

Nico Dogaer*, Mark Depauw

Mapping the Demotic Epistolary Framework through Network Visualisation

<https://doi.org/10.1515/zaes-2017-0011>

Summary: In this article, network visualisation is presented as a new way to explore the formulaic framework of the Demotic papyrus letters. It also serves as an introduction of formal network analysis to the field of Demotic studies. This approach presents the evidence in a new light, focussing on the combination of formulae rather than on the phrases themselves, resulting in fresh observations on epistolary practices.

Keywords: Demotic studies – Epistolography – Network Analysis

Demotic letters possess a clear structure signalling their epistolary nature. They almost always include an opening formula (also called interior address) and a closing formula. Probably in the great majority of papyrus letters there also was an exterior address legible on the outside of the folded package, but this is now often lost. These set phrases served to establish contact between the two parties, while ensuring that the letters arrived at the right destination. Because of their formulaic nature, the framework they form is particularly suited to quantitative analysis. In Depauw's study of Demotic letters¹, this approach is limited to the frequency of individual formulae. This article goes a step further and presents a new way to conceptualise the Demotic epistolary framework: through network visualisation, focusing on co-occurrence.

In recent years, network analysis has frequently been applied in historical disciplines². In the field of Egyptology, however, this technique hardly seems to have been explored yet. Networks have been used as a concept, in e. g. road 'networks' or trade 'networks', but the application of formal network analysis, involving software and statistics,

¹ Cf. Depauw 2006.

² Cf. www.historicalnetworkresearch.org/resources/bibliography for a comprehensive and up-to-date bibliography on network analysis in history and archaeology.

* Corresponding author: Nico Dogaer,

E-Mail: nico.dogaer@kuleuven.be

Mark Depauw, E-Mail: mark.depauw@kuleuven.be

has remained limited to prevalently Greek datasets of the Graeco-Roman and the Byzantine period³. For the earlier periods, only the Amarna letters, largely written in cuneiform script, have been subjected to network analysis⁴. This paper thus also aims to introduce the technique to the ancient Egyptian textual material, in particular in the domain of Demotic studies.

First, the formulae constituting the Demotic epistolary framework will be discussed, and the structure of the database will be outlined. In the next section, formal network analysis is introduced and applied to the source material. Attention will also be devoted to various network measures facilitating the detection of patterns in our data. Finally, issues of continuity and change in the selection and combination of epistolary formulae will be explored from a network perspective.

The Demotic epistolary framework⁵

The basic structure of Demotic letters consists of an exterior address, an interior address, the body of the letter and a closing formula. Some elements can be omitted, but the majority of preserved letters includes the last three⁶. 333 texts with at least one of the phrases constituting the Demotic epistolary framework are incorporated in Depauw's database of Demotic letters⁷. Each record in this database, which forms the basis for the figures and the observations

³ Ruffini (2008) introduced social network analysis to the field of papyrology. Other studies include e. g. Broux (2016) on Egyptian place names or Dogaer 2015 on Egyptian hybrid names.

⁴ Cf. Cline and Cline 2015.

⁵ The formulae are discussed in greater detail in Depauw 2006, 113–175. The present study only deals with standard formulae, while supplements, e. g. *b3k* (servant) or *hry* (master), are not taken into account.

⁶ The inclusion or omission of an exterior address is often difficult to verify because it is often badly damaged due to various reasons, e. g. the vulnerability of the outside of the papyrus, the reuse of the verso of old papyri and the way papyri are preserved in the present age. Therefore no figures are provided for the lack of an exterior address in the remainder of the article, as opposed to the omission of other elements.

⁷ This Filemaker database, now fully integrated in the Trismegistos (www.trismegistos.org) conglomerate and soon accessible online,

below, contains information about sender, addressee, formal aspects of the document, find spot, date and the formulae used in the letter. As for the abbreviations used, 'A' corresponds to the sender, 'B' to the addressee, 'G' to a god and 'P', which is only used for closing formulae, to the scribe, often identical to the sender.

The exterior address is written on the outside of the folded papyrus, and its primary function was to ensure that the letter arrived at its destination. For this reason, the formula almost always identifies the addressee (B). As the text of ostraca remained visible during transport, letters written on this material mostly did not include an exterior address⁸. Three formulae make up the majority of attested exterior addresses: *ỉ.ỉr-ỉr B* (41), *r dy.t s n B* (22) and *ỉrw A m-b3h B* (17). Other phrases appear only a handful of times. Table 1 provides all known exterior addresses, along with a translation and the number of attestations.

Tab. 1: Exterior addresses attested in Demotic letters

Formula	Translation	Attestations
<i>ỉ.ỉr-ỉr B</i>	To B	41
<i>r dy.t s n B</i>	To give to B	22
<i>ỉrw A m-b3h B</i>	Voice of A before B	17
<i>(n) B</i>	To B	9
<i>B n-dr.t A</i>	B from A	2
<i>A sm3' r B</i>	A greets B	3
<i>sm3' r B</i>	Greetings to B	1
<i>sm3' r B n-dr.t A</i>	Greetings to B from A	1
<i>n-dr.t A</i>	From A	1
<i>r dy A n B</i>	What A has given to B	1

The interior address served as the actual opening of the letter. In most languages, letters do not immediately start with the message they aim to convey, but first establish contact between the sender and the addressee. In Demotic, some seven opening formulae make up most of the attestations. These usually identify the sender (A) or the addressee (B) by means of names or titles. Some formulae communicate only this information: *ỉ.ỉr-ỉr B* (65), *ỉrw A m-b3h B* (38), *A p3 nty dd n B* (27), *A p3 nty dd* (22) and *A or B* (15). Others have an additional (slightly) philopronetic connotation, e. g. *A sm3' r B* (75) and its more obviously religious counterpart *A sm3' r B m-b3h G* (29). The remain-

ing eight formulae only occur once or twice. A comprehensive overview of interior addresses is provided in table 2. Eight fully preserved letters do not include an opening formula. These are very informal communications or drafts⁹.

Tab. 2: Interior addresses attested in Demotic letters

Formula	Translation	Attestations
<i>ỉrw A m-b3h B</i>	Voice of A before B	38
<i>A sm3' r B m-b3h G</i>	A blesses B before G	29
<i>A sm3' r B</i>	A blesses B	75
<i>ỉ.ỉr-ỉr B</i>	To B	65
<i>A p3 nty dd n B</i>	A is the one who says to B	27
<i>A p3 nty dd</i>	A is the one who says	22
<i>A or B</i>	(From) A or (To) B	15
<i>A n B</i>	A to B	2
<i>A dd n B n t3y ỉty A</i>	A says to B at this moment: A is the one who says	2
<i>p3 nty dd</i>		
<i>A p3 nty ỉr wh3 n</i>	A is the one who writes a letter	1
<i>smw3 r n3 'wy.w n</i>	of blessing to the houses of	
<i>wđy mtw B n-ỉm=w</i>	health B is in	
<i>A p3 nty ỉb r n3</i>	A is the one who writes to the	2
<i>'wy.w n wđy mtw B</i>	houses of health B is in	
<i>n-ỉm=w</i>		
<i>ỉ.ỉr=y ỉb s ỉ.ỉr-ỉr B</i>	It is to B that I am writing	1
<i>A nty š n r p3 gy n</i>	It is A who enquires after the	2
<i>wđy n B</i>	state of health of B	
<i>n-dr.t A</i>	From A	1
<i>p3y=k* b3k A m-b3h B</i>	Your servant A before B	2
<i>∅</i>	<i>∅</i>	8

Following the body of the document, a closing formula is added to mark the end of the letter. Almost all of these include the element '*sh*', meaning 'written'. This could be complemented with the date (103), with the name of the person who wrote the letter (13) or with both (90). Other closing formulae repeat the word *sh*, in combination with one or more of the optional elements. The date is usually provided in the form of the regnal year of the ruling pharaoh, the month, the season and the day¹⁰. Table 3 lists the different combinations and their frequency. Closing formulae are omitted much more frequently than interior addresses, especially on ostraca¹¹.

was also used for Depauw 2006, where further information about the corpus can be found.

