Ancient regional transportation infrastructures as co-evolving complex systems: a multiplex network approach

Pablo Candelas
Master on Physics of Complex Systems and Biophysics
Universitat de Barcelona
(Dated: July 6, 2022)

Network theory has demonstrated to be an invaluable tool for studying complex systems, and it has been widely used in many disciplines. Archaeology is no exception. Ancient transportation infrastructures (TIs) are often represented as networks of interacting settlements. However, these various TIs are different in nature, and the network they form requires a different approach. Here we show that ancient TIs as a whole can be characterized by a multiplex. After an analysis of the dynamic performance and structural properties of the individual road and river networks of Southern Etruria and Latium Vetus in the Iron Age (950 − 580 BC), we overlap both TIs by regarding them as layers in a multiplex. We then study the multiplex with the aim to assess how good the interplay between layers is. Our work thus serves as a case study of an empirical multiplex network. We prove that the tools we use are good to characterize ancient TIs as one multiplex and could become very powerful if further studied and refined.

I. INTRODUCTION

It is common nowadays that scholars from different disciplines, regardless of the specificities of their research domains, find in network science a valuable ally when tackling complexity. This is the case for archaeology, too, though a special case: due to the incompleteness of the data, the large time spans and the complexity of human interactions, network science in archaeology is a tool that needs to be handled carefully [1]. Nonetheless, it has proven to be effective if used correctly [2], [3]. But the more uncommon tools that network science provides are still not broadly used in archaeological problems [4]. Thus, in this work, following the work of [2] and [3], we explore the transportation infrastructures (TIs) of Southern Etruria and Latium Vetus in terms of a multiplex, a multi-layer network in which inter-layer connections exist only between nodes that represent the same entity, settlements in our case. However, before tackling the multiplex, a better understanding of the individual layers and the metrics we apply is needed.

First, we will introduce the global efficiency and the algebraic connectivity, which we then use to study the dynamic performance and structural properties of both road (see Figure 1) and river transportation networks. With a good understanding of the properties of the individual layers, we then tackle the multiplex. Since we do not know for certain the properties of the two transportation infrastructures, we need two parameters to account for their different nature. Depending on the metric we are dealing with, the parameters will come with a different interpretation. After this interpretation is clear, we apply the results of [5] and [6] on the algebraic connectivity of a multiplex to our case study, and are able to get an idea on the structural properties of the multiplex and the interplay between layers. As for global efficiency, since it is not well defined for a multiplex, we provide an alternate, provisional definition that, although imperfect, allows us to extract information from the network.

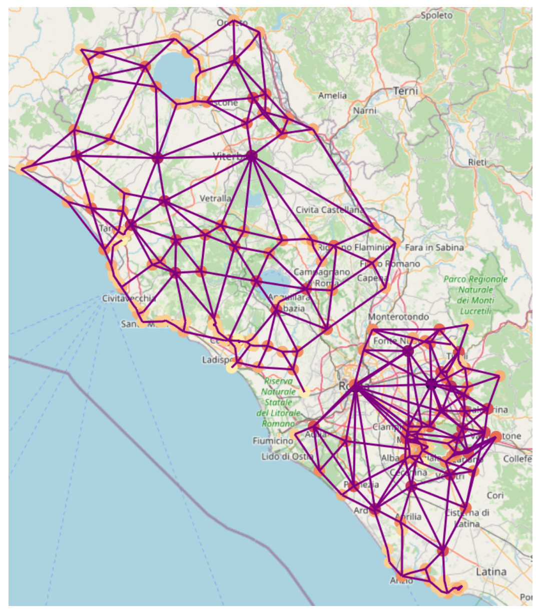

FIG. 1: Road networks of Southern Etruria (upper region) and Latium Vetus (lower region), EIA1E.

II. ON THE DATA

Road networks were not static. They continuously evolved through time, both because roads were built or neglected, because new settlements were formed or abandoned or because of many other eventualities. River networks posed a similar problem: rivers did not change in time so much —although they could, be it by drying up or by becoming no longer navigable—, but settlements did, and thus the network. However, precise information on this process is incredibly difficult to obtain. In [2], five large periods are determined in which a set of settlements coexisted without any major changes:

1. Early Iron Age 1 Early (EIA1E): Latial Period IIA

(950/925–900 BC)

2. Early Iron Age 1 Late (EIA1L): Latial Period IIB (900–850/825 BC)

3. Early Iron Age 2 (EIA2): Latial Period III (850/825–730/720 BC)

4. Orientalizing Age (OA): Latial Period IVA and IVB (730/720–580 BC)

5. Archaic Age (AA): 580–509 BC.

Within each one of these periods, the set of settlements can be regarded as unchanging, that is, all the centers existed during most of the corresponding time interval and did not suffer any major change. It is also assumed that the same applies to the routes between them. In this way, the analysis of the continuous evolution of settlements and routes is reduced to the study of five static time stamps [2]. For the road networks of Latium Vetus we used the data utilized in [3]. For the road networks of Southern Etruria, we started using the data of [2], but realized it was not correct while working on it. The coordinates of the settlements were off, not in a random manner, but in a systematic way, displaced by some factor in the conversion from GIS to longitude and latitude. We trained a polynomial regression algorithm with a set of settlements for which we had the correct coordinates and applied it then to the rest of the settlements to correct their coordinates. In Sec. IV we will quantify the effect of this correction and assess its relevance. As for the river networks, the data has not been yet published and all the results provided here are new. We were also provided with the links between regions, so, in addition to the networks of the individual regions, we are able to create a single network of the combined regions.

The process for translating roads and river maps into networks is stated in [2]. The set of settlements is represented as geo-localized nodes. As for the links, for modelling purposes it is not practical to assigning a weight only to the links that are present in the empirical networks. A mathematical function able to assign a cost to any potential links, i.e, to any connection between any pair of settlements, is needed. For simplicity, it was decided in [2] to assign weights to the links according to the geodesic distance between the nodes they connect.

It is important to note that the road networks are as accurate as archaeological data allow. A node is present when a settlement existed and was connected to other settlements through roads. This is not the case, however, for the river network. Here a node is taken when a settlement is close to a river, even if there is no strict evidence that it used the river for transportation purposes. In other words, the empirical river network is an upper bound to the real network.

III. ON THE METRICS

After creating the networks —one for every age of the road and river networks of both regions and, since we were provided with the links between regions, also for the set of settlements of the combined regions, so thirty in total—, we need to be able to extract information out of them. There are two quantities that we will use to characterize the networks, and each of them will come with a different interpretation of the network distances and parameters:

1. The global efficiency measures how close the nodes are to each other in the network L_{ij}, i.e., the shortest path length, compared to their geographical distance —in the case of geographical networks— d_{ij}:

$$E_{glob} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{d_{ij}}{L_{ij}} \tag{1}$$

   The more distant nodes are in the network, the smaller the quotient in the sum and the less efficient their communication will be. When the graph is complete, i.e., every node is connected to all other nodes, the global efficiency will be equal to unity. The global efficiency was used among other quantities in [2] and [3] with these very same networks to determine how good some models were in reproducing the empirical networks, by comparing the modeled networks with their empirical counterparts. However, it does not provide complete information on its own [7]. Furthermore, even if it is well defined for a single layered network, that is not the case for a multiplex.

2. The algebraic connectivity is the second-smallest eigenvalue of the Laplacian of a connected graph, which we can construct from the adjacency matrix $A_{\mathcal{G}}$ —a matrix with elements $a_{ij} = 1$ if there is a link between i and j and $a_{ij} = 0$ otherwise— of the graph as $L_{\mathcal{G}} = D - A_{\mathcal{G}}$, with D the diagonal matrix of the node degrees, i.e., the number of links of the node [8]. If the network is weighted, like in our case, $A_{\mathcal{G}}$ has the weight of the links w_{ij}, and D is the diagonal matrix of the node strength, the sum of the weights of the links connected to the node. The algebraic connectivity is a lower bound for both the edge connectivity and node connectivity of a graph, i.e. the minimal number of edges and nodes that have to be removed to disconnect the graph. It also sets the time scale for diffusion processes in the network, as well as the time needed to synchronize a network of oscillators. Thus, in this sense, the algebraic connectivity represents the connection between the structural and the dynamical robustness of a network [6]. In this work, we focus on the structural nature of the algebraic connectivity.

