A Social Network of the *Prosopography of the Neo-Assyrian Empire*

HEIDI JAUHIAINEN (iD)

TERO ALSTOLA (iD)

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

]u[ubiquity press

ABSTRACT

The dataset is a social network of over 17,000 individuals who lived during the so-called Neo-Assyrian period of Mesopotamian history, primarily in the eighth and seventh centuries BCE. The undirected network of individuals connected by co-occurrences in cuneiform documents was semi-automatically extracted from the *Prosopography of the Neo-Assyrian Empire*. In addition to two weighted versions of the one-mode co-occurrence network, the dataset also contains a two-mode person-text network and rich metadata for each individual. For the first time, the dataset allows large-scale computational analysis of social structures in the Assyrian Empire. The data is primarily stored as plain text and CSV files, inviting scholars to further expand and enrich it. The scripts and files used for creating and standardizing the data are also available in the Zenodo repository.

CORRESPONDING AUTHOR:

Heidi Jauhiainen

Department of Digital Humanities, University of Helsinki, Helsinki, Finland

heidi.jauhiainen@helsinki.fi

(1) OVERVIEW
REPOSITORY LOCATION

https://doi.org/10.5281/zenodo.5862904

CONTEXT

This dataset is a social network of over 17,000 individuals attested in cuneiform documents from the Neo-Assyrian period, primarily in the eighth and seventh centuries BCE. The data originates from the *Prosopography of the Neo-Assyrian Empire* (PNA), which has been available only as a printed edition (Radner & Baker, 1998–2011). Heather D. Baker's Prosopography of the Neo-Assyrian Empire Online project[1] has the long-term goal of making the PNA data digitally available, but it currently provides only additions and corrections to the printed PNA volumes.

We extracted the data from the text and PDF file versions of the PNA that we received from Simo Parpola, the Editor-in-Chief of the series. The dataset published here was produced for the research purposes of the Centre of Excellence in Ancient Near Eastern Empires (University of Helsinki).

(2) METHOD
STEPS

The earlier PNA volumes (1/I–3/I) were available to us as plain text files that were used to typeset the printed publications. As the last volume (3/II) was laid out using different software, it was available to us only as a PDF file. We wrote a number of scripts in Java to extract and process the data. Their source codes are available in the repository.

Entries in the PNA consist of two parts (see *Figure 1*): 1) an entry starts with a personal name and its linguistic analysis; and 2) then follows a list of the individuals who used this name. In some cases, there is only a single person who used a name, whereas some names were borne by dozens of people. It is often not entirely clear if two or more cuneiform texts refer to the same person or several homonymous individuals. In the PNA, the decision to connect an attestation of a name to a historical individual is made by a trained Assyriologist. We followed these identifications but could not take their level of uncertainty into account, although this is sometimes expressed in the PNA. For each individual, the PNA provides a short description (e.g., "Individual from Assur (reign of Sennacherib)" or "Tammaritu II, king of Elam c. 652–649 (reign of Assurbanipal)"), followed by the attestations of this person in texts and a short description of the person's role in each text.

The dataset was produced in four steps. In steps one and two, we worked with the text files, because their structure was ideal for automated processing. First, we extracted the names of the Neo-Assyrian cuneiform documents in which persons are attested. The document names were indicated by "@@" at the beginning of a line in the text files (see *Figure 1*). Since the PNA was published over a period of thirteen years and the entries were written by numerous scholars, there are inconsistencies in how the documents are referred to. Several typing errors were also detected. Furthermore, only some of the document names indicate where the name of the document ends and a possible line number starts. Using rules and reference lists, we compiled a list that connects standardized document names to the original "@@" lines in the text files.