⁸ Cf. Depauw 2006, 120. He considers TM 49380 (O. Leiden 354) the only exception. This letter mentions the name of the addressee on an otherwise blank line preceding a conventional interior address.

⁹ Cf. Depauw 2006, 155.

¹⁰ Cf. Depauw 2006, 159.

¹¹ Cf. Depauw 2006, 169.

Tab. 3: Closing formulae attested in Demotic letters

Formula	Translation	Attestations
$s\dot{h} P n DATE$	Written by P in DATE	90
$s\dot{h} P$	Written by P	13
$s\dot{h} n DATE$	Written in DATE	103
$s\dot{h}$	Written	16
$s\dot{h} s\dot{h} P n DATE$	Written. Written by P in DATE	1
$s\dot{h} s\dot{h} DATE$	Written. Written in DATE	2
$s\dot{h} P s\dot{h} n DATE$	Written by P. Written in DATE	1
$s\dot{h} n DATE s\dot{h} P$	Written in DATE. Written by P	1
DATE	DATE	1
∅	∅	43

Network visualisation

The above tables are not too divergent from the summaries for each type of the elements of the epistolary framework in Depauw's book, except that these include a chronological component. In the relevant sections, he also looked into combinations of the three elements of the epistolary framework. The interior address is clearly treated as the most central element. This is illustrated by the eight tables on pp. 173–174, one for each of the most common types of interior address, with the combined closing formulae in columns and the exterior address in rows (e.g. table 4 below). Although they contain all essential information, these tables do not provide a clear overview of the combinations, not even for the interior address but certainly not for the exterior addresses and the closing formulae.

This is where network analysis comes in. This technique, borrowed from mathematics, studies relations between entities¹². It can be used to analyse virtually any type of relation, between any kind of actor. In historical disciplines, it is most often applied to relationships between people (so-called 'social network analysis')¹³. Co-occurrence can also be considered a type of relation, however, and network analysis can be used to study combinations of linguistic features in a textual corpus, such as the Demotic letters¹⁴. Whatever the nature of the relation or the entities under consideration, network analysis employs common concepts and terminology. The entities are represented by 'nodes', and these constitute the core

elements of the network. In a visualisation, they are usually rendered as dots. If the relation under scrutiny exists between a pair of nodes, these are connected through an 'edge', visualised as a line running from one node to the other. In this case, the nodes represent formulae, which are connected through an edge if they occur together in the same letter (cf. figure 1¹⁵).

The thickness of the edges is in this figure determined by the 'edge weight', which is a measure for the frequency of the relation. The line connecting two formulae is in other words thicker when formulae are attested together in many letters. The size of the nodes reflects the frequency of the formulae themselves: the more times a formula is attested, the larger the corresponding node is¹⁶. The nodes are also coloured according to the type of formula. Exterior addresses are dark grey, interior addresses grey, and closing formulae light grey¹⁷. In network terms, these node characteristics are called 'attributes', introducing non-network properties in the graph.

The position of the nodes and edges is not arbitrary, but it is flexible, and various layouts can be chosen to graphically represent the network¹⁸. There is no inherent link between the properties of a node and its position in the chart; only its placement in relation to other nodes is significant. This might seem counterintuitive at first, but it is important to keep in mind when comparing different networks provided in this article.

One of the major advantages of this type of visualisation is that all formulae attested in the entire epistolary corpus are represented in a single image, with basic information about their nature and frequency (cf. figure 1). Such visualisations are particularly helpful when studying the co-occurrence of formulae. Common combinations jump out thanks to the thick lines connecting the formulae and their proximity (although the latter is not

¹⁵ This and all other networks in this article were rendered using the open source visualisation software Gephi (www.gephi.org). If you should like to indulge in network analysis yourself, we recommend the blog 'Six Degrees of Spaghetti Monsters' (spaghetti-os.blogspot.be) to get started.

¹⁶ For reasons of intelligibility, the nodes are scaled down. If they would be sized on a scale from 1 to 103, some nodes would dominate the image to such a degree that smaller nodes become virtually invisible. In particular their colour, and consequently their type, would no longer be distinguishable in the printed image.

¹⁷ In the full-colour versions of the networks available online (cf. infra), exterior addresses are represented by blue nodes, interior addresses by red nodes and closing formulae by yellow nodes.

¹⁸ We used the Gephi layout algorithm 'ForceAtlas 2', as it clusters strongly connected nodes together and pushes outliers to the periphery of the network. This way, the topology of the network aids in its interpretation. We also ran 'Noverlap' to prevent node overlap.

¹² Cf. Barabási 2002 for an accessible introduction to network theory.

¹³ Cf. Düring et.al. 2016 for an overview of historical network analysis.

¹⁴ Cf. Mehler et. al. 2013.

Tab. 4: Reproduction of one of Depauw's tables of combinations¹⁹

	<i>A p̄ n̄ ty d̄ d̄ n̄ B</i>								?
	<i>s̄h̄ P n̄</i> DATE	<i>s̄h̄ P</i>	<i>s̄h̄ n̄</i> DATE	<i>s̄h̄</i>	<i>s̄h̄ s̄h̄ P</i> n̄ DATE	<i>s̄h̄ s̄h̄ n̄</i> DATE	<i>s̄h̄ P s̄h̄</i> n̄ DATE	∅	
<i>ī.īr-h̄r B</i>	–	–	2	–	–	–	1	–	–
<i>r dy s n B</i>	–	–	1	–	–	2	–	–	–
none or ?	1	2	5	2	1	–	–	4	6

strictly necessary and depends on the layout opted for). Also, the network graph allows to switch focus easily from one particular formula and its combinations to the next. Questions such as ‘How common is the interior address attested in this combination found in conjunction with other closing formulae?’ now not necessarily lead to a splitting headache. Thus in figure 1 attention is drawn to the triangle formed by the closing formula *s̄h̄ P n̄ DATE* and the interior and exterior address *ī.īr-h̄r B*. It also shows, however, that the exterior address *ī.īr-h̄r B* is almost never combined with the interior addresses *A sm̄³ r B* or *A sm̄³ r B m-b̄h̄ G*. The former of these normally does not have an exterior address, and for the latter *r dy.t s n B* is the preferred combination. This kind of observations would be very difficult in a traditional visualisation.

