The weird, wired past. The challenges of applying network science to archaeology and ancient history

Luce Prignano
IPHES Catalan Institute for Human Palaeoecology and Social Evolution

Sergi Lozano
University of Barcelona

Jordi Perez Gonzalez
Universitat de Girona

Francesca Fulminante
Università Degli Studi Roma Tre

Col·lecció
INSTRUMENTA 55

ECONOMÍA ROMANA. NUEVAS PERSPECTIVAS

THE ROMAN ECONOMY. NEW PERSPECTIVES

José Remesal Rodríguez (ed.)

UNIVERSITAT DE BARCELONA
Edicions

Índice general

Prólogo.

Recogemos en este volumen una serie de trabajos dedicados, unos a mostrar algunos aspectos vinculados a la investigación actual, relacionados con el estudio del *instrumentum domesticum* y la economía romana. Otros son la primera muestra del desarrollo de los nuevos enfoques surgidos del proyecto ERC Advanced Grant *Production and Distribution of Food during the Roman Empire: Economic and Political Dynamics* (EPNet) (ERC-2013-ADG 340828).

Hasta ahora, la aplicación de métodos formales, nacidos fuera del ámbito de la investigación histórica, está poco desarrollada dentro de nuestra especialidad. La "ominosa cuestión" de los estudios de Historia Antigua es la falta de datos. Los modelos interpretativos de la economía antigua han partido siempre de análisis deductivos, que dependen siempre del grado de conocimientos del investigador y de sus apriorismos. A lo largo de estos años hemos conseguido reunir una gran cantidad de datos, muchos de los cuales pueden ser presentados como datos seriales gracias a la información obtenida en el Monte Testaccio. Es ésta una circunstancia, la abundancia de datos y el poder ordenar cronológicamente muchos de ellos, es lo que permite los nuevos enfoques propuestos. En última instancia se trata de confrontar los modelos y explicaciones hasta ahora ofrecidas dentro del ámbito histórico, con modelos formales nacidos dentro del ámbito de las ciencias matemáticas y en el ámbito de las ciencias de redes.

Además, estamos haciendo migrar nuestra base de datos CEIPAC, ya puesta en Internet en 1995, a un sistema ontológico de bases de datos, en el que, gracias a un sistema de metadatos, podamos interrelacionar diversas bases de datos que amplien nuestros conocimientos y la capacidad de relacionar multiples aspectos de la investigación.

Dado que los trabajos presentados proceden de ámbitos científicos en los que los sistemas de citación son diversos, se han respetado los sistemas propuestos por cada uno de los autores.

The weird, wired past. The challenges of applying network science to archaeology and ancient history.

Luce Prignano (CLabB, UB), Ignacio Morer (CLabB, UB),
Sergi Lozano (Iphes - URV), Jordi Pérez González (CEIPAC, UB)[1],
Francesca Fulminante (Univ. Roma Tre & Univ. of Cambridge)[2],
Albert Díaz-Guilera (CLabB, UB).

Introduction.

Food webs, neurons in the brain, the Internet and the World Wide Web, have very little in common except for the fact that they all are systems composed by a large number of interconnected elements. Since the last decade of twentieth century, it has been suggested - and it is nowadays largely accepted - that, if one is interested in understanding anything with similar characteristics, it is appropriate to consider the possibility to model it as a *network*.

A network is a set of items, called *nodes* or *vertices*, whose pairwise relations are represented as connections between them, called *edges* or *links*.

The study of networks, in the form of mathematical graph theory, has developed - since the eighteenth century, but mostly during the last century - into a substantial body of knowledge. At the same time, already in the 1930s, sociologists realized the importance of the patterns of connection between people to the understanding of the functioning of human society. [cit 1]

[1] El presente trabajo se inserta dentro de los siguientes proyectos: Centro para el Estudio de la Interdependencia Provincial en la Antigüedad Clásica (CEIPAC) (2014 SGR 218) y HAR2015-66771-P (MINECO/FEDER, UE)
[2] Francesca Fulminante is grateful to the European Community support to her career and research through the FP7-IEF 628818, Marie Sklodowska Curie Actions Past-People-Nets.

Recent years, however, have witnessed a substantial new movement in network research, with the focus shifting away from the analysis of the properties of individual vertices or edges within small systems to consideration of the statistical properties of networks [cit 2, cit 3, cit 4, cit 5, cit 6].

In this new phase, the communities of physics, computer science, and applied mathematics have taken the role of the principal actors. Since the beginning, these researchers - to whom we will refer as network scientists - have exploited the possibility to study plenty of large databases, directing their efforts to the identification and characterization of universal classes into which, in principle, any real networks can be assigned.

As a consequence, the fields of application of the new analytical techniques have increased enormously and currently include biology, ecology, chemistry, neuroscience, logistics, among others.

Nowadays, it is a common knowledge that, regardless of the specificities of their research domains, scholars from different disciplines can find in this approach a valuable ally when tackling complexity.

If ecologists want to know how critical the survival of some species is for the stability of a given ecosystem with intricate trophic relationships, network science can be helpful [cit 7]. If physicians need to identify genes playing a major role in determining the clinical outcome of a disease, they can find network based techniques that have been designed for this aim [cit 8, cit 9]. If public health agencies would like to know to whom they should recommend vaccination in order to prevent some virus from spreading, there are plenty of network models to suggest them how to select people for immunization [cit 10].

Ideally, network science advances through the combination of two complementary research approaches. The first one corresponds to when network scientists, looking at networks as abstract mathematical objects, identify a general question or problem and develop a method for addressing it. The second one is what researchers from any other field do when, trying to extract information from some data, find that the limitations of other existing methodologies prevent them from reaching their goal and come to the conclusion that adopting a network science approach may be the solution. In the first case, a “universal toolbox” (or theory) grows by abstracting from individual case-studies. In the second one, the understanding of a particular case-study (application) advances by applying the appropriate universal tool.

The main reason this linear way to progress is not realistic is a semantic issue. It is not a trivial task to translate into the specific language of each discipline questions that are expressed in terms as general and abstract as those normally used in network science. Moreover, if the choice of the appropriate technique is a difficult matter, determining the reliability of the output and interpreting it is even more complicated.

Additionally, such techniques have been developed for the analysis of available real networks, that is, publicly accessible digital data whose features have inspired the questions that network scientists considered worth to be addressed. Therefore, if the new set of data that one wants to analyse is too different from those studied till that moment, the appropriated technique may not have been invented yet. For instance, in some cases nodes are entities located in a geographical space (spatial networks [cit 11]). In some other cases, nodes represent objects that belong to different classes (e.g. affiliation networks [cit 12]) and can only connect with elements of the other class (bipartite networks). For this kind of situations, tools developed to deal with data that do not have the same properties may lead to erroneous results. Sometimes, it is up to the researchers working on a particular case-study to adapt the existing techniques to the particular features of their data. However, they may prefer to design a way to pre-process the data in order to make existing analytic tools suitable for them. Some other times, network scientists may work on newly available data with novel properties, elaborate the questions and then create new tools. In many cases, both things happen at the same time and we have redundancy of questions and techniques expressed in different terms.

Another relevant aspect is that data do not come in network shape by themselves. It may happen that there is only one way to map them into vertices and edges, but this is not always the case. Consider, for instance, any of the online social networks. The most obvious approach is probably to consider users as nodes and friendships or following relationships as links. Although, it makes equally sense to take into account the interactions (like/favorite, share/reblog, messages etc.). In principle, depending on what the question is, researchers would choose the most appropriate network representation. However, if those carrying out the study are network scientists, they may not know enough about the details of the information embedded (or discarded) in the data and how it is connected to the possible research questions. On the other hand, if they are the experts who collected the data, but do not know much about the technicalities of network science, their choice may be influenced by the need to "keep it simple".

At a more general level, it is worth noticing that it is not always immediately clear if a network approach can be useful in any way. In some cases, it is self-evident because interactions constitute the very basis of the phenomenon under study (contagion processes, trophic relationships, cascading failure...). However, in some other scenarios, the relational aspect of the data is just a secondary feature that can be included as a refinement and adopting a network perspective may or may not represent an improvement compared to previously existing methods. There exist multiple ways to process and extract knowledge or insights from data in various forms and most of them are older than modern network science. It is important to know the possible alternatives and to compare the results obtained through different methods. As a paradigmatic example, we can consider the task of grouping a set of objects. Such task can be performed through cluster analysis [cit 13, cit 14], collecting all the relevant information that define the objects and then measuring their similarity in order to classify them. The main idea is that that those belonging to the same group (or *cluster*) are more similar to each other than to those in other clusters. But if such objects are nodes of a system mapped onto a network, we can also apply community detection algorithms [cit 15] to perform the same task. Community structure, i.e. the organization of vertices in clusters, with many edges joining vertices of the same cluster and comparatively few edges joining vertices of different clusters, is one of the most relevant features of graphs representing real systems. Cluster analysis relies on the characterization of individual objects, while looking for communities within a network implies that the focus shifts to their relations. Sometimes the results obtained with the two approaches are similar, some other times they are not. In any case, their interpretation is not the same. In order to choose the most appropriate approach, assessing the reliability of their outputs in abstract, statistical terms is crucial. Nevertheless, it is important not to forget the ideas underlying each approach, and their different algorithms, and how suitable they are for addressing a specific research question. Performing such evaluation requires expertise from different fields that cannot be easily found in a single researcher or research team. Obviously, this is just an example, but similar situations are ubiquitous when it come to applying network science, or any other quantitative method, to a new class of case-studies.

Quite obviously, in order to avoid most of the shortcomings mentioned above, interdisciplinary collaboration is the way to go, even if a slow one. Researches carried out within collaborations between network scientists and experts of the case-study provide, arguably, the most interesting and reliable results. Unfortunately, the initial phase of mutual learning can get unfairly time demanding, especially the first time two disciplines try to communicate, when a common language needs to be established. The complexity of the task increases somehow proportionally to the distance between the discrete, often binary, quantitative language of network science and the language in which raw information is expressed.