Because the global efficiency is not well defined in the multiplex formalism, we initially planned on using only the algebraic connectivity, but the new definition of global efficiency for a two-layered multiplex proposed in Sec. V encouraged us to use it for the entire work. This way, even if the aim of the work is completely different, the metrics are not that far from the ones used in [2] and [3], where they used, along with other quantities, the global and local efficiencies in parallel. The local efficiency is also in a way a measure of the structural resilience of the network. However, like the global efficiency, it is not defined for a multiplex, so using the algebraic connectivity was the obvious choice.

On the other hand, the global efficiency helps determine the network dynamic performance. It accounts for how fast the network can get information from point to point. If distances in the network are close to the empirical distances, the effectiveness of communications will approach the optimal value one. Because of this, when calculating the global efficiency we weight the links with the distances between nodes. We can view the distance as approximately the cost of travelling from one settlement to another. Then, the calculation of the global efficiency is straightforward.

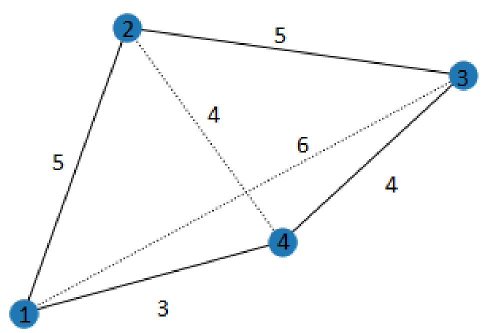

FIG. 2: Simple example network. The weight of the links have been assigned arbitrarily for illustrative purposes. Discontinuous lines account for potential links that do not exist.

Let us consider a toy graph like the one in Figure 2. There are six possible links. Four of them are present and will add one to the sum. The other two are not. Thus, they will add a fraction of one to the sum, according to Eq. 1. Because each link is considered twice in the sum, we have that:

$$E_{glob} = \frac{2}{12}\left(4 + \frac{6}{7} + \frac{4}{8}\right) = 0.89 \tag{2}$$

The algebraic connectivity, however, requires a different approach. In the formalism of networks, the weight of the network is proportional to the strength of the connection. The algebraic connectivity, which is a measure of the resilience of the network, of how hard it is to separate it into two similar components, benefits from strong connections. That is, a link with a large weight will be harder to remove, or equivalently, easier to add. It does not make sense, though, that longer roads are easier to build than shorter ones, but this is what would happen if we weighted the networks with distance. Thus, when dealing with algebraic connectivity we take the weight of the links as the inverse of the distance. This way the strongest links are the shortest ones, which are the easier to place and to maintain. In this sense, distances in the network can be regarded as maintenance costs: a link between far settlements will be more prone to failure, thus more expensive to maintain. Then, the Laplacian matrix of Figure 2 becomes

$$L_{\mathcal{G}} = \begin{pmatrix} 8/15 & -1/5 & 0 & -1/3 \\ -1/5 & 2/5 & -1/5 & 0 \\ 0 & -1/5 & 9/20 & -1/4 \\ -1/3 & 0 & -1/4 & 7/12 \end{pmatrix} \tag{3}$$

The resulting eigenvalues from the eigenvalue equation $L_{\mathcal{G}}\mathbf{v} = \lambda\mathbf{v}$ are $\lambda_1 = 0$, $\lambda_2 = 1/2$, $\lambda_3 = \frac{-\sqrt{19}+11}{15}$ and $\lambda_4 = \frac{\sqrt{19}+11}{15}$. The algebraic connectivity of the graph in Figure 2 is the second smallest eigenvalue. i.e. λ_2. The smallest eigenvalue of the Laplacian matrix being zero is a consequence of its definition: every row sum and column sum of $L_{\mathcal{G}}$ is zero, so the vector $\mathbf{v}_0 = (1, 1, \ldots, 1)$ always satisfies $L_{\mathcal{G}}\mathbf{v}_0 = 0$. In fact, there will be a zero eigenvalue of the Laplacian matrix for every separated component in the network, because their Laplacian matrix is ordered in blocks. Then, if the network is disconnected, i.e., it does not form one giant component, the algebraic connectivity will be zero. We will encounter such cases in our networks. Fortunately, most of the cases it will be a matter of isolated settlements or small clusters, which we will disregard in favor of the giant component. Furthermore, when implementing the multiplex this issue solves itself, because the new layer usually connects isolated clusters.

The global efficiency and the algebraic connectivity, although overlapping in some aspects, make us take on very different interpretations of the networks. But, in the end, the different interpretations do not forbid us from using them simultaneously. They provide complementary information, and by using both of them we are able to get additional information by comparing the results and not jump to conclusions.

IV. SINGLE LAYER APPROACH

During ancient times, roads were the primary way of transportation over short regional distances. A well connected road network allowed travellers, merchants or armies to swiftly get from one place to another. It also tied settlements together, improving communications and the spreading of news. This increase in social interaction could then result in the emergence of more complex behaviours such as collaboration. Indeed, a new road was not only beneficial to the settlements on its

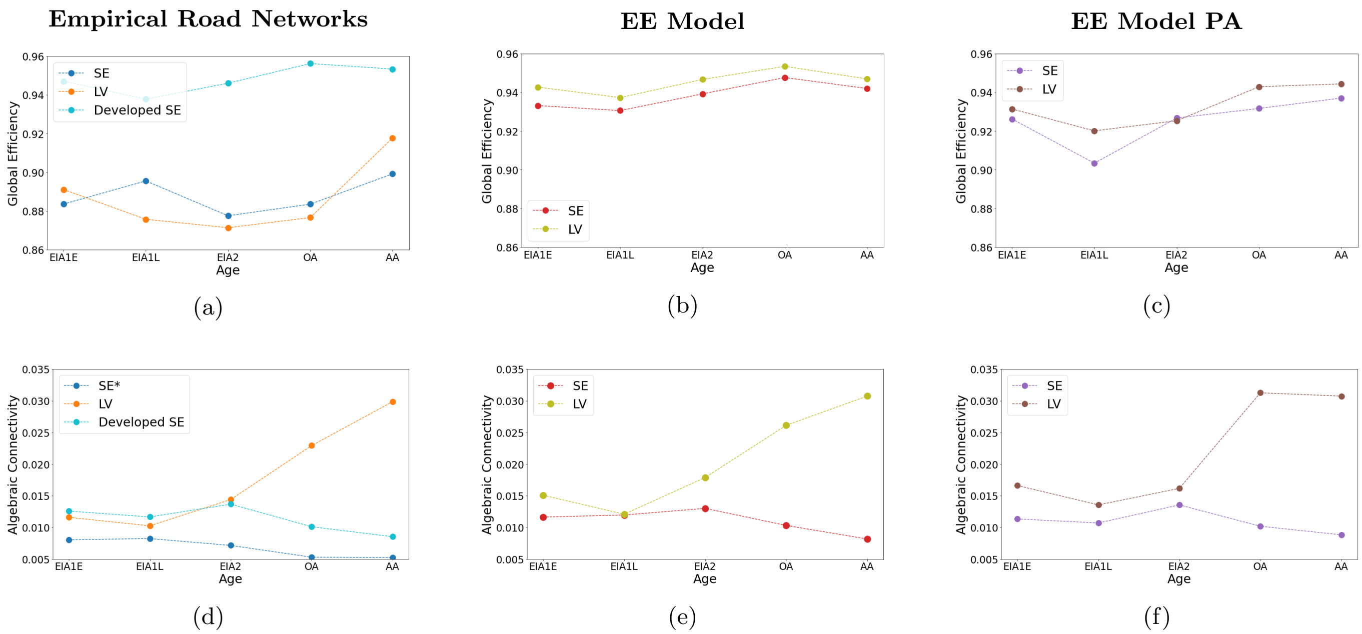

FIG. 3: Road network properties of Southern Etruria and Latium vetus. The first row shows global efficiency and the second row shows algebraic connectivity. Figures (a) and (d) show the empirical networks, along with a modified version of the empirical network of SE, developed with the EE Model until the distance ratio coincides with that of LV as shown in table I. Figures (b) and (e) show both regions developed with EE Model from scratch until the distance ratio reaches that in table I. Same for figures (c) and (f), but developed with the EE Model modified with preferential attachment. The preferential attachment parameter used is the same as in [3]. The * on (d) accounts for taking only the giant component of SE, not the complete network.

ends, but enhanced the overall network, proving to be useful for every settlement on the network. Since it is common practice to assume that roads were the output of a collective effort, it is straightforward to notice the circular property of cooperation in road building: a higher cooperation yields a better output of the resources in the form of more roads, which allows the settlements to cooperate further. A highly connected region enjoyed more efficient exchanges, which means maintaining a good road network was a global —read, the whole network— priority.