In step two, individual persons were extracted from the text files with the name of each document they were attested in. It was not possible to collect information about the individual's role in a document, because this is given in unstructured running text. Homonymous individuals were distinguished by consecutive numbers added after their names (e.g., Aššūr-iddin_1), following the numbering of individuals in the PNA. The general description and dating of each individual as well as the language and gender of the name were also extracted. It can be very difficult to assign a language and gender to an ancient name (PNA 1/I, p. xxii; Földi, 2019), and some analyses given in the PNA could be contested. As with all other data, we simply extracted the information as it is in the PNA.

```
 *A@@u_r-re_mu_ti_
 ("A@@ur is my mercy"); Akk.; masc.;
 wr. [1]=a@-@ur--rem2-ut-tu, 1=a@-@ur--rem2-u-te;
 AHw 971.


 Individual from Assur (reign of Esarhaddon and early reign of
 Assurbanipal):
 1=<a@-@ur?>--rem2-u-t[e]
 acts as a witness
 @@As8617b S002 (675);
 [1]=a@-@ur--<rem2-ut?>-tu
 is mentioned in a broken legal document of unclear function
 @@SAAB 5 40 002 (date lost);
 The name of a witness acting for Abba^,
 1=a@-@ur--re,
 is incomplete due either to a mistake of the writer of the tablet or the
 modern copyist (the tablet is unaccessible). The name is followed by the curious phrase
 @a AN`E=KUR.RA-ME` ta-da-na-ma.
 @@Rfdn 17 05 S001 (666).
 Even though the restoration as A@@u_r-re_mu_ti_ seems to be likely because of
 the space available and chronological reasons, there are other possibilities:
 A@@u_r-re_htu-u$ur (but the name seems to be too long), A@@u_r-re_manni
 (in a rarely used spelling variant) or A@@u_r-re_$u_wa (as suggested by
 Ahmad [1996] 223. However, the only attestation of the latter in Assur is as the
 recipient of the letter
 @@As13846q).


 <<E. Frahm


 *A@@u_r-re_$u_wa
 ("A@@ur is my helper"); Akk.; masc.;
 wr. 1=a@-@ur--re-$u-u-a, 1=a@-@ur--re-$i-u-a;
 AHw 972a; Saporetti (1970) 138 (MA A@@u_r-re_$u_ja); Tallqvist (1918) 45.


 1. Eunuch and scribe of the queen, from Kalhu (reign of Sargon II):
 1=a@-@ur--re-$u-u-a LU2=SAG LU2=DUB.SAR @a MI2=E2.GAL
 serves as a witness when Nabu^-kabti-ahhe@u, Sargon's palace scribe,
 buys land, presumably the area on which Dur-`arruken is to be erected
 @@TCL 9 58 R014 (709).


 2. High-ranking intelligence agent based in Kumme reporting on Urar#ian activities
 (reign of Sargon II):
 Most of his letters are concerned with information about the country of Urar#u,
 Assyria's rival in the north, the surveillance of which he seems to
 coordinate. He is based in Kumme, a town in the north ruled at that time by
 Arije and Ari$a^. Located there also is a royal delegate ({qe_pu|) whose
 identity is unknown (see
 @@CT 53 138 and
 @@CT 53 098 below).
 It is very probable that A@@ur-re$uwa, who is mentioned in both these letters,
 is this delegate.
 Whether he survived the revolt in Kumme or not (see below), is not clear from
```

Figure 1 A sample extract from the PNA text files. The standardized names of these individuals are: Aššūr-rēmūtī_1, Aššūr-rēṣūwa_1, and Aššūr-rēṣūwa_2. The processed metadata for Aššūr-rēṣūwa_1 in our dataset is "Akkadian, masc., 721–705 Sargon II, Eunuch, Kalhu."

In the third step, we extracted information from the PDF version of the last PNA volume. As preprocessing, we opened the file in a Safari browser and copied the text into Microsoft Word, an operation that kept the formatting of the font style. Entries for personal names and individuals could be identified and were marked before turning the file into plain text. Copying the text from two columns of the PDF file resulted in many inconsistencies, and the new text file was manually curated before information about documents and individuals was extracted. Entries for document names could not be identified in the structure of the PDF file, and the identification and extraction of documents is thus based on concordance lists and document names attested in other PNA volumes.