Network measures²⁰

In addition to its visualising properties, network analysis offers statistical tools to facilitate the study of large corpora, which can be applied to the Demotic papyri as well. Some of these measures inform us about the position of a node in the network structure; others allow us to characterise the entire network; finally, algorithms can be used to identify subgroups within the general network. Of each category, one or two examples relevant to the study of letters will be discussed.

Measuring centrality: degree

Centrality measures inform us about the position of a node in the network structure. Different concepts of centrality have been developed, focussing on different ways in which nodes contribute to the general structure of the

network. For example, a node might be central because it has the most connections to other nodes. Nodes with only a few connections can be central in other ways as well, e. g. because of their position at an intersection between groups of nodes which would otherwise have been unconnected. What centrality actually means in reality depends on the nature of the network: a central position in a friendship network would be a good thing, but the same cannot be said for a network mapping conflicts²¹.

The most basic centrality measure is called ‘degree’. The degree of a node corresponds to the number of edges the node has. In our case, the degree of a node representing a formula tells us how many other formulae it is combined with. Degree can also be weighted, taking into account the weight of the edges. As our purpose here is to study the amount of combinations, regardless of their frequency, we will be using regular degree. After calculating the degree of each node, these figures can then be compared to each other and to the frequency of the formulae in general.

Figure 2 shows the same network as figure 1, but this time the nodes are sized according to degree. Figure 1 displays which formulae are attested most; figure 2 reveals which formulae were combined with most other formulae. If we compare the two figures, some key differences are highlighted. In both graphs, closing formulae (light grey nodes) stand out. This is a logical consequence of their limited number: as there were less formulae to choose from, the existing ones were both used more frequently and in combination with more other formulae. There are some significant differences, however, concerning degree and frequency within this type. The formula attested most, *s̄h̄ n̄ DATE*, is relatively less important in figure 2, and it is combined with significantly less formulae than *s̄h̄ P n̄ DATE*, the second most frequent phrase. The lack of a closing formula (∅) was not particularly frequent, but according to this graph, it occurred in combination with

¹⁹ Cf. Depauw 2006, 174.

²⁰ Cf. Wasserman and Faust 1994 for the mathematics behind these and other measures.

²¹ Cf. Borgatti, Everett and Johnson 2013, 163–180 for a detailed explanation about centrality concepts and their applications.

all kinds of other formulae. The same is true for *sh*, the degree of which is much more significant than its frequency.

Exterior addresses (dark grey nodes) feature more prominently in figure 2 compared to figure 1. They are attested (or preserved) less often than the other categories, but they are combined with a good deal of other formulae. In particular (*n*) *B* and *ỉ.ỉr-ħr B* stand out. (*n*) *B* is only attested nine times, but it occurs together with nine different other formulae, hence its size in figure 2. *ỉ.ỉr-ħr B* is an equally revealing case: it is most often combined with the interior address *ỉ.ỉr-ħr B*, which is attested much more frequently than its exterior counterpart. As it turns out, however, the exterior variant is combined with double the amount of other formulae compared to interior *ỉ.ỉr-ħr B*. Interestingly, the same does not apply to the other duo of the same interior and exterior address, *ħrw A m-bʒħ B*: interior *ħrw A m-bʒħ B* is both attested most frequently and in combination with most other formulae. Both of these patterns would have been far less obvious without this network visualisation involving degree.

As for the interior addresses (grey nodes), some nodes are somewhat larger in figure 2, mainly *A pʒ nty đđ n B*, *A smʒʳ r B m-bʒħ G* and the lack of an opening formula (\emptyset). These addresses are combined with several other formulae, regardless of their frequency. On the other hand, *ỉ.ỉr-ħr B* (cf. supra) and to a lesser extent *A smʒʳ r B* and *ħrw A m-bʒħ B* feature less prominently when the nodes are sized according to degree. *ỉ.ỉr-ħr B* was only ever combined with its exterior equivalent and the two most frequent closing formulae. Although *A smʒʳ r B* is attested quite frequently, it is the standard opening formula of ostraca, which display less variety overall as to the epistolary framework (cf. infra). The degree of *ħrw A m-bʒħ B* is also rather unimpressive relative to its frequency. The opening formulae attested most thus remained limited to relatively narrow frameworks, whereas combinations involving less frequent formulae were more varied.

The overview presented in this section is hard to attain in a traditional way. Although the number of combinations each formula is involved in can be outlined in a table, as well as the specific phrases a particular formula is combined with, it is impossible to capture both aspects in one table. Depauw (2006) provides a number of tables which attempt to do this, but only the major formulae are taken into account there, and each table is drawn up from the perspective of one single formula (cf. supra). For a comprehensive overview, 36 of such tables should be composed, and subsequently compared. Network analysis presents both the amount of combinations and the combinations themselves in one elegant graph. Furthermore, this can be achieved at the push of a button, as the

software calculates degree based on the information we already provided to generate the network presented in figure 1²².

Macro-analysis: calculating density and average degree

Some network measures are used to characterise whole networks, as opposed to single nodes²³. Many of those inform us about the cohesion or connectedness of the network. To calculate this aspect, ‘density’ is often used. Density is defined as the number of existing edges expressed as a proportion of the total possible edges. This computation yields a number between 0 and 1, whereby a density of 0 corresponds to a network without edges and 1 to a network in which every node is connected to every other node. In itself, this number is not particularly revealing, and density is best used in a comparative way. The advantage of using this measure lies in the quantification of the difference between networks, which can then be compared more objectively.

In our case, density indicates the existing number of combinations of epistolary formulae as a proportion of the number of potential combinations. A density of 0 would correspond to a collection of letters in which every letter would have a single, different formula, whereas a density of 1 would mean that all formulae were combined, either in one or in many letters. As in this case the three distinct categories of nodes (exterior addresses, interior addresses and closing formulae) are only connected across categories, combinations within a single category being virtually non-existent²⁴, the figures cannot be distorted by a single letter containing all formulae. This particular network structure is also expected to yield a low density.

As an illustration, we compare the density of two subsets: letters written on papyrus (figure 3) versus letters

²² Other centrality measures include ‘betweenness’ (the more often a node lies along the shortest path between two other nodes, the higher its betweenness centrality; such nodes often function as bridges connecting otherwise disconnected parts of the network), ‘closeness’ (this is simply the sum of all the shortest paths from one node to all others) and ‘eigenvector’ (this measure takes the centrality of a node’s connections into account) centrality. These measures are, however, not applicable to the epistolary corpus. They are only included here to introduce them to scholars who might find them useful for their own purposes. These readers are advised to turn to the works on network analysis in general cited in this article.

²³ Cf. Borgatti, Everett and Johnson 2013, 149–162 for a detailed explanation of whole-network measures.

²⁴ The only exception is TM 43380 (P. Loeb 4), which contains a postscript followed by a second closing formula.

written on ostraca (figure 4). The epistolary framework differs considerably depending upon the material used. As no less than 33 out of the total of 36 formulae are used for papyrus letters and these constitute a clear majority, figure 3 resembles the general overview presented in figure 1. The nodes corresponding to the closing formulae *sh n DATE* and *sh P n DATE* are again the most prominent. These are frequently connected to combinations of interior and exterior addresses such as *i.ir-hr B - i.ir-hr B*, *hrw A m-b3h B - hrw A m-b3h B* or *A sm3' r B m-b3h G - r dy.t s n B*.