By calculating the algebraic connectivity and the global efficiency we can measure the effectiveness of the road networks and their evolution in both regions through the ages, as well as compare them to see which region had the better connected network and the most resilient one. Although the proposed cornerstone of this work was the multiplex approach, I decided to study the single layer case in more detail. The algebraic connectivity had not been yet calculated for the individual networks, so I thought a deeper single layer analysis of both this quantity and the global efficiency of the networks would be helpful when tackling the multiplex. Some of the results in this section, like Figures 3a, 3b and 3c already are presented in [2] and [3], but the rest are still unpublished.

A. Empirical Networks

Let us begin with Table I. The road networks of Latium and Etruria are different through the ages. In the later ages, for example, more roads were built and the total length of the networks L_{tot}, the sum of the length of all edges, is larger. But just referring to the total length can be misdirecting. Since Southern Etruria is bigger than Latium Vetus, we would not be able to compare them if we used L_{tot}. Instead, we use the relative spanning ratio L_{rsr}, which is just L_{tot}/L_{mst}, L_{mst} being the distance of the minimum spanning tree of the network. If N is the number of nodes in a network, the minimum spanning tree of the network is formed with the $N-1$ edges with the lowest possible weight that connect all nodes. This way, we are able to compare how developed the networks were for the different ages and regions. The road networks of LV and SE are similar enough in form so that the decision of using the L_{rsr} is justified. In fact, [2] and [3] demonstrate that the the minimum spanning tree was mostly present in the empirical road networks.

Let us look now into Figure 3. Southern Etruria does not undergo major changes to its algebraic connectivity during the first three periods, but goes to zero in the later Orientalizing Age (OA) and Archaic Age (AA) because the network becomes disconnected. Even if we remove

Region	EIA1E	EIA1L	EIA2	OA	AA
SE	3.17	3.35	3.17	3.19	3.61
LV	4.67	4.51	5.02	5.62	5.26

TABLE I: Relative spanning ratio L_{rsr} of the road networks.

	EIA1E	EIA1L	EIA2	OA	AA
a	0.09	0.11	0.10	0.08	0.06

TABLE II: Preferential attachment exponent of modified EE model.

the isolated components, the algebraic connectivity of the giant component suffers a large drop, as shown in Figure 3d. This means that its not a matter of part of the region becoming isolated, but of a larger issue in the structure of the regional network. It is an interesting result, considering the increase in road density of the Etruscan network in the AA (Table I), which does not seem to increase the algebraic connectivity of the region. It does increase, however, the global efficiency (see Figure 3a). This could mean that the region is optimizing the efficiency but disregarding possible failures, and it is an example of why it is important to use both metrics.

On the contrary, the algebraic connectivity of Latium Vetus increases considerably through the ages, as a result of a more robust, better connected network (Figure 3d). This behaviour is mirrored by the global efficiency in the later ages. In the earlier ages, however, the global efficiency gives us different insight than the algebraic connectivity, for, according to Figure 3a, SE was more efficiently communicated.

The road network of Latium Vetus was actually more developed: the relative spanning ratio L_{rsr} of the network is higher throughout the ages (see Table I). This is interpreted in [9] as Latium having access to a larger pool of resources that allowed the region to develop its infrastructure further, so the larger algebraic connectivity is expected. Because LV wasted the excess resources in a rich-get-richer bias, i.e., the settlements with greater node strength were favoured when building roads, the global efficiency is not as large. The fact that SE was able to keep up in efficiency with a lower L_{rsr} shows a better management of the available resources. We would expect a very different result if SE had access to a similar amount of resources. In that case, the road network could have been developed up to the same L_{rsr} as Latium in Table I. We will provide Etruria with this surplus of resources in order to compare the performance with LV. But for that, we need to build roads.

Roads were expensive. Building them required a huge amount of resources and manpower, which would be unlikely for one settlement to take on alone unless it had a lot to gain. Instead, there is evidence that for Etruria the road network was a result of a collective regional effort [2]. This gave insight on the type of behaviour that could have directed the road building in southern Etruria during this time, labelled in [2] as the Equitable Efficiency (EE) model. In this model, we calculate at each step the routing factor $R_i(j) = \frac{d_{ij}}{L_{ij}}$ of every pair of nodes. We then add the link with the minimum value of all the R values. In this way, the whole network cooperates in the decision of where to allocate the resources.

Since [2] suggests this model best reproduces the Etruscan empirical network, it is reasonable to assume that, had Etruria more resources available, the region would have been further developed following the same process. Thus, we use the EE model to develop the Etruscan network until the relative spanning ratio L_{rsr} is the same of that of Latium Vetus (Table I). From this results it would seem that the Etruscan network is better designed: with the same amount of resources the global efficiency of SE greatly surpasses that of LV, while the algebraic connectivity remains similar until later ages (see Figures 3a and 3d). The fact that Latium Vetus remains mostly with the larger algebraic connectivity even with significant difference in the global efficiency highlights the importance of using both quantities. Still, the later advantage of LV in algebraic connectivity cannot be explained with a difference in resources since we have already developed SE. The network of Latium must have some other source for this structural advantage.

B. EE Networks

In this approach, however, we have given Southern Etruria an advantage in terms of global efficiency, because the EE model favors efficiency in the creation of new edges. Meanwhile, Latium Vetus remained strictly empirical. We can remedy this by developing both networks from their minimum spanning tree following the same model. This also allows us to compare both networks with other networks developed using the same algorithm. In addition to the EE model, we also used the model proposed in [3], which did a better job at reproducing the empirical network of Latium Vetus from scratch, the EE PA. This model is a modification of the EE model with preferential attachment (PA), accomplished by weighting the normalized distance with a negative power of the weighted degree of the node $R_i(j) = \frac{d_{ij}}{L_{ij}} k_w(i)^{-a}$. The exponents used (Table II) are the same as in [3], which are the ones that best reproduced the empirical network of LV. Building the networks using this models gives a couple of remarkable results.

After developing the networks following these two models, both regions are better connected than their empirical counterpart. This was expected, especially for the global efficiency. As stated before, in [9] it is suggested that, although Latium had a resource advantage, some of this advantage would be wasted because of the pref-

	EIA1E	EIA1L	EIA2	OA	AA
Empirical E_{glob} diff.	0.008	0.007	0.007	0.01	0.01
EE E_{glob} diff.	0.033	0.022	0.047	0.048	0.027

TABLE III: Global efficiency difference between empirical and EE networks of SE with old coordinates and new corrected coordinates.

erential attachment in its road building method. In 3b we can see that indeed, if both regions have the same resources and Latium does not waste in PA, their global efficiency is more or less the same. Also, the algebraic connectivity of developed Etruria is larger in 3a than in 3b. One should consider that developed SE in 3a is developed from the empirical network. Thus, the EE model has had less influence on the network than in 3b. The difference is small, but it is still surprising, because the EE model adds at each step the link that constitutes the best improvement in the communication among any pair of nodes. However, the chance exists that different, non instantly optimal choices could lead to a better final result. There might be something else at work here, but a more detailed study would be necessary to reach any relevant results. Also, note that the structural advantage of LV in the later ages is maintained after developing the network with both models (see Figures 3e and 3f).

