges ($|E| = 168$) colors stand for relationship types.

References

[1] A. Wagner, Y. Levavi, S. Kedar, K. Abraham, Y. Cohen, R. Zadok, Quantitative social network analysis (sna) and the study of cuneiform archives: A test-case based on the murašû archive, Akkadica 134 (2013) 117–134.

[2] V. A. Traag, L. Waltman, N. J. Van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Scientific reports 9 (2019) 1–12.

[3] T. Alstola, S. Zaia, A. Sahala, H. Jauhiainen, S. Svärd, K. Lindén, Aššur and his friends: a statistical analysis of neo-assyrian texts, Journal of Cuneiform Studies 71 (2019) 159–180.

[4] H. De Weerdt, B. Ho, A. Wagner, J. Qiao, M. Chu, Is there a faction in this list?, Journal of Chinese History 中國歷史學刊 4 (2020) 347–389.

[5] E. Hyvönen, P. Leskinen, M. Tamper, H. Rantala, E. Ikkala, J. Tuominen, K. Keravuori, Linked data: A paradigm shift for publishing and using biography collections on the semantic web, in: Proceedings of the Third Conference on Biographical Data in a Digital World (BD 2019), CEUR-WS. org, 2022.

[6] P. Ronzino, Harmonizing the crmba and crmarchaeo models, International Journal on Digital Libraries 18 (2017) 253–261.

[7] M. Pasin, J. Bradley, Factoid-based prosopography and computer ontologies: towards an integrated approach, Digital Scholarship in the Humanities 30 (2015) 86–97.

[8] E. Devecchi, Middle Babylonian Texts in the Cornell Collections, Part 2, Penn State University Press, 2020.

[9] T. Van den Hout, Administration and writing in hittite society, in: Balza, ME; Giorgieri, M.; Mora, C.(a cura di), Archivi, depositi, magazzini presso gli Ittiti. Nuovi materiali e nuove ricerche= Proceedings of the Workshop held at Pavia, 2009, pp. 41–58.

[10] D. Bamman, A. Anderson, N. A. Smith, Inferring social rank in an old assyrian trade network, arXiv preprint arXiv:1303.2873 (2013).

[11] C. Waerzeggers, Social network analysis of cuneiform archives: A new approach, Documentary Sources in Ancient Near Eastern and Greco-Roman Economic History: Methodology and Practice (2014) 207–233.

[12] M. Newman, A.-L. Barabasi, D. J. Watts, The structure and dynamics of networks, volume 12, Princeton University Press, 2011.

[13] A. Li, S. P. Cornelius, Y.-Y. Liu, L. Wang, A.-L. Barabási, The fundamental advantages of temporal networks, Science 358 (2017) 1042–1046.

[14] T. Zhou, J. Ren, M. Medo, Y.-C. Zhang, Bipartite network projection and personal recommendation, Physical review E 76 (2007) 046115.

[15] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, M. A. Porter, Multilayer networks, Journal of complex networks 2 (2014) 203–271.

[16] A. Aleta, Y. Moreno, Multilayer networks in a nutshell, Annual Review of Condensed Matter Physics (2018).

[17] S. Bornhofen, M. Düring, Exploring dynamic multilayer graphs for digital humanities, Applied Network Science 5 (2020) 1–13.

[18] J. Novak, I. Micheel, M. Melenhorst, L. Wieneke, M. Düring, J. G. Morón, C. Pasini, M. Tagliasacchi, P. Fraternali, Histograph – a visualization tool for collaborative analysis of networks from historical social multimedia collections, in: 2014 18th International

Conference on Information Visualisation, 2014, pp. 241–250. doi:10.1109/IV.2014.47.
[19] K. Das, S. Samanta, M. Pal, Study on centrality measures in social networks: a survey, Social network analysis and mining 8 (2018) 1–11.
[20] S. Fortunato, Community detection in graphs, Physics reports 486 (2010) 75–174.
[21] R. L. Breiger, P. E. Pattison, Cumulated social roles: The duality of persons and their algebras, Social networks 8 (1986) 215–256.
[22] K. Mizoguchi, Nodes and edges: a network approach to hierarchisation and state formation in japan, Journal of Anthropological Archaeology 28 (2009) 14–26.
[23] M. Schich, C. Song, Y.-Y. Ahn, A. Mirsky, M. Martino, A.-L. Barabási, D. Helbing, A network framework of cultural history, science 345 (2014) 558–562.
[24] P. Ribeiro, F. Silva, Discovering colored network motifs, in: Complex networks V, Springer, 2014, pp. 107–118.