In letters written on ostraca, only 17 different formulae are used. The most obvious feature of figure 4 is the absence of dark grey nodes in the network. As letters on ostraca were not folded and sealed, no exterior address was required. The most prominent node corresponds to the interior address *A sm3' r B*, which was most often combined with the closing formula *sh n DATE* or with no closing formula at all (\emptyset). The combination of *i.ir-hr B* and *sh P n DATE* is attested quite frequently as well. As a result of the limited writing surface, no space is wasted on double identification of the parties: if the name of the sender is specified in the opening formula, this is rarely repeated in the closing formula.

Clearly, more formulae were used in papyri, but the degree of 'combinedness' of these formulae relative to each other is hard to describe. This is where density comes in, here indicating the proportion of the existing number of combinations of epistolary formulae to the number of potential combinations. The network with the highest density represents the epistolary framework for which most possible combinations of formulae are actually attested. This turns out to be the network of formulae attested in ostraca (0.191 versus 0.148 in the network of formulae attested in papyri).

One could argue that this comparison is unfair, however, as the size of the respective frameworks is not the same. Although density adjusts for the number of nodes in the network, in most cases it is true that the smaller the network, the more likely that all nodes are connected. For this reason, some researchers prefer another measure, called 'average degree'. Average degree is calculated by averaging the degree (number of edges, cf. supra) of every node in the network. For papyri, the average degree is 4.727, meaning that every formula is on average combined with 4.727 other formulae. In the network of ostraca, this figure is 3.059, and every formula is on average attested in combination with 3.059 other formulae.

The average degree of both networks seems to contradict the results of the density analysis. The explanation for this lies in the structure of the networks, and consequent-

ly the nature of the respective epistolary frameworks. Ostraca usually include two formulae, whereas the standard papyrus letter contains three. As connections are nearly only attested between categories and not within types of formula, it is not simply a matter of network size. The addition of an extra category to one of the networks (exterior addresses in the case of papyri) greatly expands the number of possible connections; hence the higher average degree of the network of formulae in papyri. On density, on the other hand, this addition has the opposite effect, for the same structural reason. This is why formulae are on average combined with more different formulae in papyri, even though the actually attested proportion of possible combinations is larger in the network of ostraca.

If we eliminate this structural difference by only taking interior addresses and closing formulae into consideration, the figures are closer to each other. This is particularly true for the average degree, which is most sensitive to network size. For papyri, the average degree decreases to only 3.13, versus 3.0 for the network of letters written on ostraca. On average, formulae are thus combined with the same amount of other formulae in both datasets. As for density, on the other hand, the disparity increases: the figure for the ostraca network amounts to 0.2 versus 0.142 for the network of papyrus letters. This points to the relative importance of the combination of formulae with exterior addresses in the papyrus network for the diversity of the epistolary framework, as we would normally expect density to increase as the size of the network is reduced.

In sum, different measures calculate different things and yield different results. Not every statistical tool is equally well suited to every dataset, and scholars need to decide which ones suit their data best. Different analyses can also be compared, as we did. In this case, the combination of statistics led to counterintuitive results, due to structural differences between the epistolary frameworks. This does not mean, however, that the results should be discarded. On the contrary, it demonstrates the advantage of network measures: quantification allows for comparison of different datasets in a uniform way²⁵.

²⁵ Other measures to characterise the whole network exist, which might be relevant to other egyptological questions. 'Transitivity', for example, indicates the degree to which a node A is connected to a node C if node A and C are both connected to the same node B. In a friendship network, this translates to friends of friends: if the person represented by node B is friends with both A and C, are A and C friends as well? The more of these triads exist, the higher the transitivity of the network. Another form of macro-analysis is offered by 'network diameter', which is the shortest distance between the two most distant nodes. This can be compared for different networks, e.g. to study the spread of information in a social network. These

Identifying subgroups: community detection

Another useful tool provided by network analysis software is the automatic detection of so-called ‘communities’, i.e. subgroups within the network. In a social network, for example, such strongly connected subgroups might consist of friends or co-workers, who interact more frequently among each other than with other actors in the network. Computer algorithms calculate these communities by comparing the number of edges within groups of nodes to the number of edges between different groups²⁶. Once subgroups have been detected, underlying factors causing the network structure can be examined, e. g. in the form of shared attributes²⁷.

In our network of epistolary formulae, these communities consist of groups of phrases used more frequently together than with others. They constitute distinct sets of combined formulae, which may have been used on certain occasions, in certain periods or to write certain kinds of letters. The analysis thus allows us to identify standard frameworks. It is also useful to compare the results of this quantitative analysis to our prior observations regarding the network structure, as the algorithms often detect patterns not immediately obvious to the human eye. The resolution of the analysis can be adapted as well, resulting in a greater or smaller number of communities. This way, different configurations can be compared. The nodes are coloured according to the communities the algorithm has identified²⁸.

Figure 6 shows the result of the community detection with the basic resolution 1.0 (5 communities). The nodes are coloured according to the subgroup the algorithm has assigned them to. Two of the communities consist of only one node, as these nodes are isolates. They are not connected to any other formula, and as such they are irrelevant to this discussion. This leaves us with three major subgroups. The first community is dominated by the exterior address $\dot{\iota}\dot{\iota}r\text{-}hr B$, the interior address $\dot{\iota}\dot{\iota}r\text{-}hr B$ and the closing formula $sh P n DATE$. The second consists of the

exterior address $hrw A m\text{-}b\dot{\iota}h B$ and the interior address $hrw A m\text{-}b\dot{\iota}h B$. The final subgroup comprises mainly the interior addresses $A sm^{\dot{\iota}c} r B$, $A p^{\dot{\iota}} nty \dot{d}d n B$, $A p^{\dot{\iota}} nty \dot{d}d$, $A sm^{\dot{\iota}c} r B m\text{-}b\dot{\iota}h G$; the closing formula $sh n DATE$ and the lack of closing formula; and the exterior address $r dy.t s n B$, along with numerous minor formulae.

These three communities represent different epistolary frameworks. The next step is to determine which set of formulae was used on what occasion. As it turns out, the division between the communities is primarily a chronological one (cf. infra, the section on time & space). The combination of the addresses $hrw A m\text{-}b\dot{\iota}h B$ is characteristic for the earliest period, whereas $sh P n DATE$ and $\dot{\iota}\dot{\iota}r\text{-}hr B$ constituted the most popular framework in the fourth and third century BC. The last community consists of formulae popular in the second and first century BC, especially the combination of $A sm^{\dot{\iota}c} r B$ and $sh n DATE$ (26 out of 28 letters) and attestations of $A sm^{\dot{\iota}c} r B$ without a closing formula (16 out of 18).