Humanities are usually the hardest and History and Archaeology are no exception. Among the networks used as benchmarks for testing new techniques, there are networks constructed from data of airports and flights, of web pages connected by hyperlinks, of physical contacts between proteins, of functional parts in the cortical brain of the cat or other mammalians; data about social grooming behaviour among primates, about the world trade web, about who hangs out with whom and many more. None of them have been published in history or archaeology journals, not even the one whose nodes are Florentine families of the XV Century [cit 16] which was built by political scientists meddling with history. Networks about Ancient History or Prehistory are nowhere to be found in papers and conferences (co-)authored by members of the network science's community.

However, this does not mean they do not exist. On the contrary, the number of papers on network applications to archaeological case-studies appeared on archaeology journals has been increasing continuously during the last decade [cit 17].

Formal network methods have been applied to explore research topics as diverse as the transmission of ideas [cit 18], the movement of people and objects [cit 19, cit 20], the identification of social and cultural boundaries [cit 21], interregional interaction [cit 22], and maritime connectivity [cit 23]. Besides the increasing number of journal articles, in the last four of five years, books and special issues have also started to come out, collecting contributions from tens of authors [cit 24, cit 25, cit 26, cit 27].

Nevertheless, there are several reasons why these works occupy an extremely peripheral position within the ecosystem of network science. Some of them can be regarded as "non-scientific" factors. For instance, the fact that a large majority of these papers were published quite recently, when the big hunt for benchmarks was already over. Moreover, it is not a common practice among researchers in humanities to publish datasets along with the results in their article and therefore such networks have very few chances to circulate. Other factors are instead entirely inherent to the peculiar nature of archaeological networks and their construction.

The resistance offered by raw historical or archaeological data is so difficult to overcome that it is almost impossible for network scientists to build networks by their own. To build a network from the archaeological record is challenging because, in general, one has to face all the typical issues of the other classes of data at once.

The main difference is that, normally, natural sciences and, to a less extent, social sciences collect the data they need by carrying out appropriate experiments, while this is not the case for historians and archaeologists. Even though occasionally scientists have no other option but relying on observation alone, it goes without saying that it is when one is investigating past processes that the exception becomes the rule. Therefore, difficulties that researchers from other fields have to face more or less sporadically, archeologists and historians have to deal with them all the time.

Network inference and the study of human past

In order to outline the most common difficulties, let us summarize here the basic ingredients for a proper network representation of a system:

1. A definition of such system that allows to identify its boundaries, separating what is within from what is outside.
2. A definition of the elemental parts that will constitute the nodes of the network
3. A definition of what the connections are supposed to mean and a well-defined way to determine whether they exist or not, that is, a way to measure them, or, if not possible, to infer them.

Depending on the circumstances, each one of these three ingredients may present different challenges. Before entering into details on the data-related questions, we would like to briefly discuss some ambiguities in the concepts and definitions that may be problematic even when the best possible data are available.

A matter of borders

The first ingredient may seem trivial, but it is not. It is not so infrequent that the system under consideration is indeed a part of a larger one with blurred borders. Such borders can be conceptual, spatial, or temporal, being the last situation especially relevant for historical and archaeological case-studies. In order to clarify what conceptually blurred borders are, let us consider an example involving acknowledged "good data". If we want to follow the rise and fall of a topic on Twitter, we can consider all the tweets containing a given hashtag. Alternatively, we can include those containing a counter-hashtag as well. And perhaps other related ones. If we are interested in the behavior of a specific age-class, then we will exclude all the users whose age is not in the required range. Is it the right choice? What are we doing with those whose age is unknown? Should we leave them out? If we exclude users that are not in the category of interest, what happens with the interaction between those that belong in our system and those who do not? The typical issue related to spatial borders concerns the interactions with what is outside such borders. In a network where nodes represent settlements and links some kind interaction between them - for instance, routes or commercial exchanges - the decision about where to draw the frontiers of the system can be crucial. Even if the system under study is a political entity with well defined geographical borders, it can nonetheless be unwise to cut out everything that does not belong to that entity. Immagine one is interested in knowing whether settlements that are known to be important from other evidences or sources are also the most central ones according to some network analysis measure. Disregarding everything that is outside the borders, will make the node representing an important city connecting two regions as peripheral as any small village close to a desertic area. In a network made up of many nodes, from thousands to millions, vertices at the border represent a very small fraction and this kind of issues are are just unimportant nuisances. On the contrary, when dealing with small systems, issues related to spatial borders need to be carefully tackled. Finally, establishing limits at the temporal dimension also give rise to some challenging questions.

Nodes can be created and destroyed; sometimes one splits into two, sometimes two merge into one. Connections appear and disappear; links may increase or weaken their strength. When dealing with systems that evolve over time, the so-called longitudinal networks, it is difficult to capture meaningful information in a simplified manner. Imagine someone trying to take a picture of something that is moving. The photographer surely will choose a fast shutter speed. But the temporal resolution of archaeological data is limited. It is like being in a dark place, unable to see subject clearly. The challenge is how to find the best tradeoff between a blurred and a dark image, that is, to select the appropriate time window when trying to reconstruct an evolving network.

The choice of the building blocks

The definition of the nodes may represent a real difficulty if there are more than two scales (local and global) involved and are not clearly separated. If there are buildings, blocks, neighbourhoods and cities, it is crucial to make a choice. But is also important to know what to do with small towns, rural areas, or the outer districts of large metropolis. The spatial resolution can be not homogeneous enough. Geolocalized data from mobile devices have precise coordinates associated to them, but in many situations it can be more interesting to aggregate them into larger geographical units.

In most archaeological applications, nodes are contexts or attributes of contexts, that is, any kind of archaeological evidence. Archaeological findings in some cases can be naturally grouped together depending on the context they belong to (tombs, houses, settlements, etc.), but that may also be scattered over areas where no other remains have been found. Is it better to discard such records or should we aggregate them according to some criterion? Spatial nodes are not the only ones facing these dilemmas. If we are considering amphoric types or ceramic compositional groups, how are we supposed to deal with geographic variations or imitations? Basically, it is the issue of discretizing a nearly continuous spectrum of differences. We have already explained how cluster analysis algorithms can be helpful to group or classify objects based on their individual properties. Alternatively, it is also possible to accept different hypotheses, defining for each of them a different set of nodes and, consequently, a different network. The properties shared by a majority of such networks are robust to the uncertainty in the node definition and hence can be regarded as reliable.

From interactions and similarities to connections

Connections present an even broader range of issues. Obviously, in order to determine if a link does or does not exist, we have to rely on the available data. However, data do not always have univocal interpretation since, in principle, what defines the meaning of connections is the research question. Some networks, as for instance infrastructure networks, have connections that represent material objects: streets, cables, railways. Some others have links whose existence is established through an operational definition, that is, through a process, something the nodes do: there is a links from specie A to specie B if members of A eat members of B; there is a link between author A and author B is they have co-authored at least one research paper; there is a link from user A to user B if A follows B, etc. In all these cases, little room is left for ambiguities and interpretations. Other kinds of networks, however, have connections that represent a more ontological relationship between nodes: friendship, commercial partnership, political affinity, among many others. The existence of such relationships has to be inferred from the information available which need to be filtered and interpreted. Suppose that we have collected all the emails that people working at some company sent to their colleagues during the past month. The links extracted from this dataset will not be the same if one intends to build a friendship network or a network of work-related cooperation. Then we have biochemical networks, which belong to the second category in theory, but in practice they do not, simply because not all the relations are directly measurable, not even in natural sciences. Think for example of a protein-protein interaction network, whose links are clearly identified through an operational definition, that is, physical contacts established between two proteins as a result of biochemical events and/or electrostatic forces. In order to build a network it is necessary to integrate data from different experiments that provide indirect evidence of lasting physical contacts that are difficult to detect directly.

Whenever direct observation is out of reach, regardless whether the cause is the nature of the relations or technical issues, links need to be inferred.

Typically, when trying to reconstruct past networks from archaeological evidences and historical sources, we have to face both difficulties. Many of the research questions that can be addressed through a network science approach are such that connections have an operational definition. From migratory flows and colonization processes to trade dynamics and commercial routes, to the study of the diffusion of a technological innovation or a cultural trait, nodes are generally social groups associated to places and links are interactions defined by a specific process. Hence it should be easy to determine if a connection did or did not exist and, in principle, to estimate its strength. In practice, gathering clear evidence of a certain kind of interaction is not always feasible. One possible option is to rely on statistics in order to determine the probability associated to existence of each link [cit 28]. Alternatively, one can relax the restriction on the type of interaction, a choice that allows to combine together all the available evidences but that, at the same time, gives rise to other issues.

It leads to building generalized networks [cit 29] where the meaning of the links has a less precise definition.

Suppose we have some data about a set of settlements that coexisted in time when a given process of interest was taking place. By taking into account any class of remains within some chronology and, when available, historical sources, we can deduce how culturally similar they were at those times. However, links established from cultural affinity are not derived by an operational definition. It is more the opposite: they are similar to the ontological type of connection we have discussed above. Similarity, in its broadest sense, is not a measurable quantity. One can quantify it, obviously, but there are many ways to do so [cit 30] and the procedure contains an ineludible amount of arbitrariness. Beside these technical issues, it is worth noticing that the relation between the ongoing process of interest and the similarities observed needs to be demonstrated, or at least upheld by heuristic arguments. Hence, basically, while the amount of data increases, drawing conclusions from them becomes more complicate.

Real case-studies, however, often enough allow for a mixed approach, something that is midway between looking for a direct proof of interactions of a specific type and including any kind of evidence, thus obtaining a too general cultural similarity measure. In many cases it is possible to use a variety of evidences somehow related to the process under study, then trying to infer the connections by means of a combination of tailored similarity measure and statistical tests. In the next sections, we will further explore this scenario through few illustrative datasets.