Finally, we created a two-mode network of persons and texts and a one-mode network of all individuals. Both networks are undirected. In the two-mode network, persons were connected to the texts in which they are attested. The one-mode network connects two persons if they are attested in the same document (*Figures 2* and *3*). Isolates (persons without any connections) were not included in the network. There are two versions of the one-mode network: one uses the number of co-occurrences as edge weights, and another one calculates the edge weights in relation to the total number of persons appearing in the same document (see below). As the PNA entries for Assyrian kings do not include all their attestations in the extant texts, we collected their attestations from the State Archives of Assyria book series and the corresponding online editions from the Open Richly Annotated Cuneiform Corpus (Oracc). We also supplied some metadata for individuals in the one-mode network. In addition to the unique id number, each

individual has the following attributes as far as they could be automatically extracted from the PNA files: personal name, language and gender of the name, profession or similar description, place of activity, ethnicity or geographic origin, period of attestation (corresponding to the reigns of Assyrian kings), and start and end dates for the period of attestation (for filtering and creation of timelines in networks). This metadata was extracted from the PNA files and standardized using several lists.
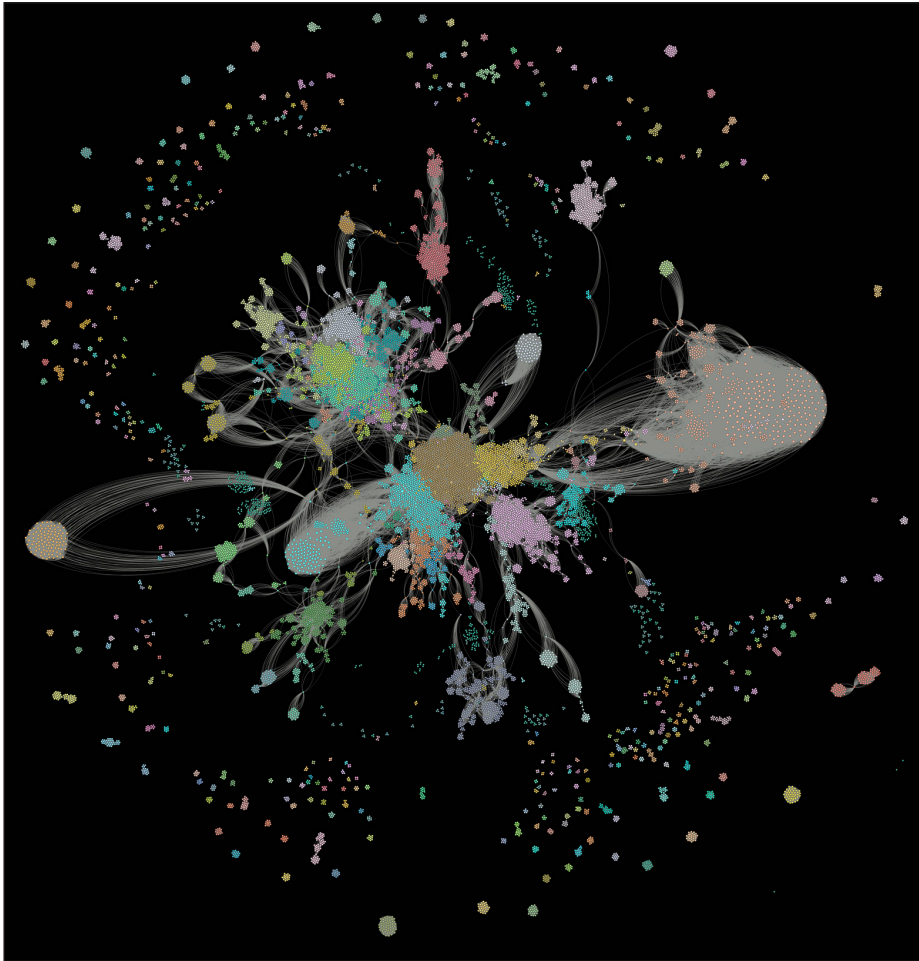

Figure 2 PNA one-mode network visualized with Gephi software (Bastian, Heymann, & Jacomy, 2009).