C. Coordinate Correction

It should be noted that the values of empirical (Fig. 3a) and EE modeled (Fig. 3b) global efficiency are different from the ones obtained in [2]. The difference in global efficiency in both the empirical and EE networks is quantified in Table III. This is due to a previous error in the coordinates of Etruria that has now been corrected, as noted in Sec. II. For the empirical network, the new values of global efficiency barely change after the correction. This is evident considering the paths are the same ones, only with a slight variation in distance. As for the EE networks, the correction is larger, about an order of magnitude larger in the earlier ages. Here the networks are developed with the EE model from the different coordinates, thus producing different networks. The larger discrepancy in global efficiency is expected, but it is still small. This means that the perturbed set of coordinates, although it builds a different network, still manages to obtain a similar global efficiency. In conclusion, this correction, although it was important to do, should not affect the results obtained in [2] or its conclusions.

D. Perturbative Analysis

To check whether or not the results were consistent we decided to make a raw perturbative analysis. We perturbed the coordinates —latitude and longitude— of the settlements with a Gaussian white noise with zero mean and 0.01 standard deviation. This amplitude was handpicked. The average displacement of one realization of the perturbation

$$\bar{\Delta} = \frac{1}{N}\sum_{i}^{N}\sqrt{(Long_i - Long'_i)^2 + (Lat_i - Lat'_i)^2} \quad (4)$$

where N is the number of settlements and the primed coordinates correspond to the perturbed coordinates, is small compared to most distances in the networks, but for the closest settlement clusters (see Figure 1) the perturbation will certainly distort the landscape. Nonetheless, it is just one value. A more complete perturbation protocol should be devised and implemented. For now it is just exploratory work. The perturbation could account, for example, for misinformation about the costs when building the road networks, or for particularities of the terrain that our models do not contemplate. In Figure 4, for both the empirical and the EE modeled networks, the difference in global efficiency and algebraic connectivity with their perturbed counterpart is plotted, normalized to the non-perturbed value. For each age and region:

$$\delta E_{glob} = \frac{\langle E_{glob} - E'_{glob}\rangle}{E_{glob}} \quad (5)$$

where E'_{glob} stands for the global efficiency of an individual realization of the perturbation. We did not perform the perturbation on the EE PA (Figures 3c and 3f) because we did not expect it to provide any new information.

The perturbation makes the values of algebraic connectivity and global efficiency drop. It is worth noticing that the perturbation affects similarly both quantities. In fact, although not shown in Figure 4, the stronger the perturbation, the larger the drop. Also, in general, the EE modeled networks seem to hold better against it. The difference for the empirical values is non-negligible. Meanwhile, for the global efficiency of the corrected coordinates (Table III), the correction, which can also be seen as a perturbation, caused much smaller differences in the case of empirical networks. The fact that these two different perturbations in the coordinates cause different alterations in the networks is interesting, though not surprising. The correction of the coordinates is a systematic displacement, while the noise is random and thus easily makes distances in the network larger.

To assess the robustness of the results of this section, in Table IV we write the p-values of the perturbation: the percentage of times the perturbation compensates for the difference in algebraic connectivity or global efficiency between regions. In other words, if after a perturbation of the coordinates the global efficiency of the EE modeled LV network is lower than the global efficiency of SE in Figure 3b or vice-versa, the p-value of the EE modeled global efficiency increases by a factor $1/n$, where n is the number of realizations of the perturbation.

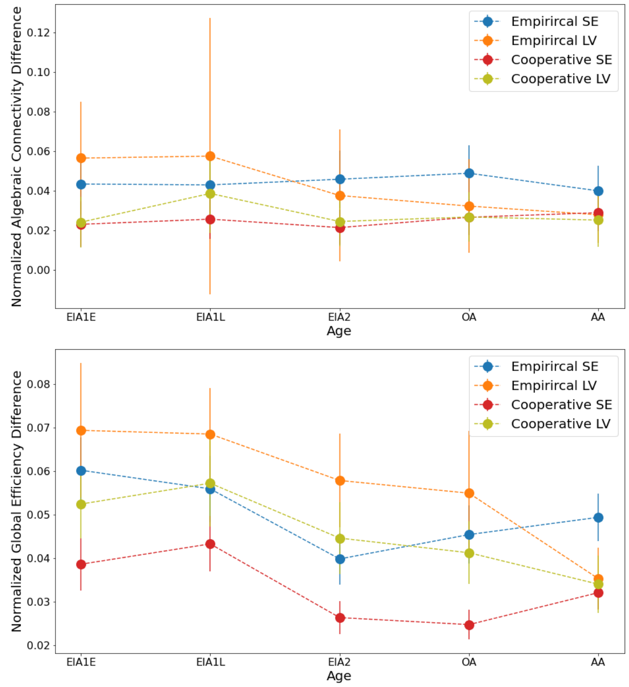

FIG. 4: Difference in algebraic connectivity and global efficiency between the empirical and perturbed networks of LV and SE (blue and orange) and between the empirical and perturbed EE networks (red and olive) with values normalized to the non-perturbed value of the network. These results are for 50 realizations of the perturbation.

	EIA1E	EIA1L	EIA2	OA	AA
Empirical E_{glob}	1	1	1	1	0.98
EE modeled E_{glob}	1	1	1	1	1
Empirical λ_2	0	0	0	0	0
EE modeled λ_2	0	0.9	0	0	0

TABLE IV: p-values after 50 realizations of the perturbations for Figures 3a, 3b, 3d and 3e.

The first two rows in Table IV tell us that the differences in global efficiency between LV and SE in 3a and 3b are not robust. When perturbing the EE modeled network, we are effectively worsening the EE model. This is realistic in the sense that the model assumes the settlements have perfect information on the cost of building roads to their neighbours. However, it is easy to see that this information would probably not have been so complete. Thus, we can say that both networks are able to attain effectively the same global efficiency after being developed with the EE model. This result, together with Figure 3a supports the argument in [9] that Latium wastes their resource excess with the preferential attachment bias. It also highlights the similarities of both regions, supporting our decision of using the L_{rsr}. As for the algebraic connectivity, the only time the perturbation compensates the difference is for the EE modeled EIA1L age, where the values almost overlap (see Figure 3e. Perhaps a larger perturbation could increase the p-values of the algebraic connectivity in EIA1E, EIA1L or EIA2, but it is clear that for OA and AA the difference is too large. The result that there is something else to the structure of Latium in these later ages holds.

E. River Networks and Combined Regions

Although our analysis of road networks yielded interesting results, it is rather incomplete: ancient transportation infrastructures were composed not only of roads, but of rivers as well. If we want to consider a more accurate picture of the ancient transportation networks, we need to account for other ways of transport.

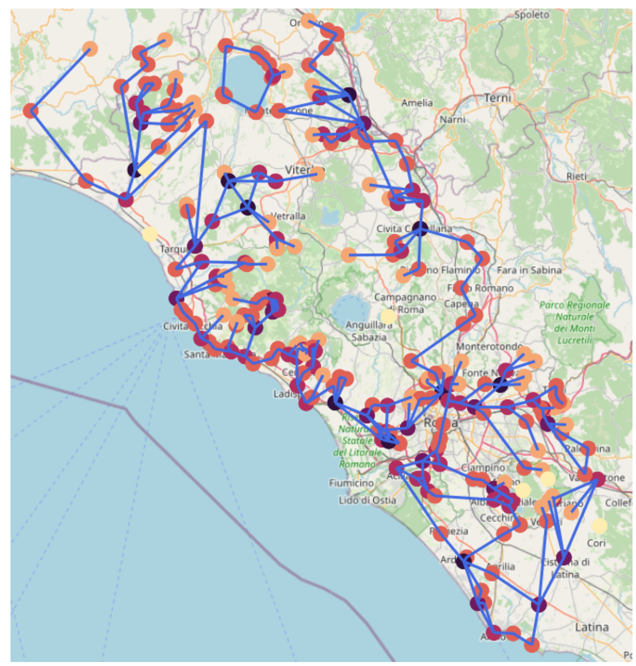

FIG. 5: River networks of Latium Vetus and Southern Etruria, Archaic Period.