When it comes to nodes (sites) and attributes (artifact typologies), it is easy to identify two opposite situations between which all the possible spectrum of real cases would lie, that is, two extreme scenarios corresponding to maximal and minimal diversity among sites, respectively. We may have that each of the considered nodes has completely different attributes, which means that they are not related in any way and therefore in our similarity network there is no link that can be drawn. We would have a set of isolated nodes. The opposite to this scenario is a situation where all the attributes are equally distributed among the sites, each one of them having the same proportion of artifact typologies. In this case, the similarity measure will be maximal - regardless the details of its definition - and each pair of nodes will be connected by a link. We would have a fully connected network. None of these extreme situations is realistic or interesting. In real cases, we expect some groups of sites (clusters) to be more similar among each other than to any other node in the system. The similarity network would then have a community structure and we could analyse how its communities are organized both internally and among each other. As a paradigmatic example, let us introduce a sample data from Huntley's [cit 31,cit 32] study consisting of counts of five ceramic compositional groups from nine sites in the Zuni region.

		CERAMIC COMPOSITIONAL GROUPS					
		DLH-1	DLH-2a	DLH-2b	DLH-2c	DLH-4	Total
SITES	Atsinna	16	9	3	0	1	29
	Cienega	13	3	2	0	0	18
	Mirabal	9	5	2	5	0	21
	PdMuertos	14	12	3	0	0	29
	Hesh	0	26	4	0	0	30
	LowPesc	1	26	4	0	0	31
	BoxS	0	11	3	13	0	27
	OjoBon	0	0	17	0	16	33
	Sp170	0	0	18	0	14	32
	Total	53	92	56	18	31	

Table 1. Sites and ceramic compositional groups of the Zuni dataset.

We chose to measure the similarity between this sites by means of the Brainerd-Robinson (BR) coefficient because its definition is intuitive and it was developed within archeology specifically for comparing assemblages in terms of the proportions of types or other categorical data. BR is a city-block metric of similarity that is calculated as:

BR(i,j)=200-k=1pPik-Pjk

where, for all categories (*k*), *P* is the total percentage in assemblages *i* and *j*. This provides a scale of similarity from 0-200 where 200 is perfect similarity and 0 is no similarity [cit 33] From a cursory inspection of the frequencies histogram of the BR values (Fig. 1), we can easily observe that the pairs of sites naturally group into three classes: pairs having none or very little similarity (BR<30); pairs that are weakly similar (50<BR<120); pairs whose similarity is stronger than average (BR>130).

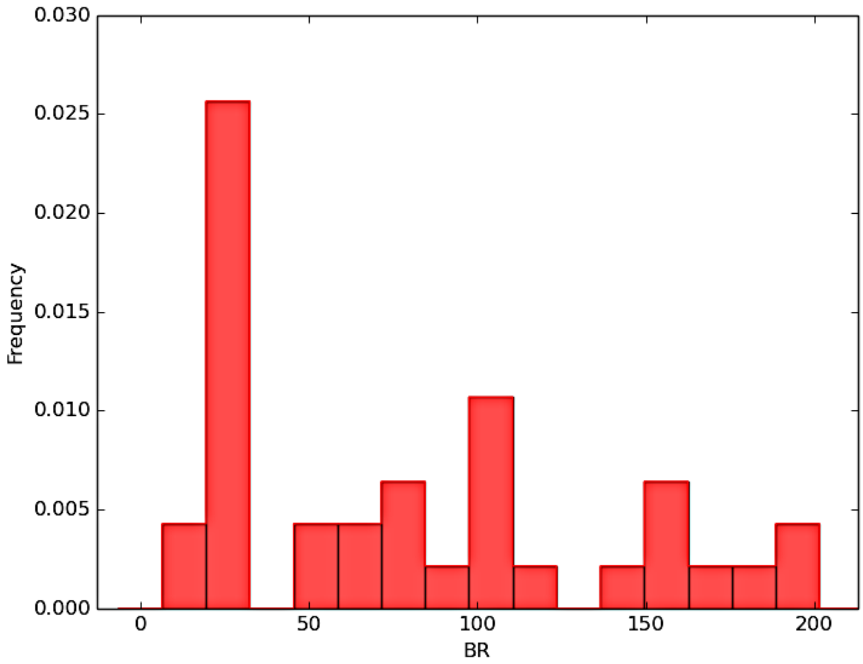

Figure 1. Frequency histogram of the BR values for the Zuni dataset.

If we build the similarity network (Fig. 2) considering only the strong similarities (thick purple links), we will recover the three known clusters of settlements defined accordingly to their geographic location. Including the weak similarities, that is, adding weaker (thinner, grey) links to the network, we can see how two of them are related to each other and integrate the only site that does not belong to any cluster, while the third one is completely separated from the rest.

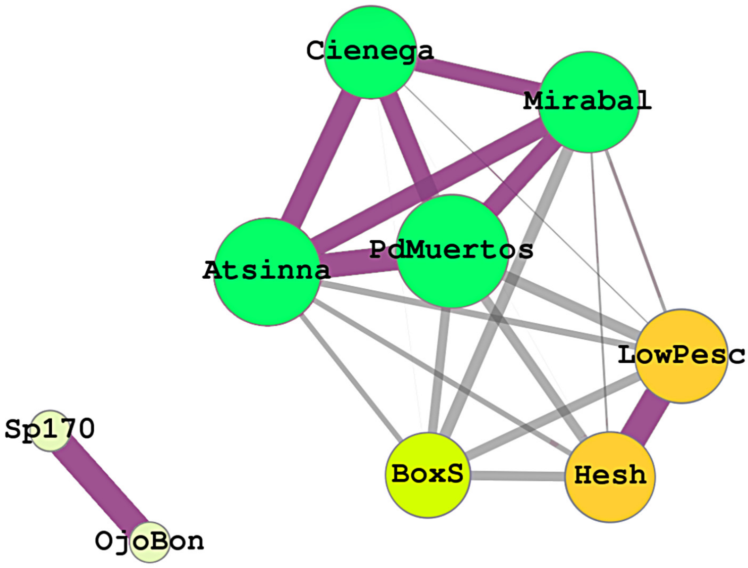

Figure 2. Zuni dataset network of sites constructed according to the BR coefficient. Thicker links correspond to higher BR values; purple links correspond to BR values larger than 130. Node size is proportional to the size (number of artifacts) in the assemblage from the corresponding site.

However, this simple example does not represent the most typical situation. Things are usually not this clean and self-evident. First of all, the appearance of three groups of BR values that can be interpreted as non-existent, weak and strong similarity is quite a rare fact. More commonly, BR values are continuously distributed in a more or less broad range of values. Moreover, in the Zuni dataset, both the five categorical attributes (compositional groups) and the nine nodes (assemblages) have a comparable number of samples, a feature that is not usual when dealing with archaeological data.

Assemblages with heterogeneous size: looking for significant similarities.

Small assemblages are problematic because their composition is likely to be non-representative of the affiliations of the site. A reduced number of samples can be the results of a poor conservation or it may be the consequence of the actual small size of the excavated area. Sometimes, it can also indicate a real lack of the considered categorical attributes. Therefore, comparing their proportions may lead to wrong conclusions. We should be able, with the help of appropriate statistical techniques, to differentiate between the first two cases and the third one. Consider the ideal scenario of perfect uniformity among the sites. If the entirety of the assemblages is conserved, all the BR values will be equal to 200, but if only a fraction of them survived, then we would obtain smaller values as a consequence of sampling error. Suppose that this fraction of survived objects is small and that the size of the sites varies considerably. How can we discern whether what we observe are random differences within a group of highly damaged identical assemblages or instead represent something significant, inherent to the sites under study? Fortunately enough, the first case can be artificially reproduced by computer simulations. Actually, it is possible to reproduce such scenario thousands of times. Random samples of a specified size, based on the actual number of objects in each site, are drawn with replacement from a population with proportions defined by the actual total number of artifacts in each category. Then, we can compare the empirical data with the synthetic ones, determining the probability that what we observe from the archaeological evidence is "signal" and not "noise", feature by feature. In particular, we will discuss the application of two statistical tests: one addressing the statistical significance of the similarity between pairs of nodes with given sizes; the other tackling the issue of the how significant the presence of certain category of data is in a certain node, given the relative abundance of objects belonging to that category and the size of the node. As a comparative example, we consider two datasets with opposite properties. The first one (from now on, DP), is not from an archaeological or historical case study. It is the result of a poll performed by the *Spanish Centro de Investigaciones Sociológicas* (CIS) about the political affiliation of the congresspersons in the Spanish Parliament after the general election of June 26th 2016. This time, the districts are the nodes, the political parties the categorial attributes, and the elected parliamentarians the objects or artifacts. The properties of this system are known, hence it can be efficiently used as a benchmark dataset. The second set of data (GS) we are going to analyse is from the EPNet project (ERC-2013-ADG 340828). The nodes are settlements in the province of Germania Superior, while the categorial attributes are the different stamp-types found on olive oil amphorae, and each piece of amphora with a stamp is an object or artifact. The first dataset has much more nodes than categories and is quite close to the scenario of perfect uniformity, except for some districts that have their own political parties. The second one, on the contrary, is a case of extremely high diversity, with many categories, a large majority of them having just one or two representatives, and few nodes. In Fig. 3 it is shown the network representation of the associations between nodes (sites, in purple) and categorial attributes (stamp-types, in light green) present in the GS dataset.