SAMPLING STRATEGY

We aimed to include all the individuals in the PNA, but we decided to ignore attestations in king lists and royal inscriptions. These are usually very long texts, and they cannot be used to automatically determine a real social interaction between two people who are attested in them. As a single royal inscription can cover several unrelated military campaigns and other political events, it cannot be assumed that the people appearing therein formed a social network or even knew one another. They are mentioned in the same inscription only because they were relevant to the Assyrian king as his allies, enemies, or servants. Letters can also refer to unrelated events, but the issue is less pronounced because of the letters' brevity and narrow scope in comparison to royal inscriptions. Letters are thus included in our dataset. Moreover, we left out instances where the name of an individual was used as an eponym for dating a document. Finally, some texts were ignored because it was not possible to determine if the document name refers to one or more texts.

QUALITY CONTROL

The quality of the extracted data was manually examined by the authors during the dataset creation, and recurring errors were corrected by changing the rules in our scripts and creating reference lists. Both document names and metadata on individuals were normalized using a number of rules and lists. The lists, scripts, and explanatory notes are available in the repository.

LIMITATIONS

Each entry in the PNA is a scholar's interpretation of the information found in one or more cuneiform texts. Our work is an interpretation of the PNA: we aimed to create a dataset that reflects the contents of the PNA as precisely as possible, not to update, correct, or supplement any information found therein. Because of limited resources, we chose to use an automated method to create our dataset. As a result, it is missing some information available in the PNA and is bound to contain errors.

For example, it was not possible to consider the nature of relationships between different people mentioned in a single text. This results in some false connections between people who had nothing to do with each other but just happened to be mentioned in the same text, such as a long list of officials. To alleviate this problem, we have created a one-mode network in which the edge weight between persons is relative to the total number of persons attested in the same document (the weighting method courtesy of Krister Lindén). The larger the number, the weaker the edge between two persons. The weight *w(c)* of the connection *c* between two persons is calculated as the sum of the weights in all documents *1...k* in which they are mentioned together. This can be formulated as

$$w(c) = \sum_{i=1}^{k} \frac{1}{|A|-1}$$

where *A* is the set of individuals mentioned in a document *i* and |A| is the size of the set. This weighting method is based on the premise that the larger the number of persons in a document, the smaller the probability that they all were in regular or intimate contact with one another. The sender of a letter had a reason to write to the addressee, while the people mentioned in an administrative list may have worked in the same institution but just known each other by name. While this model does not comply with Granovetter's (1973; 1983) suggestion that strong ties tend to form denser networks than weak ones, it follows his classic theory in assuming that the strength of interpersonal ties has a bearing on network structures on micro and macro levels.

Moreover, the role of a person in a text (debtor, recipient of a letter, etc.) and the nature and possible direction of interpersonal ties (family or business relationship, sender/recipient of a letter, etc.) could not be automatically extracted. In the case of unpublished texts and texts that were published after 1997, references to them are not consistent across the PNA volumes, and they could be semi-automatically standardized only to a certain extent. The dataset also lacks the corrections and additions to the PNA volumes made available in the Prosopography of the Neo-Assyrian Empire Online project. Despite our efforts in cleaning the text file converted from the PDF format (PNA 3/II), many conversion errors remain in the file.