Fluvial and maritime routes were, and still are, crucial ways of transportation. Unlike roads, rivers did not need to be built; they already were there. They flowed from high altitude points towards the sea, thus the tree-like structure in Fig. 5. We also consider short navigable distances between coastal settlements as links in the network. Since we were also provided with the data on the links between both regions, we thought it would be interesting to produce the results for the networks of the combined regions for both road and river networks (Table V).

When calculating the algebraic connectivity, we only keep the giant component. For most of the networks the isolated components are small. However, the rivers of SE in EIA2 are a special case. Here, the network is

Road	EIA1E	EIA1L	EIA2	OA	AA
LV E_{glob}	0.891	0.876	0.871	0.877	0.918
SE E_{glob}	0.884	0.896	0.878	0.863	0.879
Comb. Reg. E_{glob}	0.901	0.876	0.887	0.897	0.914
LV λ_2	$1.2*10^{-2}$	10^{-2}	$1.4*10^{-2}$	$2.3*10^{-2}$	$3*10^{-2}$
SE λ_2^*	$8.1*10^{-3}$	$8.3*10^{-3}$	$7.2*10^{-3}$	$5.3*10^{-3}$	$5.2*10^{-3}$
Comb. Reg. λ_2^*	$3.2*10^{-3}$	$3.5*10^{-3}$	$2.4*10^{-3}$	$2.6*10^{-3}$	$3*10^{-3}$
River	EIA1E	EIA1L	EIA2	OA	AA
LV E_{glob}	0.59	0.54	0.53	0.61	0.59
SE E_{glob}	0.46	0.59	0.62	0.49	0.44
Comb. Reg. E_{glob}	0.64	0.67	0.58	0.60	0.61
LV λ_2	$2.3*10^{-3}$	$1.9*10^{-3}$	10^{-3}	$2.1*10^{-3}$	$1.6*10^{-3}$
SE λ_2^*	$3*10^{-4}$	$5*10^{-4}$	10^{-3}	$3*10^{-4}$	$2*10^{-4}$
Comb. Reg. λ_2^*	$4*10^{-4}$	$5*10^{-4}$	$3*10^{-4}$	$3*10^{-4}$	$3*10^{-4}$

TABLE V: Global efficiency and algebraic connectivity for empirical road and river networks of Latium vetus, southern Etruria and the combined regions. The asterisk accounts for considering only the giant component.

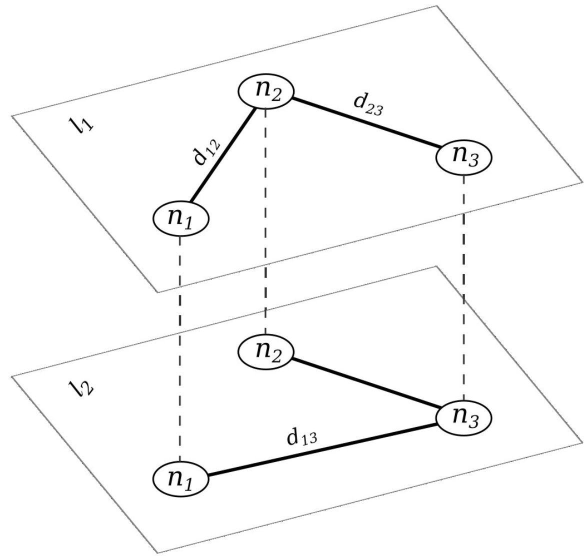

FIG. 6: Toy two layered multiplex.

separated into two large components. When calculating the algebraic connectivity, we keep the giant component, the largest of the two, and the consequence can be seen in the resulting λ_2 being an order of magnitude larger than for the rest of the ages.

In Table V it can be observed that both measured quantities are lower for river networks than for road networks because rivers were not built. One could argue that the settlements were built to maximize the benefit of rivers, but that hardly counters the effect of customizing the network at least partially, as with roads, to fit the region's needs. This difference in algebraic connectivity and global efficiency between road and rivers is important to keep in mind. As for the combined regions, it is not surprising that the algebraic connectivity is lower: the network can easily be cut into two components —the two regions—, due to the smaller number of links between regions. The global efficiency, on the other hand, is much more intriguing. Save a couple cases, the combined regions have a higher global efficiency than the isolated networks. This means that the paths between nodes of different regions are short, relative to their geodesic distance. A possible explanation for this could be Rome, which is a hub in the network of Latium and close to the border of the regions, immediately connecting many of Latium's nodes to Etruria. In any case, although informative, the single layer approach is incomplete. In order to get a complete picture of the transportation infrastructure we need a way to overlap the road and river layers into a single network.

V. THE MULTIPLEX APPROACH

A multiplex is a multilayered network in which the layers share at least part of their set of nodes. Nodes belonging to more than one layer are connected to all their counterparts in the other layers (see Fig. (6)). Multiplex networks are characterized by a supra-adjacency matrix $A_{\mathcal{M}}$, which has the adjacency matrix of the individual layers as diagonal blocks [8]. Additionally, the inter-layer connections between nodes present in multiple layers are represented by the inter-layer coupling or weight [10].

Now, as we can see in Table V, the river network has very poor properties compared to the road network. However, a multiplex formed by these two layers can have better properties than the single road network, provided the rivers either create alternative paths or create shorter paths. These two possibilities account for resilience and efficiency and are directly connected with the metrics we use: the algebraic connectivity and the global efficiency.

A. Algebraic Connectivity

Just like with single layered graphs, from the supra-adjacency matrix we can construct the supra-Laplacian matrix $L_{\mathcal{G}_{\mathcal{M}}}$ with the inter-layer weights p outside the diagonal blocks [8]. The supra-Laplacian of the multiplex in Figure 6 is

$$L_{\mathcal{G}} = \begin{pmatrix} d_{12}+p & -d_{12} & 0 & -p & 0 & 0 \\ -d_{12} & d_{12}+d_{23}+p & -d_{23} & 0 & -p & 0 \\ 0 & -d_{23} & d_{23}+p & 0 & 0 & -p \\ -p & 0 & 0 & d_{13}+p & 0 & -d_{13} \\ 0 & -p & 0 & 0 & d_{23}+p & -d_{23} \\ 0 & 0 & -p & -d_{13} & -d_{23} & d_{13}+d_{23}+p \end{pmatrix} \quad (6)$$

The algebraic connectivity of the multiplex will be the second smallest eigenvalue of the supra-Laplacian, provided the multiplex is one connected component, which will be the case of our networks.

Let us consider any two-layer multiplex. If the strength of the connection depends on distance equally on both layers, then, depending on the inter-layer coupling p, the multiplex can exist in various states [5]. In the disconnected phase, the strength of the connections between layers p is small ($p < p^*$) and the algebraic connectivity of the multiplex $\tilde{\lambda}_2$ will increase linearly as $2p$, i.e., the sparsest cut is to separate the multiplex into the two layers. If $p^* < p < p^\diamond$, $\tilde{\lambda}_2$ will be given approximately by $p + \lambda_2^{min}$, where λ_2^{min} is the algebraic connectivity of the Laplacian dominant layer [6], the layer with lowest algebraic connectivity. The easiest way to disconnect the multiplex will no longer be to separate the layers, but to cut the the Laplacian dominant layer (λ_2^{min}) and the connections between one of those halves and the other layer (p). Ultimately, if $p > p^\diamond$, the algebraic connectivity is approximated by that of the aggregated network of both layers. The aggregated network is a coarse-graining of the multiplex in which the nodes are represented in one layer, with connections present in more than one layer represented as multiple connections between the same pair of nodes and the inter-layer couplings as self loops. The transition point p^* will be approximately given by

$$\begin{aligned} 2p^* &= p^* + \lambda_2^{min} \\ p^* &= \lambda_2^{min} \end{aligned} \quad (7)$$

For simplicity, let us consider the regimes with $p > p^*$ as one, with the algebraic connectivity bounded from above by that of the aggregated network and from below approximately by $2p^* = 2\lambda_2^{min}$. So, either $p < p^*$ and the the links between layers are incredibly vulnerable, or $p > p^*$ and the multiplex is at best as vulnerable as the aggregated network. Under these circumstances, adding a lower algebraic connectivity layer to a higher algebraic connectivity layer is most likely going to be a hindrance —not necessarily, if p is large enough and the second layer is not that bad in terms of algebraic connectivity—.