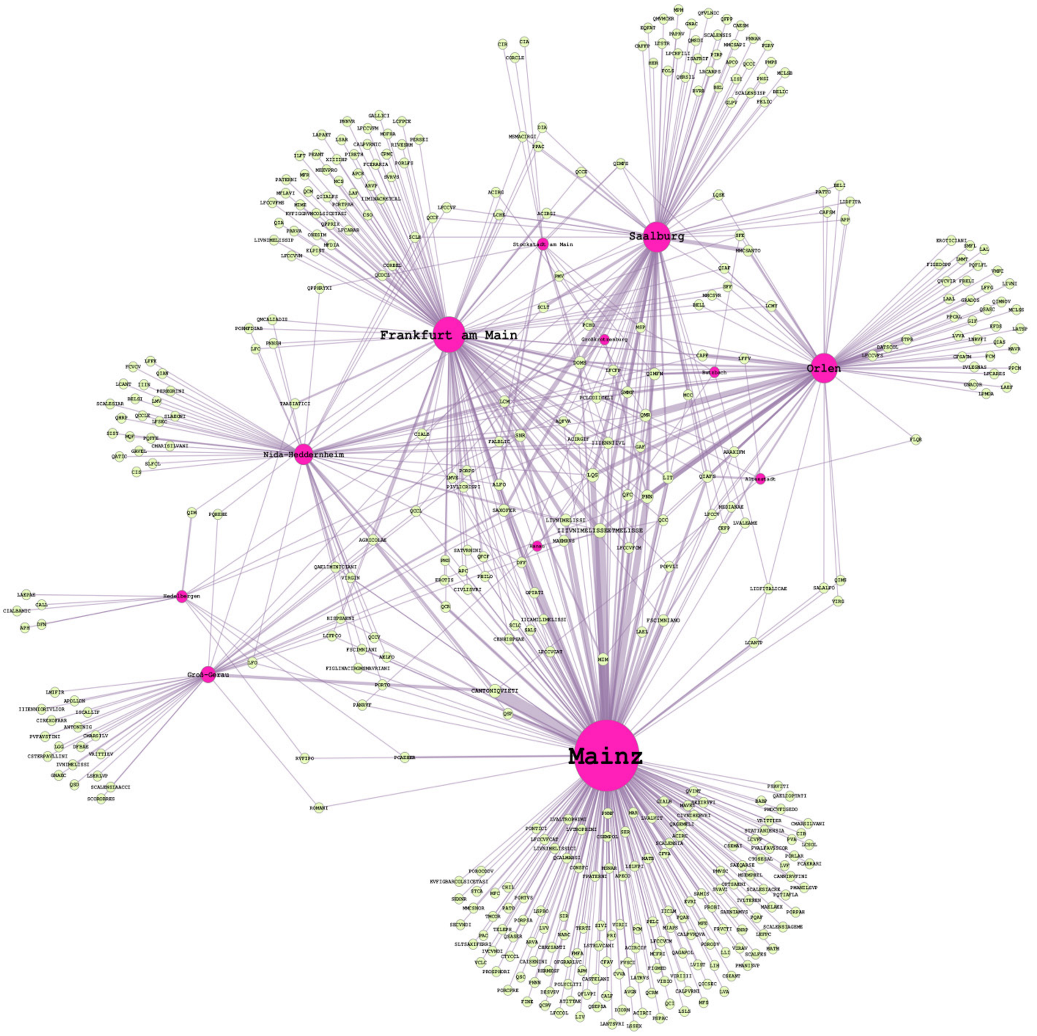

Figure 3. GS networks of places (sites) and stamp-types (categorial attributes). Each node representing a place (in purple) is connected to stamp-types (light green) that have been found in its assemblage and its size is proportional to the total number of elements (stamps) in it. The size of nodes representing stamp-types is proportional to the number of stamps in that category. The thickness of the links is proportional to the number of stamps of a given stamp-type that have be found in a certain place.

The differences between the two datasets can be illustrated effectively enough through the frequency histogram of the respective BR values (see Fig. 4).

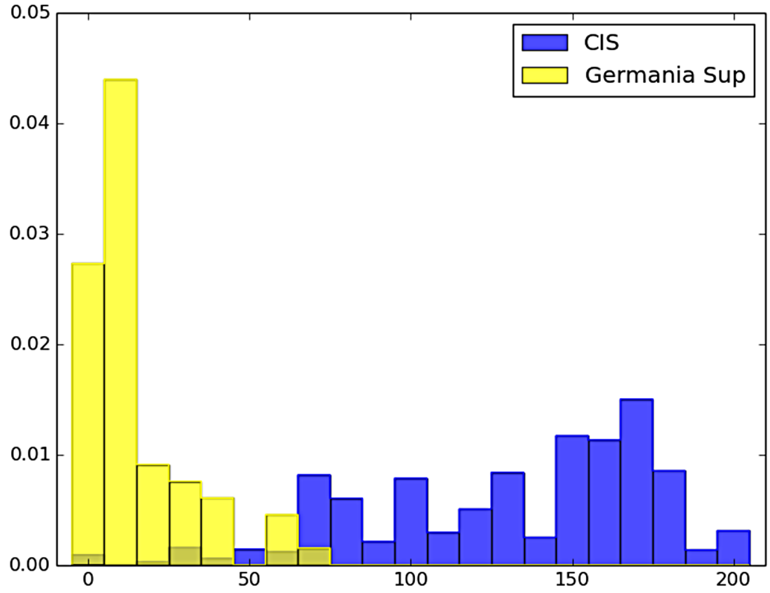

Figure 4. Frequency histogram of the BR values for the GS (in yellow) and DP (in blue) datasets.

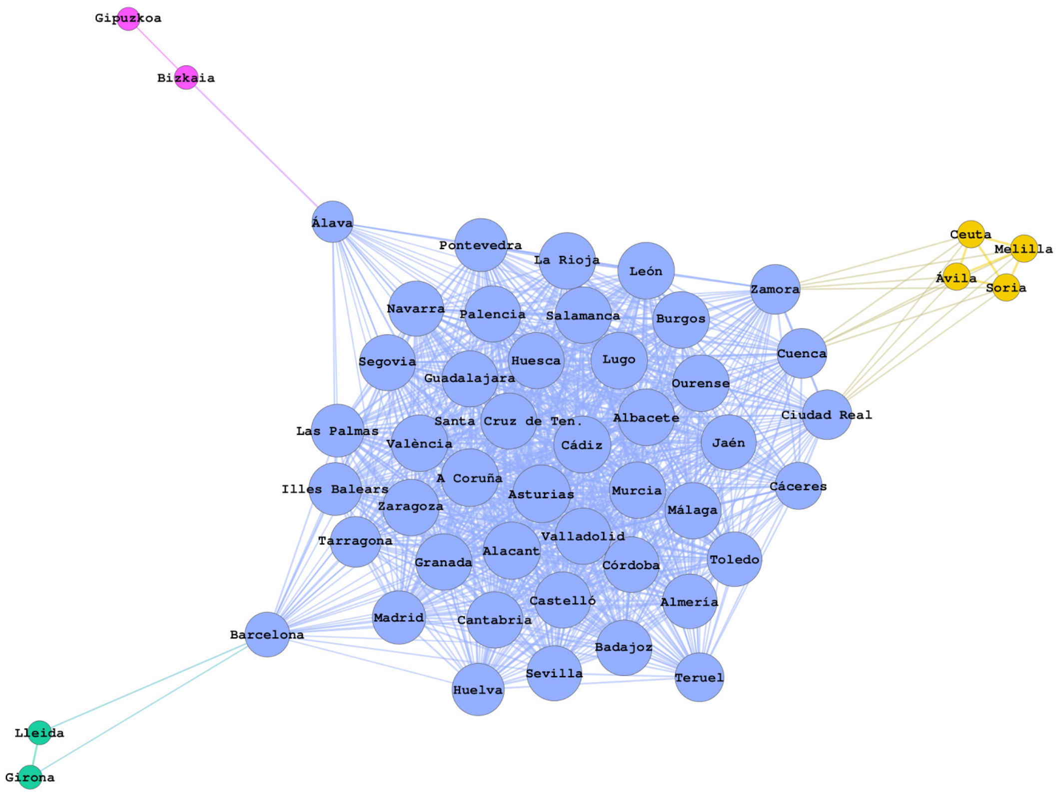

Figure 5. DP network of districts constructed according to the BR values. Node pairs whose BR coefficient is greater than 115 are connected and the thickness of the link is proportional to their similarity value.

In neither of these datasets it is clear where one should put the threshold in order to separate BR values that have to be interpreted as links and values that do not represent relevant relationships. Obviously, the threshold value cannot be the same in the two cases. If we set it in such way that each site has at least one link, we obtain the networks displayed in Fig. 5 and Fig. 6, respectively, for the Spanish districts and the archaeological sites in Germania Superior. The threshold value has been set at BR=115 for DP and BR=14 for GS.