A smaller resolution (0.5) yields 8 communities (figure 5). Again, two of these consist of one isolate each and should be disregarded. The major difference compared to the previous situation is the dissolution of the $A sm^{\dot{\iota}c} r B - sh n DATE$ -axis. This is due to the frequent combination of this interior address with the lack of a closing formula, whereas the closing formula $sh n DATE$ has strong connections to multiple other interior and exterior addresses. This is linked to the prevalence of $A sm^{\dot{\iota}c} r B$ in ostraca, which often lack a closing formula (cf. supra).

Another new community consists mostly of the interior address $A p^{\dot{\iota}} nty \dot{d}d n B$ and some rare closing formulae, along with other infrequently attested interior addresses occurring together with these closing formulae. This community thus assembles deviations from the standard frameworks revolving around $A p^{\dot{\iota}} nty \dot{d}d n B$. The peculiar set of combinations for this formula suggests that Depauw (2006) is right in stating that it probably constituted an infiltration into epistolography from another genre, i.e. contracts²⁹. The structural position of this node, and consequently the status of the formula it represents, would have been rather less obvious without the calculations by the community detection algorithm. The major advantage of network analysis, i.e. that it also takes account of the other connections of connected nodes, ‘friends of friends’, is clearly demonstrated here.

Increasing the resolution to 2.0 results in the division of the network in two parts (figure 7). Basically, the community consisting of the exterior address $hrw A m\text{-}b\dot{\iota}h B$ and the interior address $hrw A m\text{-}b\dot{\iota}h B$ from figure 6 is

measures are not relevant to the study of documentary letters and they will not be further elaborated upon here. For an in-depth explanation of these and other measures, the reader should again consult the specialised literature cited in the footnotes and the bibliography. ²⁶ For this study, the ‘Louvain’ algorithm was used. Cf. Blondel e.a. 2008.

²⁷ Cf. Borgatti, Everett and Johnson 2013, 181–205 for more information on identifying subgroups.

²⁸ Especially in this section, the reader is advised to consult the full-colour versions of the networks online, as these provide a much clearer picture, due to the multitude of communities and subsequently of different colours.

²⁹ Cf. Depauw 2006, 147.

now combined with the $\dot{i}\dot{i}r\text{-}hr\ B - sh\ P\ n\ DATE$ -cluster. As stated, the difference between the $hrw\ A\ m\text{-}b\dot{s}h\ B$ - and the $\dot{i}\dot{i}r\text{-}hr\ B$ -frameworks is chronological. The reason for their unification in this image lies in the lasting popularity of $sh\ P\ n\ DATE$, a formula frequently used in combination with each of the two other formulae. Despite the chronological gap, both frameworks were preferably used in combination with this closing formula, rather than with other closing phrases, although combinations with $sh\ n\ DATE$ occur as well.

The frameworks identified using community detection are not absolute. As the last example shows, there was also considerable freedom for scribes in their selection of formulae. Network analysis illustrates this nicely, as connections with formulae not belonging to the major structures are *visible* at all times. This does not mean, however, that it is impossible to detect clear patterns. The primary focus of this section has been to identify groups of formulae, but community detection can also be used the other way around: to determine the primary affiliation of less frequently attested formulae.

Time & space

In the final section, two important variables are examined in more detail: time and space. Chronological and geographical differences are explored using network analysis to determine whether the time or place of writing influenced the selection and combination of epistolary formulae.

Time

Although the networks presented so far have combined the situation in different time periods into a single graph, network analysis does not necessarily provide a static picture. To demonstrate this, a breakdown of the general network into three-time periods is provided here: the 6th–5th century BC, the 4th–3rd century BC and the 2nd–1st century BC³⁰. This allows us to examine the chronological evolution of the Demotic epistolary framework. The position of the nodes is based on the layout of the network of Demotic epistolary formulae in general (figure 1). The different figures can thus be compared to figure 1 and to each other,

³⁰ The same chronological subdivisions are used in the tables provided by Depauw 2006. These only present an overview per type, not of all formulae and combinations dating to the different time periods.

whereby recurring formulae and combinations can serve as visual reference points across figures. The first three centuries AD are not considered here, as only 14 letters dating to this period are preserved.

In 6th–5th century letters (figure 8), the most common framework consists of exterior and interior $hrw\ A\ m\text{-}b\dot{s}h\ B$ and the closing formula $sh\ P\ n\ DATE$. The standard addresses $hrw\ A\ m\text{-}b\dot{s}h\ B$ are also combined with $sh\ n\ DATE$. Another striking feature of this period are interior and exterior addresses containing the verb $sm\dot{s}'$. As it turns out, these forms were mainly combined with each other. Although the network in figure 8 contains the least amount of nodes, we should keep in mind that it is based on 25 letters only. In particular the variety of exterior addresses was greatest in this period.

The most important evolution in the 4th–3rd century BC (figure 9) is the gradual replacement of $hrw\ A\ m\text{-}b\dot{s}h\ B$ with the new interior and exterior address $\dot{i}\dot{i}r\text{-}hr\ B$. According to Depauw (2006), this shift is linked to the change in form from the broad to the high format³¹. The evolution is most marked for the exterior address: sometimes old interior $hrw\ A\ m\text{-}b\dot{s}h\ B$ is combined with new exterior $\dot{i}\dot{i}r\text{-}hr\ B$, whereas new interior $\dot{i}\dot{i}r\text{-}hr\ B$ is never combined with old exterior $hrw\ A\ m\text{-}b\dot{s}h\ B$. The $\dot{i}\dot{i}r\text{-}hr\ B$ addresses are most often combined with the closing formula $sh\ P\ n\ DATE$, which retains its earlier popularity. Its position is increasingly threatened, however, as $sh\ n\ DATE$ becomes more prominent. The variant with mention of the scribe still counts most attestations, but $sh\ n\ DATE$ is combined with more different formulae (degree of 11 versus 10).

Apart from two minor opening formulae, forms including the verb $sm\dot{s}'$ have virtually disappeared. Other forms emerged in this period: in addition to $\dot{i}\dot{i}r\text{-}hr\ B$, the interior addresses $A\ p\dot{s}'\ nty\ \dot{d}d$ and $A\ p\dot{s}'\ nty\ \dot{d}d\ n\ B$ were also an innovation. These forms were usually combined with the exterior address $\dot{i}\dot{i}r\text{-}hr\ B$ and the closing formula $sh\ n\ DATE$. Sometimes, the version without the name of the addressee is also combined with $sh\ P\ n\ DATE$. The deliberate lack of an opening formula is also a feature of the 4th–3rd century letters, which is only attested in this period. In two of the eight cases, the addressee is identified in the exterior address ($\dot{i}\dot{i}r\text{-}hr\ B$), and the sender is specified four times in the closing formula ($sh\ P\ n\ DATE$ or $sh\ P$).