In our case, rivers are the Laplacian dominant layer ($\lambda_2^{riv} < \lambda_2^{road}$). We have seen in Table V that the river network is worse than the road network. However, there is archaeological evidence that rivers were part of the transportation infrastructure. For that, though, the river layer needs to compensate for its low algebraic connectivity, so we thought that the strength of a river connection, if equal in distance to a road, should be higher. The parameter α accounts for this by multiplying weights in the river layer. This way, rivers are more resilient. We will effectively enhance the river layer from being equal in nature to the road layer ($\alpha = 1$) to being better ($\alpha > 1$). The algebraic connectivity of the rivers will be $\alpha\lambda_2^{riv}$, where λ_2^{riv} is the original algebraic connectivity of the river network. Then, $\lambda_2^{min}(\alpha) = min(\alpha\lambda_2^{riv}, \lambda_2^{road})$. There will be a transition point given by

$$\bar{\alpha} = \frac{\lambda_2^{road}}{\lambda_2^{river}} \quad (8)$$

for which the river network can be considered more resilient than the road network. Then we are no longer adding a "worse" layer, and the algebraic connectivity of the multiplex will increase respect to that of the road network unless, in the disconnected phase,

$$\begin{aligned} \tilde{\lambda}_2 = 2p &< \lambda_2^{min} = \lambda_2^{road} \\ p &< \frac{\lambda_2^{road}}{2} \end{aligned} \quad (9)$$

The algebraic connectivity responds to the question of how much is worth investing in order to prevent possible failures in the system. Much like in the single layer approach, the calculation of the algebraic connectivity requires a weight that is inversely proportional to the distance. In this sense, it is intuitive to associate the distance of the links to their maintenance; if we are able to make a well connected and resilient network with abundant, shorter paths —and thus easier to maintain, or rather, more difficult to remove—, then the algebraic connectivity will be large. The parameter that allows rivers to enhance the network by decreasing their maintenance costs will be regarded as the *river convenience factor* α, which divides distances and thus multiplies weights in the river block of the supra-adjacency matrix. A high river convenience factor will increase the weights of the river connections, thus making the incorporation of rivers to the multiplex increase the algebraic connectivity further. Or, in other words, a high convenience factor means that it will be hard to perturb the multiplex by removing —as in the case of a failure— river connections. Meanwhile, the inter-layer coupling p will also be the inverse of a distance, which we interpret as the cost $1/p$ of maintaining a port on the settlement, the *port cost*. A unit of cost is equivalent to one kilometer of road maintenance. Maintaining the river network and its ports might not be worth it if it does not add enough to the network to pay off for the costs.

Whether or not the river layer adds to the road layer is determined by these two parameters. This is the difficulty of the multiplex respect to the single layer networks: the need for the convenience factor and the inter-layer weight. Depending on their value, the measured properties of the multiplex will be different. Their actual value is unknown, and it probably varied from network to network. Our aim is not to determine them, but to use them to asses the relevance of the river network.

To see how good the interplay between layers is, for each age and region, including also the network of the

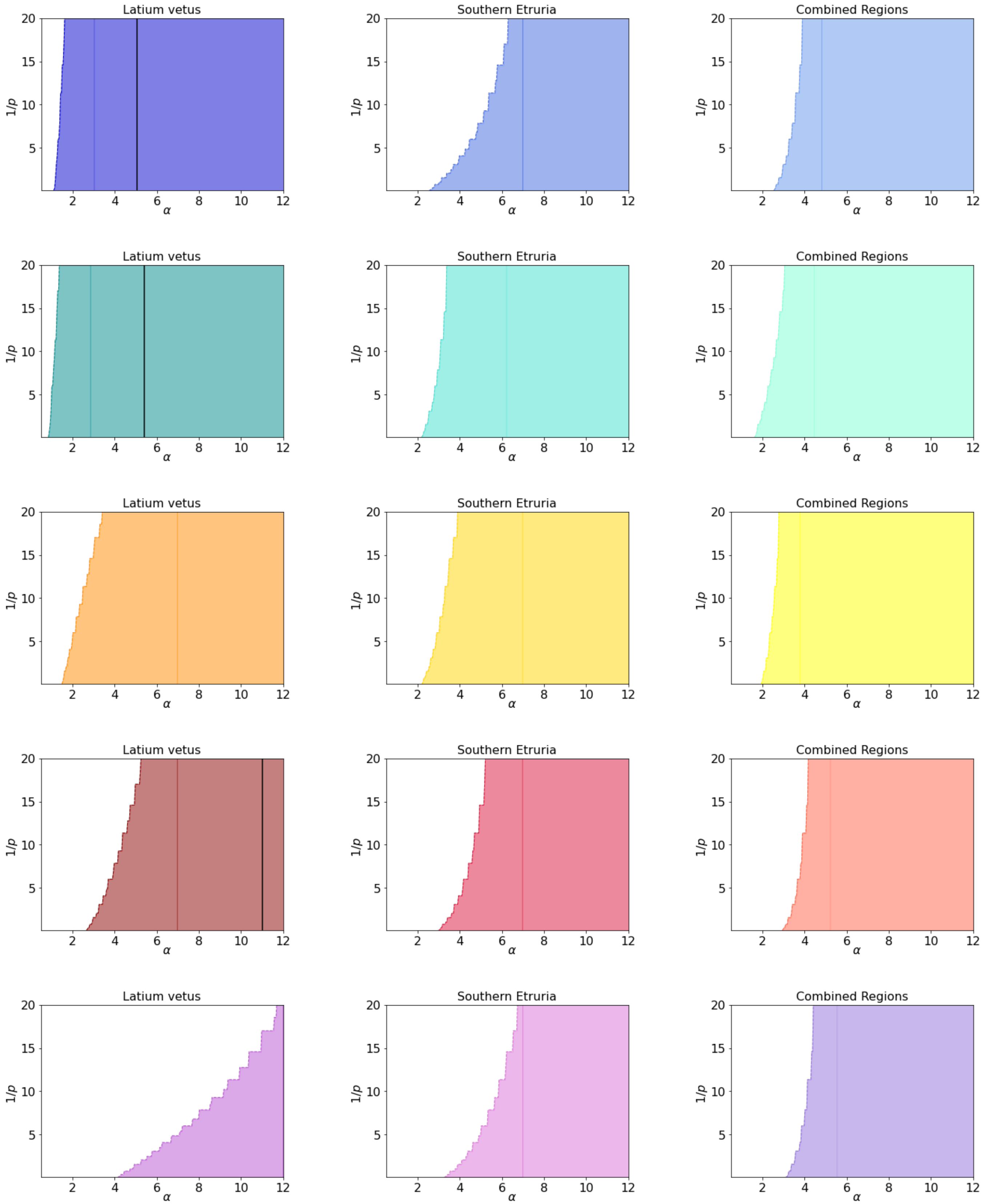

FIG. 7: Algebraic connectivity figures of road and river multiplexes. The river convenience factor α is a factor dividing distances, and thus multiplying weights, in the river layer. The port cost ($1/p$, the inverse of the inter-layer coupling) is the cost of maintaining a port in a settlement measured in kilometers of road maintenance. The colored area means an improvement in algebraic connectivity when adding the river layer. A larger colored area means good interplay between layers. The black vertical line corresponds to $\bar{\alpha}$ given by Eq. 8.

combined regions, and for the river convenience factor in the interval $[0.5, 12]$, we calculated the value of the port costs at which adding the river network increased the algebraic connectivity of the network. In Figure 7, the colored areas represent the range of the parameter values for which maintaining the multiplex network supposed an improvement respect to just the road network. The blank areas represent the conditions for which the river network did not compensate the maintenance costs. A larger colored area is an indication of a good interplay between rivers and roads, meaning the river layer will easily provide the road network with alternative paths, thus improving the overall resilience of the multiplex. We can clearly see that Latium Vetus made good use of the river network in the early ages, but it became more and more redundant as time passed and the region developed. This is not to say that later on the rivers were useless; quite on the contrary, the growing population and economy begged for alternative ways of transportation. Still, because the rivers rarely change —though the nodes do, and thus our network—, this figures provide a qualitative assessment of the development of the road network through the ages.