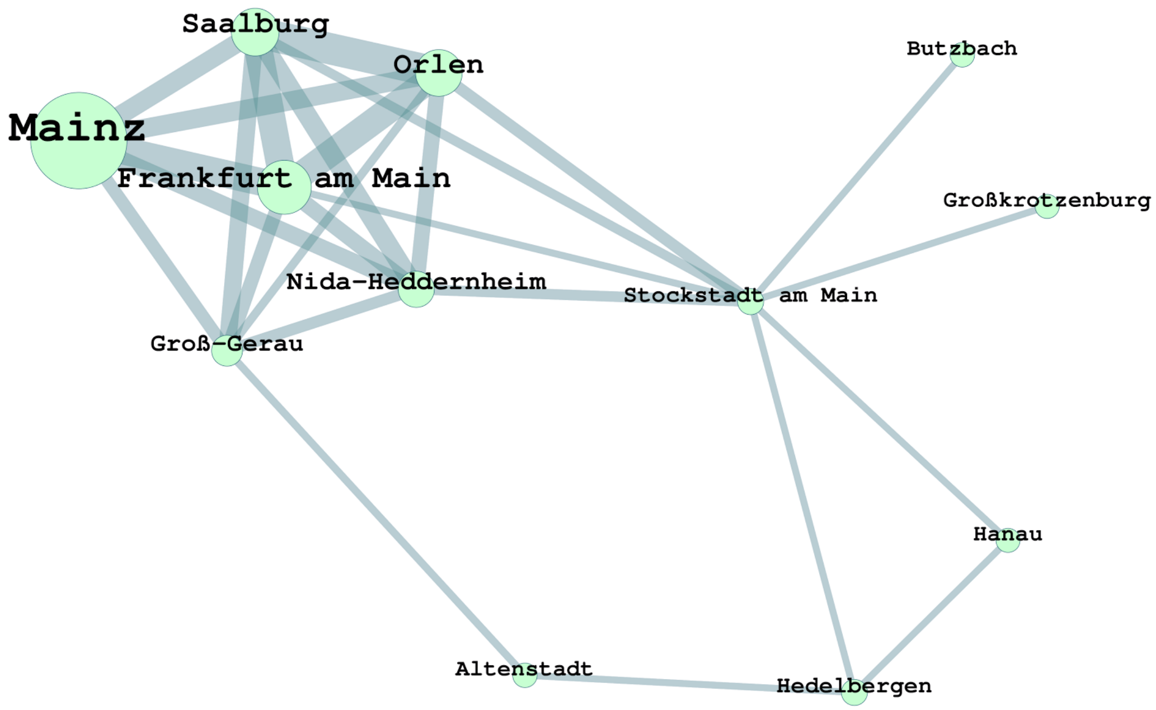

Figure 6. GS network of places constructed according to the BR values. Node pairs whose BR coefficient is greater than 14 are connected and the thickness of the link is proportional to their similarity value. Node size is proportional to the number of stamps that have been found in the corresponding site.

The statistical test concerning the significance of the similarities answers the question about how confident we can be that the measured BR values would have not been observed in random samples. A certain degree of similarity can be due to fact that, mixing the attributes randomly, some samples happen to be more alike, just by mere chance. This is especially likely to occur if there is a limited number of categories compared with the samples' size. Hence, the idea is that only the BR values that are very unlikely to be observed in randomized samples can be regarded as significant. Therefore, the correct criterion for adding links is not that of connecting pairs of sites whose similarity is above a given threshold value. A more appropriate criterion would be to add links between pairs whose similarity has a very small probability to be observed. This approach allows to take into account spurious size effects that would be neglected otherwise. Consider the case of Groß-Gerau, in the GS dataset. Its similarity with Saalburg is the highest one, being BR(Groß-Gerau, Saalburg)=35, while it is quite less alike to Altenstadt (BR=16). Nevertheless, Saalburg with its 121 stamps is a quite large site, while Altenstadt has only 6 stamps. It is not infrequent to obtain a similarity larger than that between Groß-Gerau and Saalburg in random samples of the respective sizes. On the contrary, this is not the case for a BR value equal to 16 for a sample of the size of Groß-Gerau (34 stamps) and another one as small as Altenstadt, as clearly shown in Fig. 7.

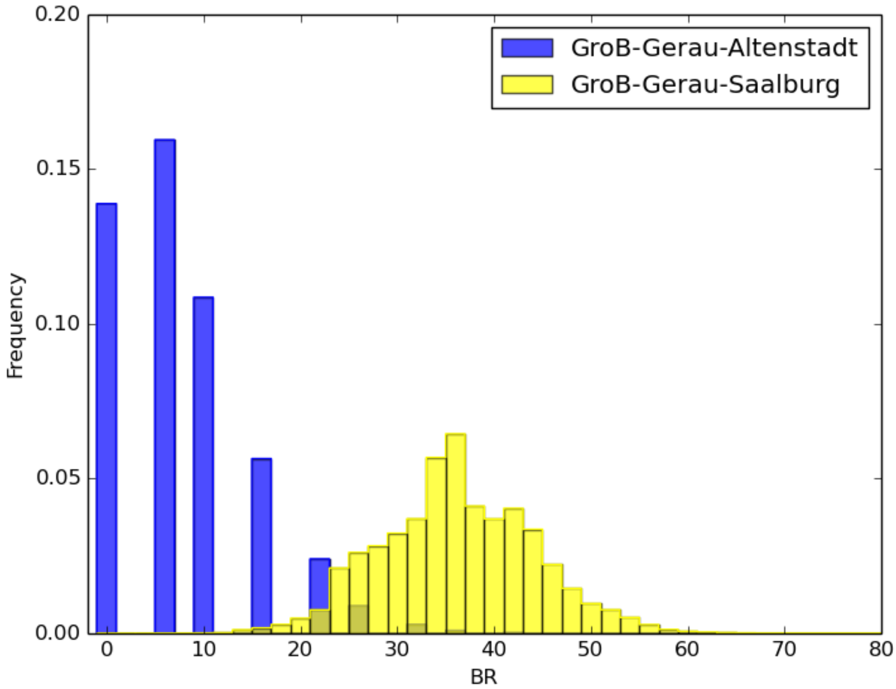

Figure 7. Frequency histogram of the BR values between randomized versions of two pair of assemblages: those of Groß-Gerau (34 stamps) and Altenstadt (6 stamps), in blue, and those of Groß-Gerau and Saalburg (134 stamps), in yellow.

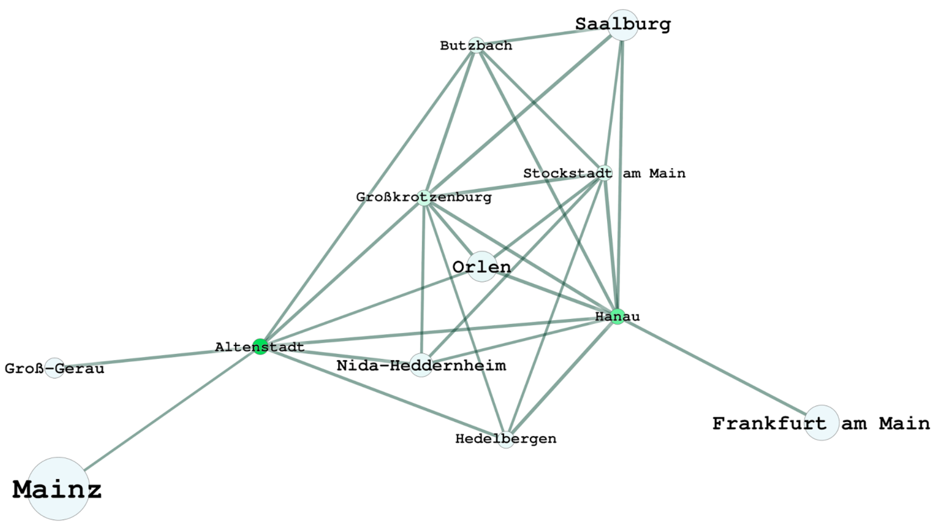

Figure 8. GS network of places constructed according to the probability that the BR value measured for a pair of nodes is larger than that of the randomized version of their assemblages. Nodes highlighted in darker green are those preventing it from becoming disconnected

Consequently, in a network constructed according to this new method, we will have a link between Groß-Gerau and Altenstadt, but no connection between the Groß-Gerau and Saalburg. In Fig. 8 we represent the similarity network of the sites in the GS dataset as inferred applying the approach we have just exposed, setting the threshold at a value such that each site has at least one link. In such network, the similarity between two connected nodes is higher than what we measured in two thirds of the corresponding randomized samples.

The size of the node is proportional to the total number of stamps found in the site. Comparing the pictures in Fig. 6 and Fig. 8, we can notice how all the largest nodes are no longer connected to each other. Beside, not all the small sites are in peripheral position. The nodes highlighted in darker green have 6 and 4 stamps, respectively, and are of fundamental importance for the system since they are preventing it from becoming disconnected. This is a further proof that assessing the possible role of chance does indeed allow us to correct a negative bias toward small assemblages, a bias that is implicit in the BR definition when measuring the similarity of samples with different size.

Categories with heterogeneous size: looking for locally overrepresented attributes

Beside samples size effects , there are other issues whose impact deserves to be considered. Among them, the most relevant one is probably the heterogeneity in the size of the categorial attributes, that is, the variability in the number of objects or artifacts belonging to different categories. Sharing the same high percentage of a rare attribute is much more relevant than sharing the same fraction of a very frequent one. Generally speaking, it is straightforward that the presence of a certain category in a given sample is particularly significant when such category represents a larger fraction of that sample than it represents in the whole system. In other words, it is not only the value of the BR coefficient and its likelihood, but also the way such value is achieved. Performing random permutations is the appropriate way to address any kind of size heterogeneity, being it inherent to the samples or the categories. Therefore, if we want to address the latter, we can adopt the same strategy we used for addressing the former, but we need a different comparison. It is not longer the BR value for a given pair of sites that has to be larger than in the randomized samples. Now, it is the number of objects belonging to a certain category that has to exceed the amount we can find in a sample of the same size just as an effect of the action of chance. The categories will be thus labeled as "significant" or "not significant" in each one of the sites, depending on their size, the size of the assemblage in that site, and the presence of objects of that category in it. In Fig. 9A we represent two categories that are not significant in the largest site, i.e., the site of Mainz: *PNN*, with its 19 objects, 7 of which have been found in Mainz, and *AGRICOLAE*, who has a single representative in Mainz, out of a total of 4. In Fig. 9B, we show instead stamp-types that are significant in the same site: *CANTONIQVIETI*, with 13 object out of 20, and *SER*, with 3 out of 3.