In the 2nd–1st century BC (figure 10), this picture changes considerably, in part due to the increasing number of ostraca with their own conventions (cf. supra). In particular the prominent combination of the opening formula $A\ sm\dot{s}'\ r\ B$ and the lack of a closing formula (\emptyset) reflects this. The combination of $A\ sm\dot{s}'\ r\ B$ and $sh\ n\ DATE$ is charac-

³¹ Cf. Depauw 2006, 126.

teristic for ostraca as well, but it is also used in papyrus letters. In fact, $A sm^3' r B$ is the most frequently attested interior address in both ostraca and papyri. The same is true for the closing formula $sh n DATE$. The most popular combination in papyrus letters, however, remains $i. i' r- hr B - i. i' r- hr B - sh P n DATE$. Interior $i. i' r- hr B$ is occasionally also used in letters on ostraca. The interior and exterior addresses $hrw A m- b^3 h B$ have disappeared completely. $A p^3 nty dd$ and $A p^3 nty dd n B$, on the other hand, remain in use in this period.

Another feature of the late Ptolemaic epistolary framework is the revival of formulae containing sm^3' . In addition to the ubiquitous opening formula $A sm^3' r B$, its more obviously religious counterpart $A sm^3' r B m- b^3 h G$ rises to prominence as well, especially in papyri, where it is often combined with the new exterior address $r dy. t s n B$ and the closing formula $sh n DATE$ to form an alternative framework to $i. i' r- hr B - sh P n DATE$. The exterior address $A sm^3' r B$ makes a return as well.

The Demotic epistolary framework evolved sizeably over time. In the course of six centuries, some formulae disappeared, while others emerged to take their place. Network analysis provides an excellent way to explore this chronological development. Through successive visualisations in which the constellation of the nodes remains the same, the evolution in the combination of formulae can easily be traced. The possibility to filter out nodes corresponding to formulae not attested in a specific time frame facilitates the study of the formulae themselves as well. Once acquainted with network analysis, the resulting graphs are much more intuitive and easy to interpret than a series of tables containing the same information.

Space

Factors other than chronology determined the appearance of the Demotic epistolary framework as well. In this last section, geographical variation in the selection and combination of formulae is examined, in order to assess the influence of the place the letters were written or, perhaps better, the origin of the sender. Networks are generated on the basis of two datasets, which are geographically divergent but chronologically and typologically similar. The first one consists of 32 letters from the Loeb papyri, found in Tehna, Middle Egypt. The other is made up of letters written in Elephantine dating to the same period (currently 52 published ones). Both archives fall within the 4th–3rd century BC category (cf. supra). The Loeb let-

ters were mainly written at the end of the fourth century³², whereas the exact dating of the Elephantine letters is still problematic³³. Both groups of letters stem from a similar priestly milieu, and both archives consist entirely of papyrus letters.

Despite these similarities, some significant differences can be observed. In the Tehna corpus (figure 12), there is strikingly little variation in the selection of formulae. The network is dominated by the combination of interior and exterior $i. i' r- hr B$ and the closing formula $sh P n DATE$, which is very common in the 4th–3rd century BC (cf. supra). $i. i' r- hr B$ is the only exterior addresses attested in the archive, and attestations of formulae not belonging to the common framework are equally rare. This homogeneity is not due to the influence of one particular scribe, as the letters identify multiple senders and addressees.

The Elephantine letters (figure 11), on the other hand, show considerable variety. The older combination of interior and exterior $hrw A m- b^3 h B$, which had already disappeared in Tehna, was still relatively common in this region. In fact, it is attested more than the newer $i. i' r- hr B$ duo. This pattern suggests that scribes in Elephantine were either more conservative, or that the letters were written relatively early, in the first part or the middle of the fourth century BC³⁴. On the other hand, the closing formula without mention of the scribe, the growing popularity of which is also a feature typical of the 4th–3rd century BC, is attested more often in Elephantine than in Tehna. This may be related to the prominent use of $i. i' r- hr B$ in the latter, where the sender still needs to be identified.

Most striking is the frequent combination of interior $hrw A m- b^3 h B$ and exterior $i. i' r- hr B$, which is far less common in general than the use of identical interior and exterior addresses (cf. supra). This combination seems to have been a distinctly regional feature, as it occurs only twice elsewhere. Overall, however, geographical differences seem to have had less of an influence on the selection of formulae than chronological and typological factors. Other formulae emerging in this period, such as the interior addresses $A p^3 nty dd$, $A p^3 nty dd n B$ or the lack of an opening formula (\emptyset) are attested in both archives. The same is true for the interior address $A sm^3' r B m- b^3 h G$, which was somewhat in decline. Although the closing formula $sh n DATE$ is not attested as frequently in Tehna,

³² Cf. Spiegelberg 1931, XII–XIV.

³³ About half of the letters mention a regnal year, but no ruler is included in the date.

³⁴ Broux and Depauw (2015) also argue for a date around the middle of the fourth century for a considerable portion of the Elephantine letters, using another network approach.

it was not unheard of either. The most significant feature, the surprising popularity of a specific combination of formulae in Elephantine, is highlighted through the use of network visualisation. This again underlines the advantages of a network approach to the epistolary framework, as opposed to a focus on the distribution of formulae alone.

Closing remarks

Network analysis provides a new way to explore the formulaic framework of Demotic papyrus letters. The entire corpus can be presented in one single image, taking both the frequency and the combination of formulae into account. Especially for the study of the latter, network visualisations offer numerous advantages over conventional tables. This is also true for the study of particular subsets: combining the figures of this article provides a complete overview of the Demotic epistolary framework. The influence of chronology, typology and geography are presented in a uniform way, all ultimately derived from the general network of formulae. These visualisations enable the comparison of different variables, and this has allowed us to identify distinct frameworks. These can in turn be used to provide rough estimates of the date of certain letters, for example.

The use of visualisation software and statistical algorithms greatly facilitates this enquiry. Automatically calculated centrality measures point out the importance of individual formulae. Other measures, e. g. density or average degree, characterise whole networks. Through these, differences between frameworks can be quantified and compared objectively. Community detection algorithms assist in the identification of distinct groups of formulae, and they often lead to unexpected insights.

This article has focused on the basic epistolary framework, but other topics can be explored in much the same way. Elements from the body of the letter, such as greetings or courtesies, can be examined from this new perspective. The analysis of other formulaic documents through net-

work visualisation, e. g. petitions, might also lead to fresh perspectives. Finally, studies of letters written during other stages of the Egyptian language could benefit from this approach as well.