(3) DATASET DESCRIPTION

OBJECT NAME

A Social Network of the *Prosopography of the Neo-Assyrian Empire*

FORMAT NAMES AND VERSIONS

BIN, CSV, Gephi, Java, MD, TXT

CREATION DATES

2019-03-01 – 2022-02-22

DATASET CREATORS

Heidi Jauhiainen (University of Helsinki) was responsible for research design, data extraction and validation, dataset creation, and software development. Tero Alstola (University of Helsinki) was responsible for research design, validation, and Assyriological research. Saana Svärd (University of Helsinki) was responsible for conceptualization and funding acquisition. Krister Lindén (University of Helsinki) was responsible for conceptualization and methodology. Repekka Uotila (University of Helsinki) worked on normalization and reference lists.

LANGUAGE

English, Akkadian; personal names in several ancient languages

LICENSE

Creative Commons Attribution-Share-Alike 3.0

REPOSITORY NAME

Zenodo

PUBLICATION DATE

2022-01-17

(4) REUSE POTENTIAL

One of the first – or perhaps even the first – studies in computational network analysis on historical data focused on Old Assyrian cuneiform texts (Gardin & Garelli, 1961; see also Plutniak, 2021). Recent studies have further shown the promise of social network analysis (SNA) in Assyriology (Waerzeggers, 2014; Still, 2019) and Neo-Assyrian studies in particular (Alstola et al., 2019; Jones, 2021). The present dataset will provide researchers with an exceptionally large and richly annotated network to work with (*Figures 2* and *3*). Such network data from the Neo-Assyrian period has not been openly available until now, and the dataset provides new opportunities for studies in the social structure of the Assyrian Empire. This does not need to be limited to network studies; the metadata on persons can also prompt other types of data analysis and quantitative research.
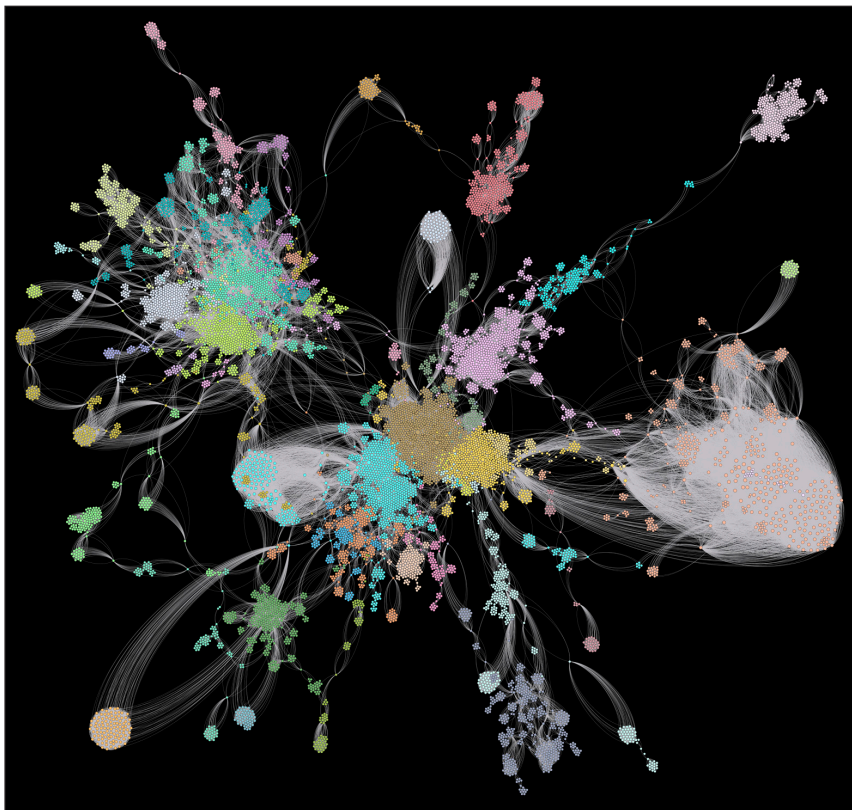
Figure 3 The giant component of the one-mode network visualized with Gephi software (Bastian, Heymann, & Jacomy, 2009).