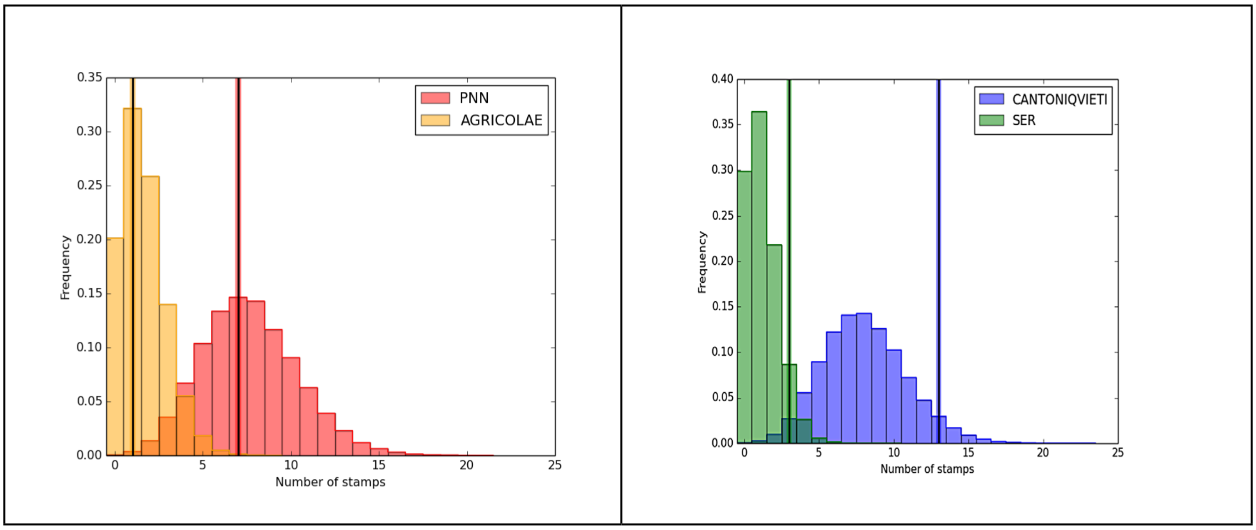

Figure 9. Frequency histogram of four stamp-types in randomized versions of Mainz' assemblage (362 stamps). In panel A (on the left), the stamp-types are PNN(19 stamps, 7 in Mainz) and AGRICOLAE (4 stamps, 2 in Mainz). In both cases, the probability of observing more in a randomized sample is high. In panel B (on the right), the stamp-types are CANTONIQVIETI (20 stamps, 13 in Mainz) and SER (3 out of 3 stamps in Mainz). The probability of observing more in a random sample of the same size is very small.

In Fig. 10 there is the comparison between two sites, Mainz (size=362) and Butzbach (size=6), where the stamp-type *QIMF* is present with 2 representatives out of a total of 6 stamps. Despite having one third of the total amount of pieces in each of the considered site, given the huge size of the assemblage in Mainz, two objects are not enough to make the presence of *QIMF* significant. This is not the case for the small site of Butzbach, where this category represent one third of the stamps in its assemblage .

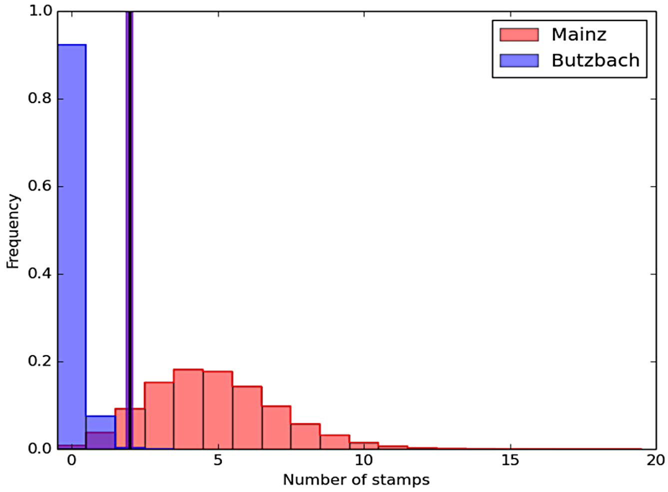

Figure 10. Frequency histogram comparing between two sites, Mainz (size=362) and Butzbach (size=6), where the stamp-type QIMF is present with 2 representatives (vertical line) out of a total of 6 stamps. The probability of finding more in a randomized sample is high for Mainz and low for Butzbach.

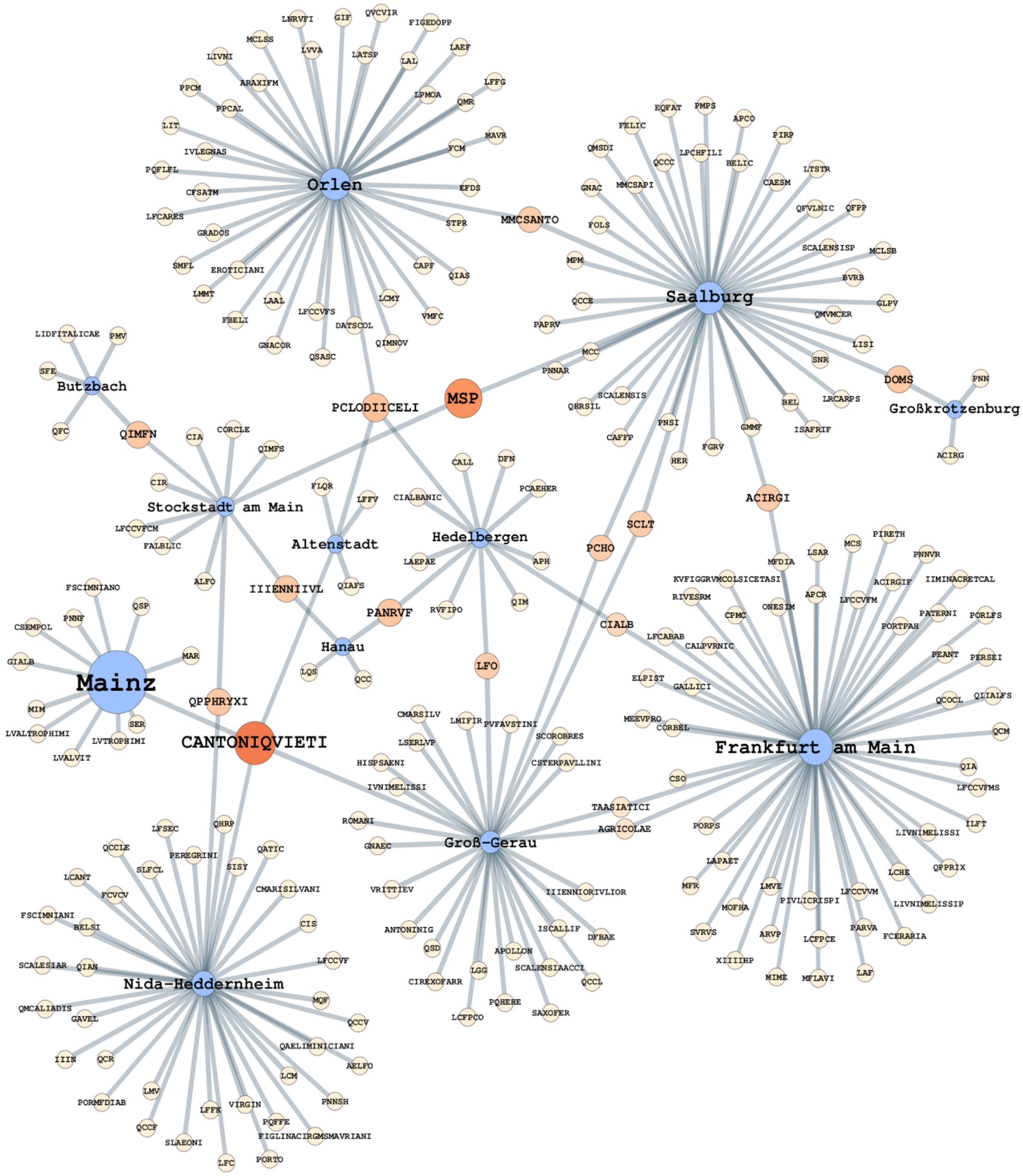

Figure 11. GS network of places (in blue) and stamps (from pale yellow to orange) with significant presence. The size of the nodes representing places are proportional to the size (number of stamps) of the corresponding assemblages. The size and the intensity of the colour of nodes representing stamp-types are proportional to their Betweenness Centrality [cit 34].

In this way, we obtain a filtered version of the network in Fig. 3, where the sites are linked only to the stamp-types whose presence in their assemblages can be regarded are significant (see Fig. 11).

Then, a new similarity measure has to be introduced in order to compare not the relative proportions of attributes, but binary lists of "significant" and "not significant" presences. We adopt the Jaccard Index (JI) [cit 35], a coefficient which measures similarity between finite sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets:

JI(i,j)={significant categorial attributes in i}{significant categorial attributes in j}{significant categorial attributes in i}{significant categorial attributes in j}

that is the ratio between the categories that are in both sites and the total number of those that are in one or another of the two places considered. This measure takes values between 0 and 1. It is equal to 1 when all the categories in the two lists are the same and 0 when there is no concurrence.