Bibliography

- Barabási, A.-L., 2002, *Linked: The New Science of Networks*, Cambridge (Massachusetts).
- Blondel, V. D. et. al., 2008, “Fast Unfolding of Communities in Large Networks”, *Journal of Statistical Mechanics: Theory and Experiment* 10 (<http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta>).
- Borgatti, S. P., Everett, M. G., Johnson, J. C., 2013, *Analyzing Social Networks*, London.
- Broux, Y., (2016), “Detecting Settlement Communities in Graeco-Roman Egypt”, *BASP* 53, 295–313.
- Broux, Y., and Depauw, M., 2015, “Developing Onomastic Gazetteers and Prosopographies for the Ancient World through Named Entity Recognition and Graph Visualization: Some Examples from Trismegistos People”, in: L. Aiello and D. McFarland (eds.), *SocInfo 2014 International Workshops*, Heidelberg, 304–313.
- Cline, D. H., and Cline, E. H., 2015, “Text Messages, Tablets, and Social Networks in the Late Bronze Age Eastern Mediterranean: The Small World of the Amarna Letters”, in: J. Mynářová, P. Onderka, P. Pavúk (eds.), *There and Back Again – the Crossroads II. Proceedings of an International Conference Held in Prague, September 15–18, 2014*, Prague, 17–44.
- Depauw, M., 2006, *The Demotic Letter: A Study of Epistolographic Scribal Traditions Against their Intra- and Intercultural Background*. *DemStud* 14, Sommerhausen.
- Dogaer, N., 2015, “Egyptian Names Derived from Foreign Elements: Innovation in Egyptian Onomastic Practice After the Roman Conquest”, *CdE* 90, 360–370.
- Düring, M. et. al., 2016, *Handbuch Historische Netzwerkforschung: Grundlagen und Anwendungen*. Schriften des Kulturwissenschaftlichen Instituts Essen (KWI) zur Methodenforschung 1, Berlin.
- Mehler, A. et. al., 2013, “Inducing Linguistic Networks from Historical Corpora: Towards a New Method in Historical Semantics”, in: P. Bennett et. al. (eds.), *New Methods in Historical Corpora*. *Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache* 3, Tübingen, 257–274.
- Ruffini, R. G., 2008, *Social Networks in Byzantine Egypt*, New York.
- Spiegelberg, W., 1931, *Die Demotischen Papyri* Loeb, Munich.
- Wasserman, S., and Faust, K., 1994, *Social Network Analysis: Methods and Applications*, Cambridge.

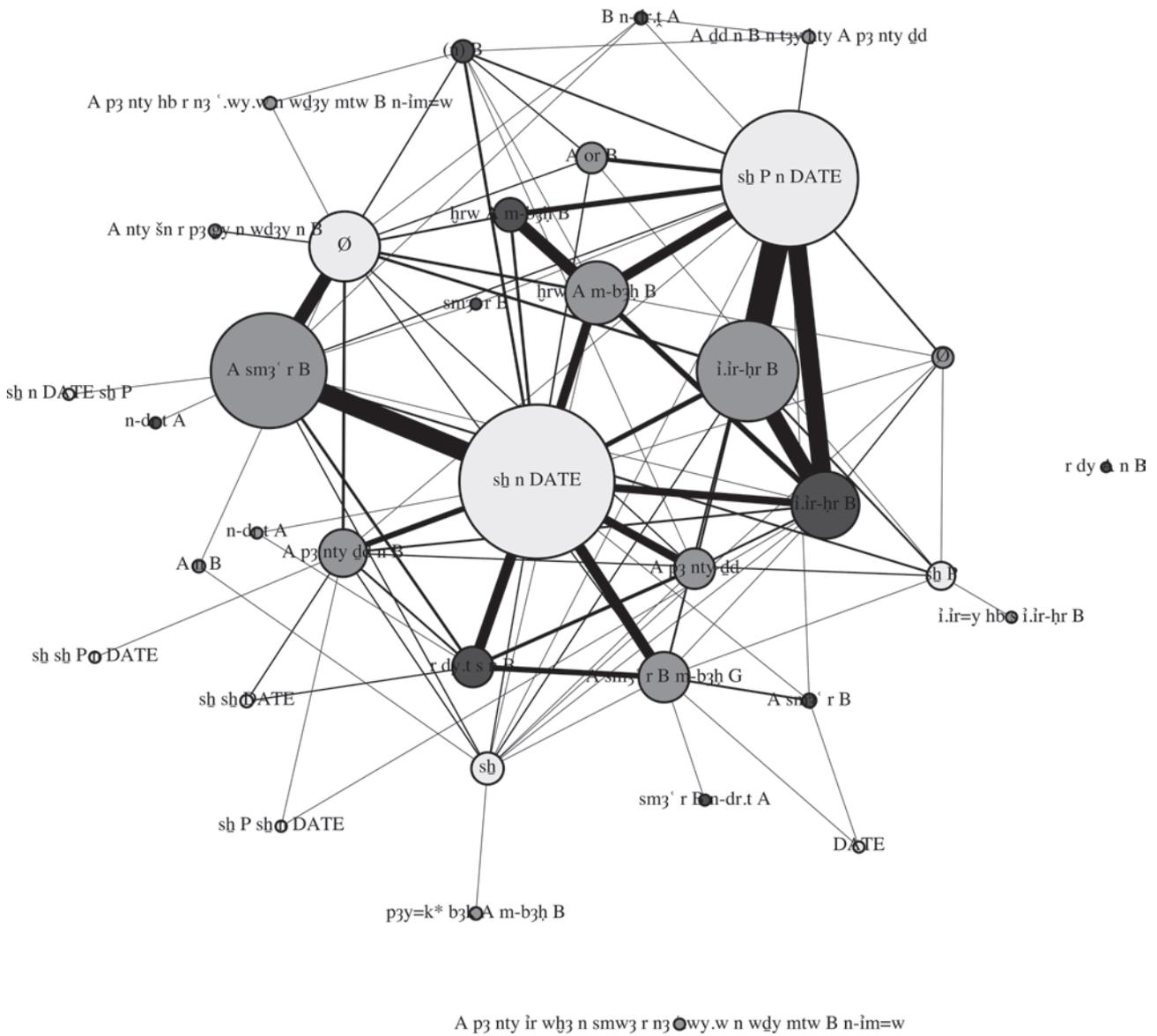


Fig. 1: Network of Demotic epistolary formulae³⁵.

³⁵ A full-colour and interactive version of this network is available online at www.trismegistos.org/network/15. Hovering over a node with the mouse pointer will result in the highlighting of the connections of said node. We strongly recommend the reader to consult the online version, here and elsewhere in the article, as it provides a much clearer picture.