In the case of Southern Etruria, there is not such a progression. From our previous analysis of the road networks, we know that the algebraic connectivity of SE was worse than Latium's. A moderately convenient —in the sense that it provides alternative paths— river network should have easily increased the algebraic connectivity of the region. Instead, we find that the colored area is smaller in the EIA1E period, which means the river network was not good. During EIA1L and EIA2 the interplay between both layers improves. This is due to a better algebraic connectivity of the river network. Finally, in the later OA and AA periods the area is reduced due to the development of the road networks. Nevertheless, SE does not reach the level of development of LV, with is why it still needed to be complemented by the river network.

As for the combined regions, we can clearly see in the case of the Archaic Age that the interplay is better than the mean interplay of both regions. This is most likely due to the poorer road connections between regions compared to the Tiber river, which crossed from SE to LV.

The y axis in Figure 7 is the port maintenance costs $c = 1/p$, the inverse of the inter-layer coupling p. It should be noted that the interval $(0, 20]$ of c is within the $p > p^*$ regime for all figures. From Eq. 7, $p^* = 1/c^* = \lambda_2^{min}$. Thus, $c^* = 1/\lambda_2^{min}$, and the maximum λ_2^{min} in Table V is given by LV AA (only relevant if $\alpha > \bar{\alpha}$, see Eq. 8), which results in $c^* = 33$, way off the limit of our figures. Thus, in our figures, the algebraic connectivity of the multiplex will be bounded from above by that of the aggregated network, which can be better or worse than the road layer depending on the quality of rivers —which in turn depends on α—. There can be no improvement if the river layer is not good enough. The algebraic connectivity will approach this value as p increases.

We also plotted in Figure 7 a vertical black line corresponding to the values of $\bar{\alpha}$ as given by Eq. 8. For the figures with no such line, either it is outside the limit of the figure (Latium) or the quotient cannot be calculated because the river network is disconnected (Etruria and the Combined Regions). It could be approximated by the giant component, which would give an upper bound of $\bar{\alpha}$. In any case, on the right side of this line the river network is more resilient than the road network, so the multiplex will be more resilient than the road networks unless Eq. 9 is fulfilled, i.e., the ports are so prone to failures that they do not compensate for increase in resilience provided by the river paths.

B. Global Efficiency

Global efficiency is not well established in the multiplex formalism: the distance between nodes from different layers is difficult to define, and thus the distances in the network, even for nodes in the same layer, considering one can traverse from layer to layer, are also not well defined. Here I propose a definition of global efficiency in a multiplex. It is only valid for two-layer multiplexes, like the ones we have, and one of the layers needs to be regarded as the main layer, with the second layer acting as a complement. In our case, the main layer is the road network, while the river layer is there to possibly provide shorter paths. It is a reasonable choice considering one would need to embark to transverse through the river layer and would also need to disembark to arrive at its destination. Then, the global efficiency is calculated by taking Eq. (1) for the main layer, but the network paths can shortcut through the second layer:

$$E_{glob} = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{d_{ij}}{L'_{ij}} \tag{10}$$

N is the number of nodes in the main layer —the road layer—. Like in Eq. (1), L'_{ij} accounts for the shortest path length between nodes i and j in the network, but because we are dealing with a multiplex, this path may go through the second layer. Let us consider the example of Figure 6: if we only consider the main layer $l1$, the distance between nodes n_1 and n_3 is the sum of the distances d_{12} and d_{23}. When we add the second layer, there is an alternate path through the secondary layer of length d_{13}. To this distance we need to sum $2t$, where t is the inter-layer distance. This accounts for hopping into the secondary layer and back to the main layer. If this path provides a shortcut respect to the path on the main layer, i.e., $2t + d_{13} < d_{12} + d_{23}$, the global efficiency will be larger. The river network will be convenient not when it provides alternative pathways, but when it shortens distances between the settlements. It is no longer a question of whether maintaining the river network provides a good backup in case of node failures, but if it

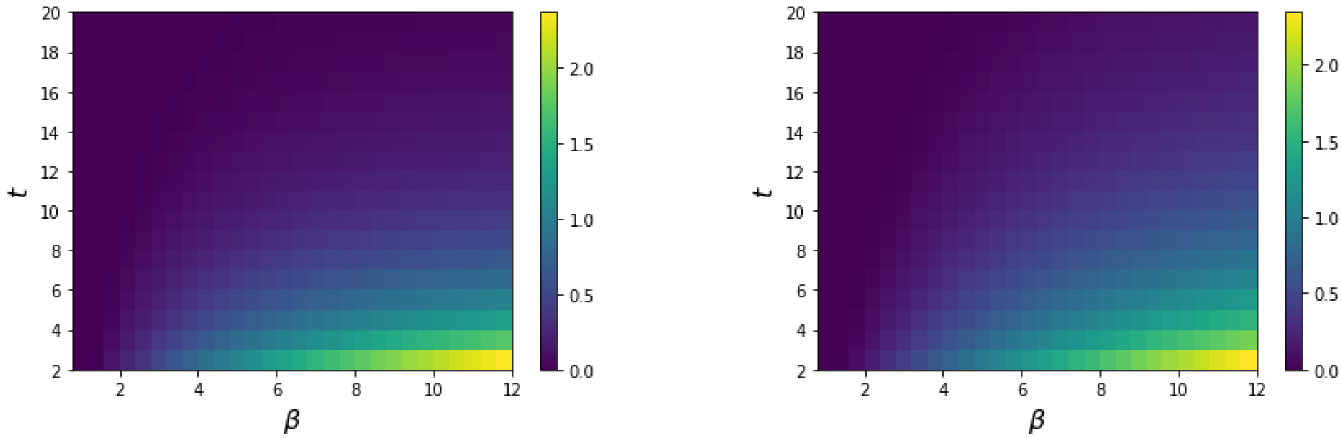

FIG. 8: Normalized global efficiency difference between multiplex and road network (Eq. 12). The relative cost β is a factor dividing distances in the river layer. The transfer cost t is the cost of traversing from one layer to the other measured in cost of traversing kilometers of road.

makes communications more efficient. In this sense, the distance in the network between two nodes can be seen as a transport cost instead of a maintenance cost, just like in the single layer approach. The inter-layer coupling t will be regarded as a *transfer cost*, the cost of traversing from one layer to the other. As well as in the case of algebraic connectivity, we need to account for the different nature of rivers. Also, there is evidence that transport through rivers had different costs than through roads. The parameter β, the *relative cost*, will divide distances in the river layer to account for this difference in transportation costs between roads and rivers. Although the parameters remain formally the same to their counterpart for the algebraic connectivity, their interpretation is completely different.

For another way in which the second layer can shorten the distances, back in Figure 6, consider (n_2, n_3). This link is present in both layers. Let us call X the distance in the main layer. Then the distance through the second layer will be $2t + \frac{X}{\beta}$. $\beta > 1$ will favor transit through the second layer. If

$$\begin{aligned} 2t + \frac{X}{\beta} &< X \\ X &> \frac{2t}{1 - \frac{1}{\beta}}, \end{aligned} \tag{11}$$

then the second layer will shorten the path, increasing the global efficiency. Note that, in this definition, adding the second layer can only increase the global efficiency of the network. In fact, this new definition of global efficiency allows values larger than one: the paths through the second layer can be shorter than the straight paths through the first layer if the relative cost β is high enough. We cannot normalize it by adding the nodes of the second layer to the normalizing factor, because then we would need to account for the paths between nodes in the second layer when applying Eq. 10. This does not make sense in our road and river framework. Having a definition of the global efficiency not well normalized is less than ideal, but for now it serves our purpose of extracting information out of the multiplex.

