Social Network Analysis, Community Detection Algorithms, and Neighbourhood Identification in Pompeii

Notarian, Matthew[1]

[1] Hiram College – Hiram, Ohio, USA

Correspondence: notarianmf@hiram.edu

ABSTRACT

The definition and identification of urban neighbourhoods in archaeological contexts remain complex and problematic, both theoretically and empirically. As constructs with both social and spatial characteristics, their detection through material culture alone remains elusive, especially within large settlements that are incompletely excavated or preserved. Thanks to its focus upon relational ties, network analysis offers a profitable path towards untangling the complexities of urban neighbourhoods, especially with respect to their often imprecise, fuzzy boundaries. Various community detection algorithms offer mathematical solutions for partitioning large graphs into communities, but these should not be applied without careful interpretation. Two of the most widely utilized community detection algorithms based on modularity optimization, Louvain and Leiden, contain a customizable resolution parameter that is often overlooked by practitioners. This controls the density of the partitioned communities, and therefore the number identified, but it is difficult to determine the optimal value for any given network. In addition, the results of community detection algorithms vary stochastically. Reliance upon a single iteration may mask potentially significant differences between runs using even a constant resolution parameter. A recently developed algorithm, the Convex Hull of Admissible Modularity Partitions (CHAMP), is designed to overcome these complications and also generates potentially useful multiscalar network community structures. Its applicability to neighbourhood archaeology is demonstrated within three networks of Pompeian housing units based on shared public fountains.

The case study examines Pompeii's public fountains as hubs of social interaction. Given their daily frequentation by nearby inhabitants, fountains represent plausible proxies for the centres of definable neighbourhoods. It expands upon a spatial network model that connected all 2000+ external doors in the city to the 40 public fountains that were likely functional in 79 CE. Three undirected, one-mode networks were constructed in which units are linked to each other by a common fountain and weighted by the number of fountains they share. The first network connects units by the closest fountain to any external door. Since many properties had side doors within reach of different water sources, these represent potential interconnections between communities. The second and third networks use incrementally larger time to fountain thresholds (30-second and one-minute walks to any fountain, respectively) to map potential choices, also expanding social integration. The results demonstrate that innovative methods for assessing the output of community detection algorithms offer new modes of analysis that are applicable not just to neighbourhood archaeology, but any archaeological network analysis that uses graph partitioning.

Keywords: network analysis, community detection, neighbourhoods, urban archaeology, Roman archaeology, Pompeii

Introduction

Neighbourhood archaeology (e.g. Pacifo & Truex, 2019a; Thompson et al., 2022; Haug et al., 2023) and network community detection (Shai et al., 2020) are two topics that have seen increasing interest in recent years within their respective disciplines. In general, community detection algorithms seek to identify clusters of nodes that are more densely connected to each other than others in the network, a process with many useful applications in archaeology. Studies have used them to explore the political affiliation of Maya polities (Scholnick et al., 2013), spatial scales of ceramic similarity networks (Peeples & Bischoff, 2023), Neolithic household community organization (Mazzucato, 2019), and dendrochronological provenance patterns (Visser, 2021). Nonetheless, advances in the quality and efficiency of community detection algorithms, and innovative methods for assessing their outputs, offer new modes of analysis that have yet to be put to widespread use in archaeological network research. This paper focuses on the potential of community detection to identify urban neighbourhoods in Pompeii, but the methodologies outlined have implications for numerous applications within archaeological network analysis that use graph partitioning.

Two of the most widely utilized community detection algorithms based on modularity optimization, Louvain and Leiden, contain a customizable resolution parameter (γ) that is often overlooked by practitioners. This controls the density of the partitioned communities, and therefore the number identified, but it is difficult to determine the optimal value for any given network. In addition, the results of community detection algorithms vary stochastically. Reliance upon a single iteration may mask potentially significant differences between runs using even a constant resolution parameter. A recently developed algorithm, the Convex Hull of Admissible Modularity Partitions (CHAMP), is designed to overcome these complications and also generates potentially useful multiscalar network community structures (Weir et al., 2017). Its applicability to neighbourhood archaeology is demonstrated within three one-mode networks of Pompeian housing units based on shared public fountains.

Neighbourhoods in Pompeii based on Shared Public Fountains

The definition and identification of urban neighbourhoods in archaeological contexts remain complex and problematic, both theoretically and empirically (Smith 2010; Pacifico & Truex, 2019b). Are neighbourhoods primarily top-down administrative units, organic bottom-up household agglomerations, or something in between? Is face-to-face interaction an important prerequisite or can larger districts exist at a more dispersed level of contact? Crucially, as constructs with both social and spatial characteristics, their detection through material culture alone remains elusive, especially within large settlements that are incompletely excavated or preserved. Nonetheless, in Pompeii, many studies have noted that public fountains were centres of daily social interaction for nearby inhabitants (Laurence, 2007; Notarian, 2023). Therefore, they can be considered plausible proxies for definable neighbourhoods. These are often defined by simple linear distances from fountains. Yet, thanks to its focus upon relational ties, network analysis offers a more profitable path towards untangling the complexities of urban neighbourhoods, especially with respect to their often imprecise, fuzzy boundaries (Poorthuis, 2018). It also generates quantitative metrics with which to explore theoretical questions regarding the nature of urban neighbourhoods.