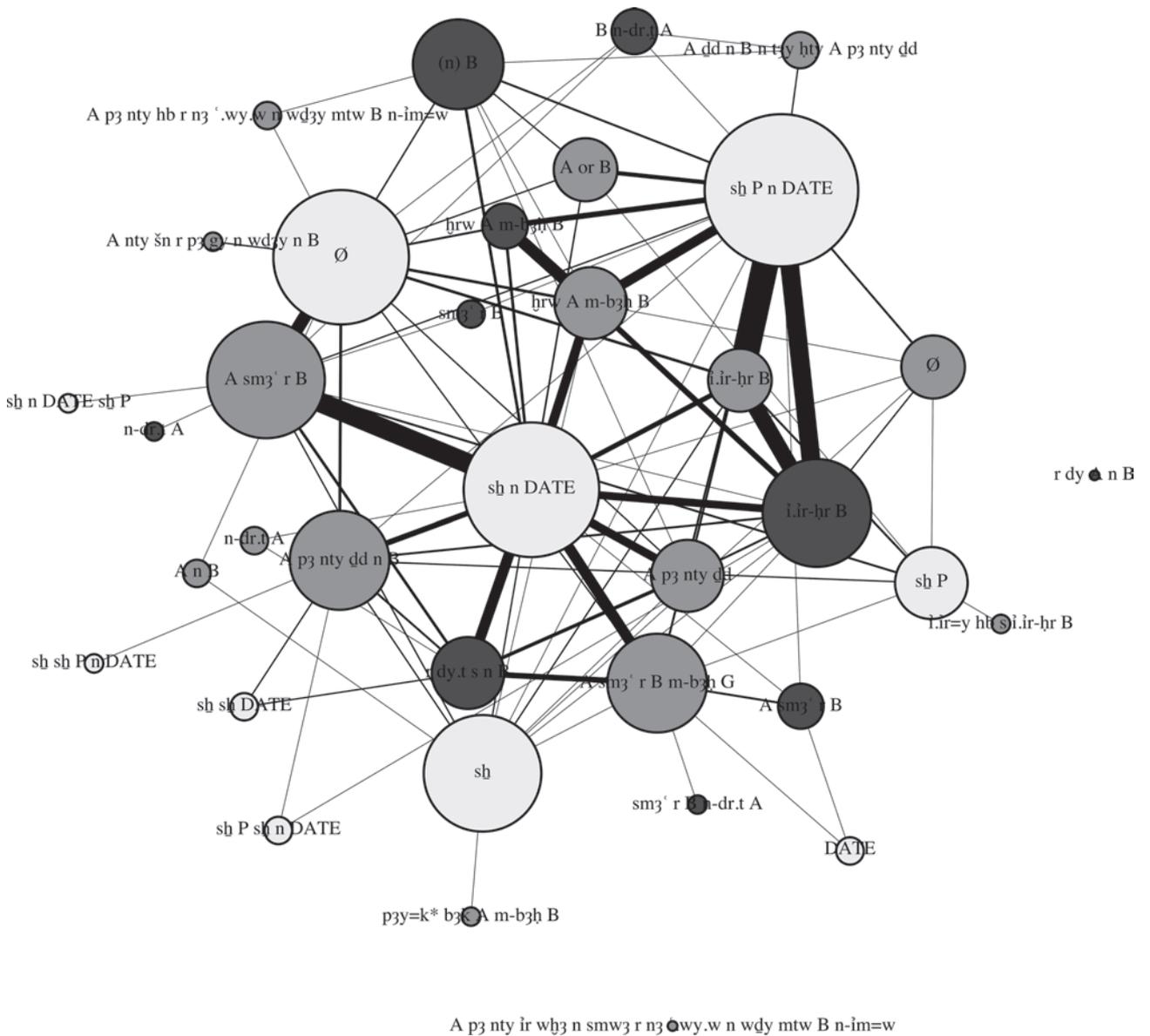


Fig. 2: Network of Demotic epistolary formulae (node size determined by node degree)³⁶.

³⁶ A full-colour and interactive version of this network is available online at www.trismegistos.org/network/16.

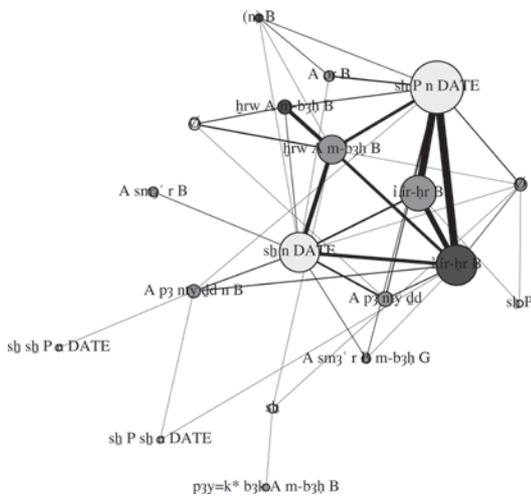


Fig. 9: 4th–3rd century BC⁴³.

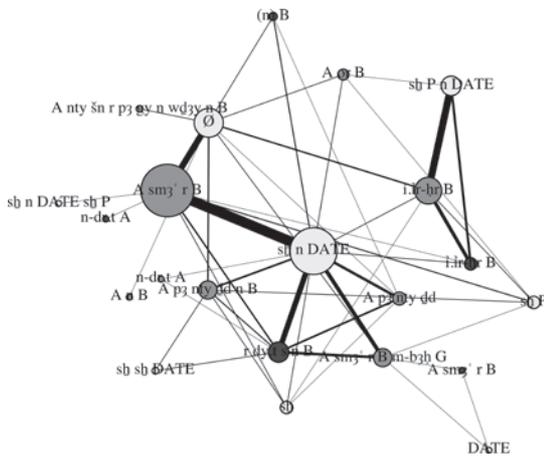


Fig. 10: 2nd–1st century BC⁴⁴.

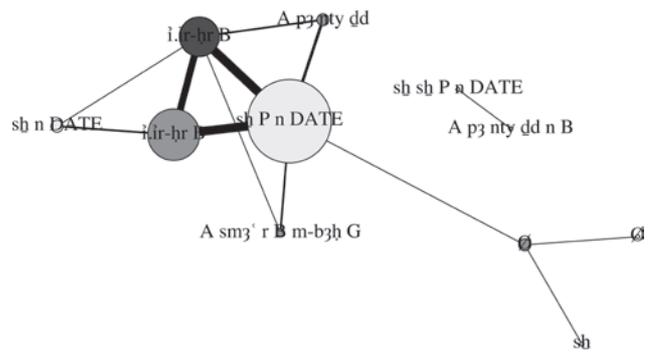


Fig. 11: Network of formulae in Tehna⁴⁵.

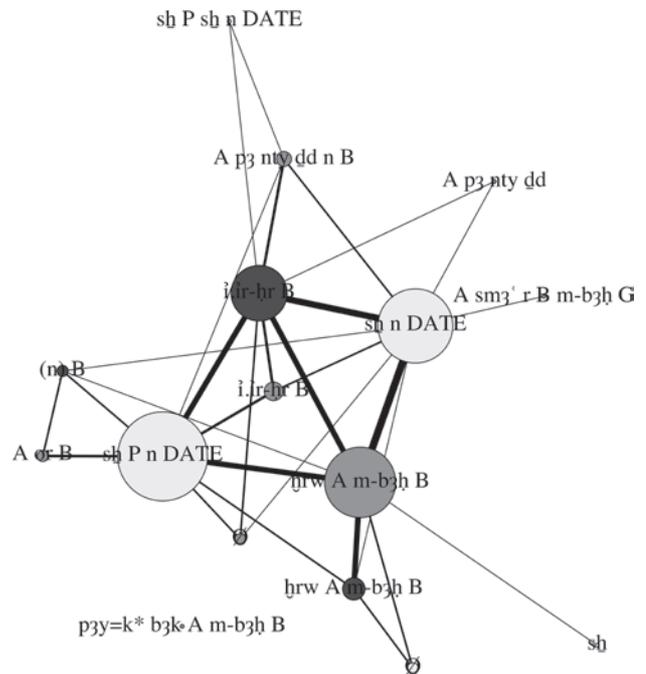


Fig. 12: Network of formulae in Elephantine⁴⁶.

⁴³ A full-colour and interactive version of this network is available online at www.trismegistos.org/network/23.

⁴⁴ A full-colour and interactive version of this network is available online at www.trismegistos.org/network/24.

⁴⁵ A full-colour and interactive version of this network is available online at www.trismegistos.org/network/25.

⁴⁶ A full-colour and interactive version of this network is available online at www.trismegistos.org/network/26.