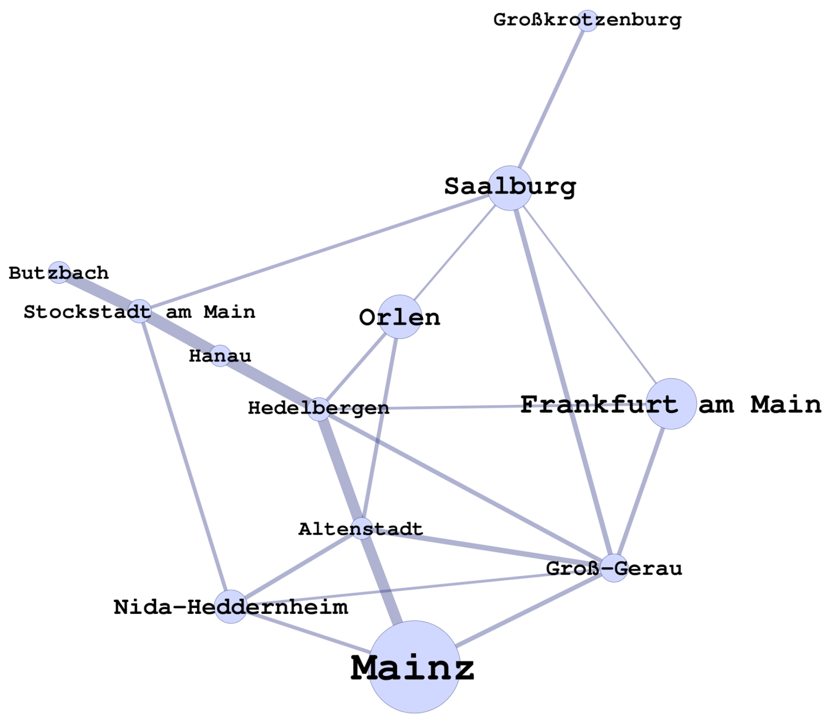

Figure 12. GS network of places constructed according to the JI. Node pairs connected by thicker links are those whose JI is larger. Disconnected pairs have null JI value. The size of the nodes are proportional to the size (number of stamps) of the corresponding assemblages.

Although this approach does not tackle size heterogeneity among assemblages directly, it does prevent large sites from acquiring a massive number of connections. This method is designed to unravel the hidden modular structure of the similarity network, if it has one. It connects nodes sharing attributes that can be regarded as typical from the corresponding sites, where typical means overrepresented compared to the random hypothesis. If there is no modular structure, if there are not groups of nodes sharing something that is not common in the rest of the system, or at least does not dominate the attributes distribution with the same force outside the boundaries of some cluster of sites, then this approach cannot be applied. Applying it to a system that has no pattern to unveil is pointless.

In the case of dataset DP, we have applied each of the three methods, thus reconstructing three different similarity networks. The first two networks resemble each other very closely and the second one can indeed be regarded as a validation of the first. We find three small clusters of outsiders. Two of them are characterized by local categorial attributes, that is, political parties that are not present in the rest of the system. The third group is characterized by minimal diversity, having only one category represented in it. The bulk of the network is composed by nodes (districts) where at least two of the most common categories (parties) can be found. The network obtained through the third method (Fig. 13) is remarkably different. Nevertheless, we can clearly interpret what we observe. This last method is able to ravel the fine structure hidden inside the bulk. Districts group together based on which parties, one or two at most, have better results compared to how they perform in the rest of the system. Clusters with two significant categories are connected to the clusters where only one of them is significant. Such intra-cluster links, however, are weaker than those connecting nodes inside the same group. The outsiders still form their own clusters, but those characterized by low diversity are now connected to other clusters where the same category dominates. By combining the two network representations, the one offered by the BR similarity measure and the other obtained by focusing on local specificities of single sites, we have the whole picture. The differences are not contradictions, they bring complementary information about the system under study.

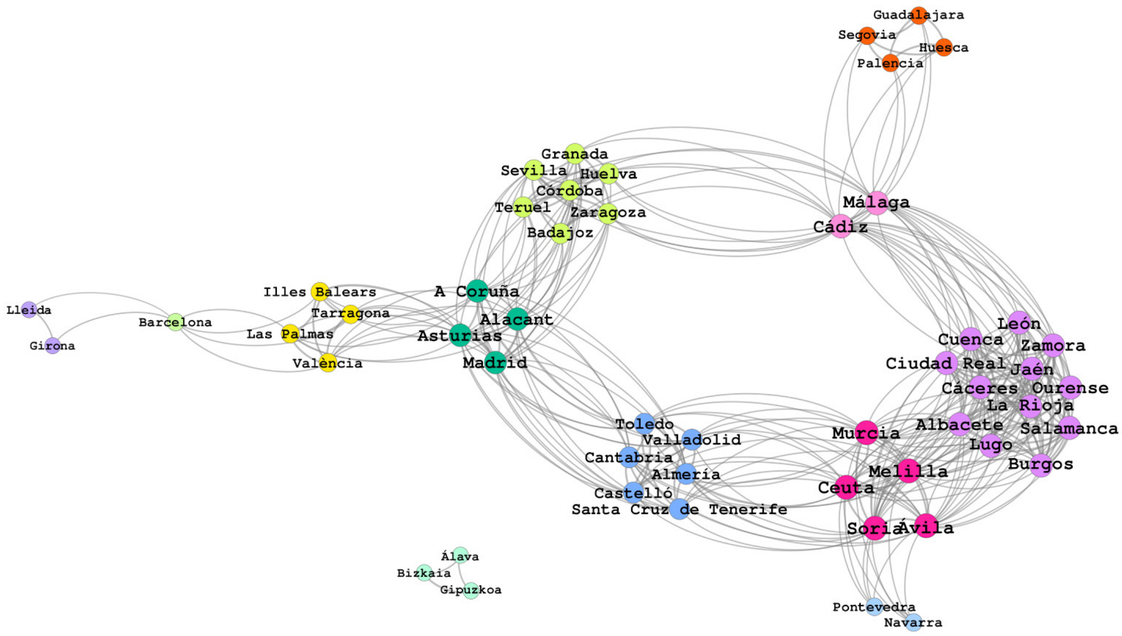

Figure 13. DP network of districts constructed according to the JI. Node pairs connected by thicker links are those whose JI is larger. Disconnected pairs have null JI value.

The case of GS is more complicate. There are no well-defined clusters and the correspondence between the two networks is not easy to read. However, all the strongest links in the network inferred accordingly to the third method are also present in the other network.

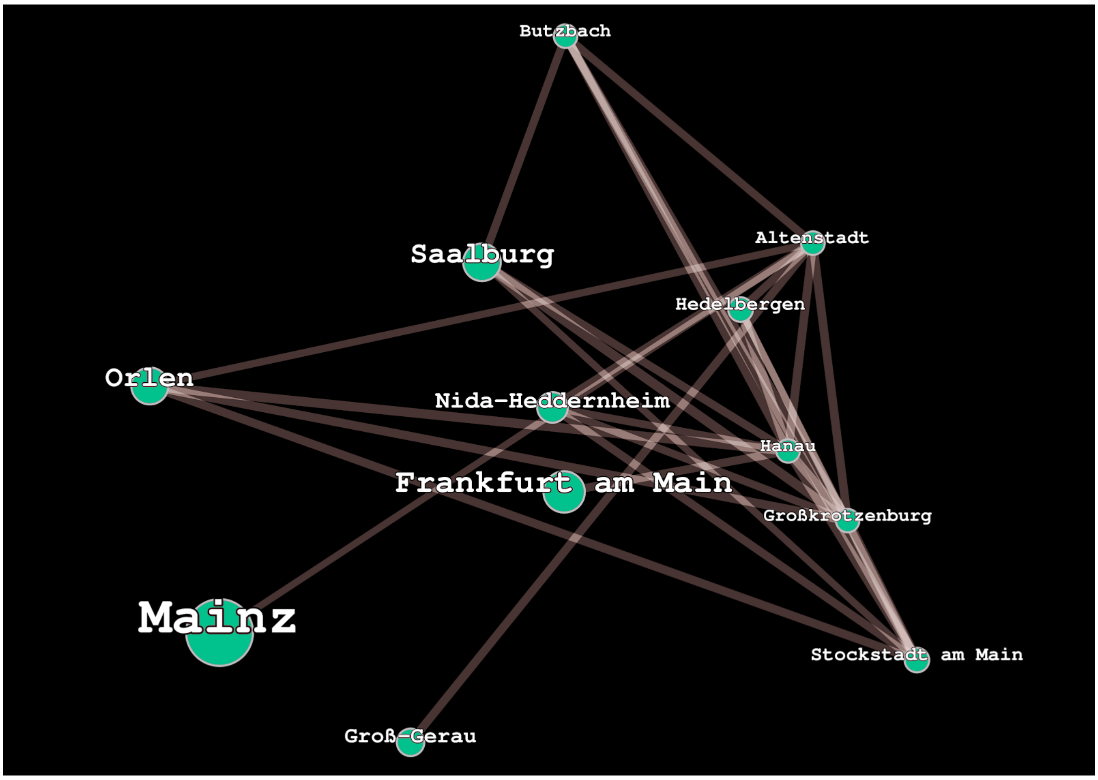

Figure 14. Geolocalized version of the network in Fig. 8.