Because the dataset is made openly available, there is great potential for future collaboration to validate, correct, and expand it. In addition to amending the existing dataset, this could include creation of new network data, such as a person's role in a document and the direction of interpersonal relationships (e.g., seller and buyer or sender and recipient of a letter). There is also potential in connecting the dataset with existing digital resources on the Assyrian Empire. The authors invite all interested researchers to enrich and further improve the dataset.

The dataset has already been used for teaching purposes. It is well suited for introducing history students to SNA, but it can also be used to teach more advanced methods of network analysis and visualization. Finally, the dataset can function as source material for MA theses and PhD dissertations in Assyriology and Ancient Near Eastern studies.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

Heidi Jauhiainen: conceptualization, data curation, investigation, methodology, software, validation, writing – original draft

Tero Alstola: conceptualization, data curation, investigation, validation, writing – original draft

AUTHOR AFFILIATIONS

Heidi Jauhiainen (iD) *orcid.org/0000-0002-8227-5627*
Department of Digital Humanities, University of Helsinki, Helsinki, Finland
Tero Alstola (iD) *orcid.org/0000-0003-4382-1733*
Department of Digital Humanities, University of Helsinki, Helsinki, Finland

REFERENCES

Alstola, T., Zaia, S., Sahala, A., Jauhiainen, H., Svärd, S., & **Lindén, K.** (2019). Aššur and His Friends: A Statistical Analysis of Neo-Assyrian Texts. *Journal of Cuneiform Studies*, *71*, 159–180. DOI: *https://doi.org/10.1086/703859*

Bastian, M., Heymann, S., & **Jacomy, M.** (2009, May). Gephi: An Open Source Software for Exploring and Manipulating Networks [Paper presentation]. In E. Adar et al. (Eds.), *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*. ICWSM-09, Menlo Park, CA (pp. 361–362). AAAI. *https://ojs.aaai.org/index.php/ICWSM/article/view/13937*

Földi, Z. J. (2019). Bullussa-rabi, Author of the Gula Hymn. In E. Jiménez (Ed.), From the Electronic Babylonian Literature Lab 1–7. *KASKAL*, *16*, 81–83. DOI: *https://doi.org/10.1400/278197*

Gardin, J.-C., & **Garelli, P.** (1961). Étude des établissements assyriens en Cappadoce par ordinateur. *Annales, 16*, 837–876. DOI: *https://doi.org/10.3406/ahess.1961.420758*

---

2   *https://www.en.ag.geschichte.uni-muenchen.de/research/mocci/index.html.*

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*, 1360–1380. DOI: *https://doi.org/10.1086/225469*

Granovetter, M. S. (1983). The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory, 1,* 201–233. DOI: *https://doi.org/10.2307/202051*

Jones, C. W. (2021). *Power and Elite Competition in the Neo-Assyrian Empire, 745–612 BC* [Unpublished doctoral dissertation]. Columbia University.

Plutniak, S. (2021). Assyrian Merchants Meet Nuclear Physicists: History of the Early Contributions from Social Sciences to Computer Science: The Case of Automatic Pattern Detection in Graphs (1950s–1970s). *Interdisciplinary Science Reviews, 46*, 547–568. DOI: *https://doi.org/10.1080/03080188.2021.1877502*

Radner, K., & **Baker, H. D.** (Eds.). (1998–2011). *The Prosopography of the Neo-Assyrian Empire*. Helsinki: The Neo-Assyrian Text Corpus Project.

Still, B. (2019). *The Social World of the Babylonian Priest*. Leiden: Brill. DOI: *https://doi.org/10.1163/9789004399969*

Waerzeggers, C. (2014). Social Network Analysis of Cuneiform Archives – a New Approach. In H. D. Baker & M. Jursa (Eds.), *Documentary Sources in Ancient Near Eastern and Greco-Roman Economic History: Methodology and Practice* (pp. 207–233). Oxford: Oxbow. DOI: *https://doi.org/10.2307/j.ctvh1dn9m.14*