Like for the algebraic connectivity, we want to know how good the interplay between layers is in terms of the newly defined global efficiency. However, in Fig. (7) both the blank and colored areas represent a variety of values of algebraic connectivity, with the borderline representing the parameter for which road network and multiplex algebraic connectivity is the same. If we were to take this approach with the global efficiency, then the blank areas would represent the same value of the global efficiency, equal to that of the road network, while the colored area would represent an improvement in efficiency provided by the river network, which is different for every pair of parameters. Thus, it feels more natural to instead use a heat map to plot the normalized difference in global efficiency between the road network and the multiplex:

$$\Delta E_{glob} = \frac{E_{glob}^{mult} - E_{glob}^{road}}{E_{glob}^{road}} \tag{12}$$

Also, because the computation takes more time, Figure 8 only corresponds to the EIA1E of Latium and Etruria.

Obviously, left of $\beta = 1$ no improvement can happen (see Figure 8). As soon as we cross that threshold, however, the global efficiency increases. Note that we measure values larger than one. As β increases, the condition in Eq. 11 is met easily, so we arrive at values of global efficiency greater than one when β is large. Thus the brighter colors as we progress to the right of the figure. In fact, increasing β is more beneficial than decreasing the t, because there are more river kilometers than ports. Also, both layers of the empirical networks are very overlapped —many rivers run parallel to roads—, which encourages this effect.

Also, a horizontal asymptote can be inferred at some value t_{max}, above which the global efficiency cannot increase. If we take a look into Eq. 11, we can see that for a value of t, no matter how high beta, due to transfer costs, it is better to just traverse through the road layer. Then, t_{max} will be bounded from below by

$$\bar{t} = (1 - \frac{1}{\beta})\frac{X_{max}}{2} \tag{13}$$

where X_{max} stands for the longest road parallel to a river. The difference between this lower bound and the real t_{max} will depend on the alternate path the river network provides.

This new global efficiency is not as good a quantity as the algebraic connectivity. Being only able to increase, the moment there is a link slightly favored by a river, the difference will be larger than zero. Any combination of parameters out of the zero area in Figure 8 must not be taken as good interplay between layers. Instead, this heat maps should be used in combination with Figure 7 to better assess the impact of adding the rivers to the transportation network. The parameters are defined differently, so it is not possible to just overlap the figures. Still, their function is formally the same, so the figures can complement one another: if the colored area in the algebraic connectivity figures accounts for good interplay between layers, the heat maps of the global efficiency indicate how good.

VI. CONCLUSIONS

In this thesis, by following the work of [2] and [3], we have successfully translated ancient transportation infrastructures into complex network formalism. Then, through the global efficiency and the algebraic connectivity, we have analyzed the road and river networks of Southern Etruria and Latium Vetus and confirmed a number of old results, as well as found some other new ones.

Firstly, we corrected the Etruscan coordinates, which had been displaced by an error in the conversion from GIS. Then, by developing Etruria from the empirical network and from scratch (Figures 3a and 3b), we showed that the EE model does not lead to the best network in terms of global efficiency. We also found that, if provided with the same amount of resources and developed from scratch, both regions arrived at more or less the same values of global efficiency (Figure 3b). A perturbation of the coordinates could compensate for the difference (Figure 4); although a more detailed study of a complete perturbation protocol is advised, it shows that the difference is not significant. This result highlights the similarities between regions and supports the result of [9] that, had Latium not wasted resources on preferential attachment, the region could have accomplished a much better efficiency than Etruria. However, there is a difference in algebraic connectivity that arises in the later ages no matter the model used. It is a structural property of the networks due to the positions of the nodes, and the perturbation did not remove this property (Table IV).

Following in this line of work, a perturbation protocol should be devised and the response of the networks measured. Beyond a more detailed analysis of a Gaussian white noise for different amplitudes, perhaps it would be interesting to apply a perturbation that depends on the node strength. Highly connected settlements are also the ones receiving more information, so it is reasonable to assume that they would be the most affected by possible information errors. This perturbations should be simple enough to implement and would not require much work.

After the road network analysis, we measured the global efficiency and algebraic connectivity of the river networks and the roads and rivers of the combined regions. As expected, river networks had much poorer properties than road networks, because they could not be build. Thus, we explored the multiplex approach in terms of adding a poorer layer to a richer layer. We provided two different interpretations of the parameters of the multiplex, each suited for the quantity at hand. Regardless of their definitions, they behaved effectively the same and thus could complement each other when extracting information from the multiplex. We showed that the algebraic connectivity suffered two —though we simplified to one— transitions depending on the value of the inter-layer weight p. We then represented the interplay between layers in Figure 7, which accounted only for a set of parameters for which $p > p^*$, i.e., the algebraic connectivity was bounded from above by that of the aggregated network.

Finally, we proposed a definition of global efficiency suitable for a two-layer multiplex. Although not well normalized and in need of a more strict definition, it has proved to be useful when applied to our multiplexes. Of course, the first thing to do next should be to apply it to all ages, not just EIA1E. A more formal study of the multiplex formalism should be next, both in terms of algebraic connectivity and the global efficiency, which should be formally defined. This would be a more difficult task and would require time. Furthermore, the results in Sec. IV could be attempted with the multiplex. However, the parameters of the multiplex would pose a problem, for we do not know their value. Instead, one could try to mirror the processes in [2] and [3] with the multiplex and find the combination of parameters that, taking the rivers into account, best reproduces the empirical road networks. This would also take time, though, either computationally by using our codes, or by translating them into a more efficient language.

Hopefully by now we have shown not only that networks are an effective tool to analyse ancient transportation infrastructures, but that the multiplex approach can be a very powerful tool for doing so. We expect this thesis to serve as a basis for future work focused on further studying the multiplex properties of this case study, and, possibly, others.

ACKNOWLEDGMENTS

We would like to thank Emanuele Cozzo for the advise provided on the mathematical formalism of the multiplex.

[1] L. Prignano, I. Morer, and A. Diaz-Guilera, Wiring the past: a network science perspective on the challenge of archeological similarity networks, Frontiers in Digital Humanities **4**, 13 (2017).

[2] L. Prignano, I. Morer, F. Fulminante, and S. Lozano, Modelling terrestrial route networks to understand inter-polity interactions (southern etruria, 950-500 bc), Journal of Archaeological Science **105**, 46 (2019).

[3] F. Fulminante, L. Prignano, I. Morer, and S. Lozano, Coordinated decisions and unbalanced power. how latin cities shaped their terrestrial transportation network, Frontiers in Digital Humanities **4**, 4 (2017).

[4] T. Brughmans and M. A. Peeples, Network science, in *The Encyclopedia of Archaeological Sciences* (John Wiley Sons, Ltd, 2018) pp. 1–4.

[5] E. Cozzo and Y. Moreno, Characterization of multiple topological scales in multiplex networks through supra-laplacian eigengaps, Physical Review E **94**, 052318 (2016).

[6] E. Cozzo, G. F. de Arruda, F. A. Rodrigues, and Y. Moreno, Layer degradation triggers an abrupt structural transition in multiplex networks, Physical Review E **100**, 012313 (2019).

[7] I. Morer, A. Cardillo, A. Díaz-Guilera, L. Prignano, and S. Lozano, Comparing spatial networks: A one-size-fits-all efficiency-driven approach, Physical Review E **101**, 042301 (2020).

[8] R. J. Sánchez-García, E. Cozzo, and Y. Moreno, Dimensionality reduction and spectral properties of multilayer networks, Physical Review E **89**, 052815 (2014).

[9] H. S. Dawson and F. Iacono, *Bridging Social and Geographical Space Through Networks* (Sidestone Press, 2021) pp. 31–46.

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