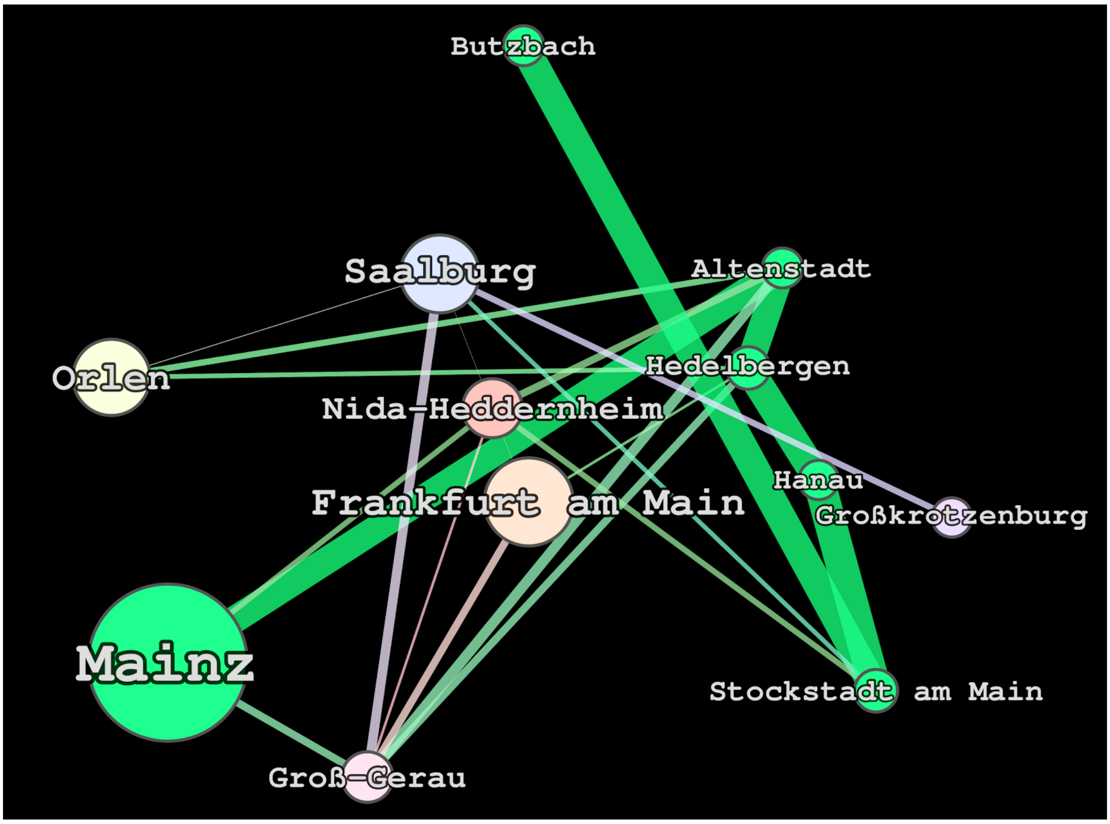

Figure 15. Geolocalized version of the network in Fig. 12.

Looking at Fig.14 and Fig. 15, we can clearly recognize a – rawly speaking – North-South axis that goes from Butzbach to Stockstadt am Main, including Altenstadt, Hedelbergen, Hanau, and Großkrotzenburg. On the other hand, the only certainty we have about East-West connections, are the links converging from Mainz and Orlen on the west side, to Altenstadt, in the east. Additionally, it is worth noting that, except for the nodes of the North-South axis, sites on the border of the province are almost never connected to each other, not even to the closest one. In particular, there is no evidence that the similarity between Mainz and Orlen is statistically significant, independently of the adopted approach. Actually, three of the four largest sites in the region – Mainz, Frankfurt am Main, and Orlen – are never connected to each other. The only exception is Saalburg, that has two weak links with Orlen and Frankfurt due to the presence of a significant amount of *MMCSANTO* and *ACIRGI* stamp-types, respectively.

Given the data available, this is as far as we can go. We can gain interesting insights, but the complete connectivity pattern is out of reach. However, this dataset demonstrate beyond any doubt that the absolute value of a similarity index is, *per se*, meaningless. The first similarity network we built has nothing to do with the other two. When we deal with data characterized by a high level of heterogeneity, both in the size of the assemblages and in the size of the categories, without knowing whether the survived samples are representative of the original proportion of attributes in the sites, statistical tests are an unavoidable part of network inference. It is the only way we have to separate information from illusions created by the action of chance, that is, actual signal from background noise due to heterogeneity, incompleteness and uncertainty.

Conclusions

Network science has great potential to offer to archaeology and history, but the foundation stone for any application of its tools and analytic techniques is a well grounded network construction. This means that we need a coherent definition for boundaries and nodes and we have to to infer the connectivity pattern in a reliable way.

Up to now, an established theoretical framework to deal with the specific combination of difficulties presented by archaeological data has not been completely developed. Network science toolbox does not include a general method to translate incomplete heterogeneous categorial data into probability of connection. Such methodology should be tested on a large number of datasets with different properties, including more recent datasets that can be artificially damaged.

Afterwards, we need to learn how to extract useful information from this probabilistic networks. Uncertainty at the link level does not, in principle, prevent from knowing which are the most important nodes, or how they group together, or whether the system is fragile or robust, easily navigable or likely to congest, centralized or distributed. Statistics may help integrating network science techniques. However, we should also be able to quantify how reliable our conclusions are.

This contribution represents a first attempt at outlining a long term research program.

A great effort needs to be devoted to this enterprise in next years. It is hard to say how long it will take, but we believe that interdisciplinary research teams, involving network scientists, archaeologists, and historians, gather the necessary expertise to accept the challenge.

References

[1] Freeman, L. C. (2011). The development of social network analysis-With an emphasis on recent events. *The SAGE handbook of social network analysis*, 26-39.

[2] Newman, M. E. (2003). The structure and function of complex networks. *SIAM review*, 45(2), 167-256.

[3] Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics reports,* 424(4), 175-308.

[4] Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1), 47.

[5] Strogatz, S. H. (2001). Exploring complex networks. *Nature*, 410(6825), 268-276.

[6] Costa, L. D. F., Rodrigues, F. A., Travieso, G., & Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in physics,* 56(1), 167-242.

[7] Montoya, J. M., Pimm, S. L., & Solé, R. V. (2006). Ecological networks and their fragility. *Nature*, 442(7100), 259-264.

[8] di Bernardo, D., Thompson, M. J., Gardner, T. S., Chobot, S. E., Eastwood, E. L., Wojtovich, A. P., ... & Collins, J. J. (2005). Chemogenomic profiling on a genome-wide scale using reverse-engineered gene networks. *Nature biotechnology*, 23(3), 377-383.

[9] Borge-Holthoefer, J., Moreno, Y., & Arenas, A. (2011). Modeling abnormal priming in Alzheimer's patients with a free association network. *PloS one,* 6(8), e22651.

[10] Gallos, L. K., Liljeros, F., Argyrakis, P., Bunde, A., & Havlin, S. (2007). Improving immunization strategies. *Physical Review* E, 75(4), 045104.

[11] Barthélemy, M. (2011). Spatial networks. *Physics Reports*, 499(1), 1-101.

[12] Borgatti, S. P., & Halgin, D. S. (2011). Analyzing affiliation networks. *The SAGE handbook of social network analysis*, 417-433.

[13] Everitt, B. S., Stahl, D., Leese, M., & Landau, S. (2011). *Cluster analysis*.

[14] Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344). John Wiley & Sons.

[15] Fortunato, S. (2010). Community detection in graphs. *Physics reports*, 486(3), 75-174.

[16] Padgett, J. F., & Ansell, C. K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American journal of sociology,* 1259-1319.

[17] Collar, A., Coward, F., Brughmans, T., & Mills, B. J. (2015). Networks in archaeology: phenomena, abstraction, representation. *Journal of Archaeological Method and Theory*, 22(1), 1-32.

[18] Graham, S. (2006). *Ex Figlinis: the network dynamics of the Tiber Valley brick industry in the Hinterland of Rome* (Vol. 1486). British Archaeological Reports.

[19] Brughmans, T. (2010). Connecting the dots: towards archaeological network analysis. *Oxford Journal of Archaeology*, 29(3), 277-303.

[20] Brughmans, T., & Poblome, J. (2012). Pots in space: understanding Roman pottery distribution from confronting exploratory and geographical network analyses.

[21] Terrell, J. E. (2010). Language and material culture on the Sepik Coast of Papua New Guinea: using social network analysis to simulate, graph, identify, and analyze social and cultural boundaries between communities. *Journal of Island & Coastal Archaeology*, 5(1), 3-32.

[22] Mizoguchi, K. (2009). Nodes and edges: A network approach to hierarchisation and state formation in Japan. *Journal of Anthropological Archaeology,* 28(1), 14-26.

[23] Sindbæk, S. M. (2015). Northern emporia and maritime networks. Modelling past communication using archaeological network analysis. *Harbours and Maritime Networks As Complex Aadptive Systems.*

[24] Knappett, C. (2013). *Network analysis in archaeology: New approaches to regional interaction*. Oxford University Press.

[25] Carayon, N. (2016). Harbours and Maritime Networks as Complex Adaptive Systems Johannes Preiser - Kapeller and Falko Daim (Eds) with 10 Contributors 139 pp., 44 colour and 12 b&w illustrations including maps and tables RGZM Monograph 23 from Römanisch-Germanisches Zentralmuseum, Ernst-Ludwig Platz 2, Mainz D-55116, 2015.I*nternational Journal of Nautical Archaeology,* 45(1), 213-214.

[26] Collar, A., Coward, F., Brughmans, T., & Mills, B. (2015). The connected past: critical and innovative approaches to networks in archaeology. *Journal of Archaeological Method and Theory Special*, (22), 1.

[27] Brughmans, T., Collar, A., & Coward, F. (Eds.). (2016). *The Connected Past: Challenges to Network Studies in Archaeology and History.* Oxford University Press.

[28] Martin, T., Ball, B., & Newman, M. E. J. (2016). Structural inference for uncertain networks. *Physical Review* E, 93(1), 012306.

[29] Östborn, P., & Gerding, H. (2014). Network analysis of archaeological data: a systematic approach. *Journal of Archaeological Science*, 46, 75-88.

[30] Liu, J. G., Hou, L., Pan, X., Guo, Q., & Zhou, T. (2016). Stability of similarity measurements for bipartite networks. *Scientific reports*, 6.

[31] Huntley, D. L. (2004). *Interaction, Boundaries and Identities: A Multiscalar Approach to the Organizational Scale of Pueblo IV Zuni Society* (Doctoral dissertation, Arizona State University).

[32] Huntley, D. L. (2008). *Ancestral Zuni glaze-decorated pottery: Viewing Pueblo IV regional organization through ceramic production and exchange* (No. 72). University of Arizona Press.

[33] Peeples, M. A. (2011). R Script for Calculating the Brainerd-Robinson Coefficient of Similarity and Assessing Sampling Error.

[34] Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

[35] Jaccard, P. (1901). *Etude comparative de la distribution florale dans une portion des Alpes et du Jura*. Impr. Corbaz.