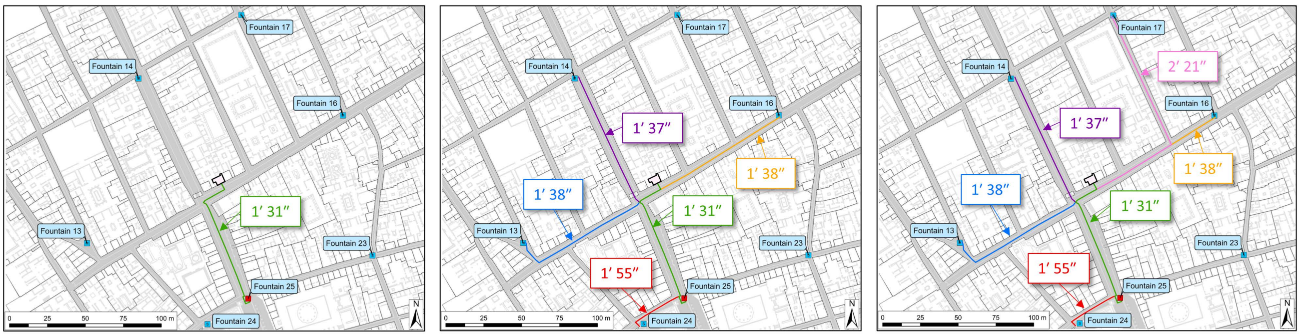
The excellent preservation of Pompeii's streets and domestic units permits the creation of a detailed social network based on shared fountain use. This study expands upon an earlier GIS spatial network model that connected all 2000+ external doors in the city to the 40 public fountains that were likely functional in 79 CE, when the city was destroyed (Notarian, 2023). Two points layers represented doors and fountains, respectively. A polyline along the street network connected these points, which was then subdivided into

3m segments. Least-cost routes to the closest fountain from every door were identified using ArcGIS' Closest Facility function, but costs were derived in units of time rather than linear distance. Elevation values from an underlying digital elevation model were imported to each network segment endpoint to calculate its average slope. Various slope-based walking functions were used to estimate speed, which was then converted to costs in minutes. The results of this analysis revealed important points that had previously been overlooked. First, many housing units with multiple exterior doors are closer to different fountains depending upon the door exited. Second, an inhabitant's choice of fountain often depended upon differences of mere seconds in walking time. In such situations it is difficult to predict which fountain a particular unit may have frequented. Individual agency must have determined the composition of the social communities tied to fountains more than previously assumed. Further, if Pompeians habitually used *multiple* fountains, the neighbourhoods formed by this activity would have been far more dynamic, fuzzy, and overlapping than a static division can represent.

Methods

To further explore these dynamics, three undirected, one-mode networks were constructed in which units are linked to each other by a common fountain and weighted by the number of fountains they share. The networks were built and analysed in Gephi (Bastian et al., 2009) and the R packages qgraph (Epskamp, 2012) and igraph (Csárdi et al., 2023). First, ArcGIS' Closest Facility analysis identified the 10 closest fountains to every external door. The door to fountain times were exported to a csv and ranked. Ranking by door and unit identified the quickest route from any unit entrance, which formed the basis of the first network. Many properties had side doors closer to different water sources – potential interconnections between communities. Ranking by unit alone was used to construct the other two networks. Based on incrementally larger time to fountain thresholds - an additional 30-seconds and one-minute of walking time beyond that to the closest fountain, respectively - these networks expand choice and also social integration. After cumulatively summing walking times to successive fountains to identify the relevant edges, each network was converted into an edge table and imported into Gephi as a bimodal network (unit to fountain), then converted to a one-mode network (unit to unit) using the multimode network transformation plugin. A node list imported key unit attributes, such as coordinates, so that the networks could be mapped spatially.

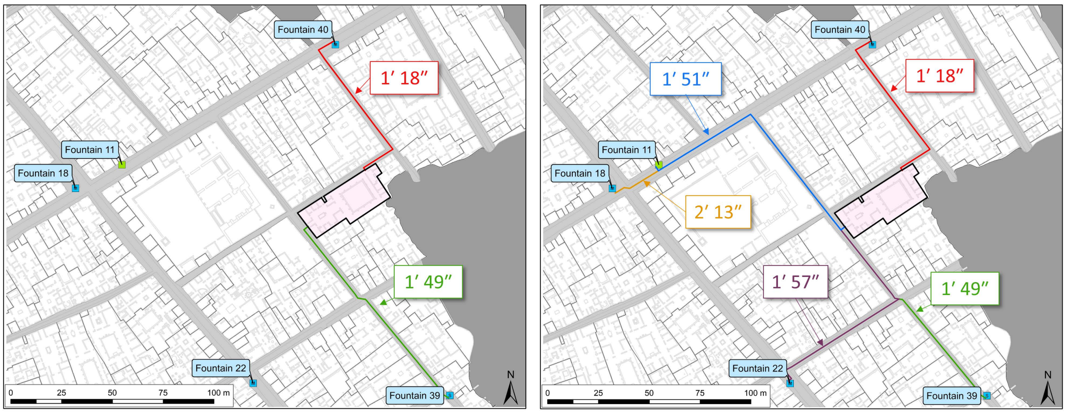
Defining the Three Networks



A. B. C.

Figure 1 – The three network scenarios as applied to a single unit, the taberna at VI.10.12 (outlined and highlighted pink). **A)** The closest fountain to its external door. **B)** Fountains within 30 additional seconds to its closest fountain. **C)** Fountains within one additional minute to its closest fountain.

A couple examples illustrate how the three network scenarios were formed. Taberna VI.10.12, with only one door, is closest to fountain 25, a walk of one and a half minutes (fig. 1a). By expanding the time threshold by 30 additional seconds, four other fountains are reached (fig. 1b). Extending the threshold by a full minute from the closest fountain connects one more (fig. 1c). In the social network, the assumption is that inhabitants of this shop frequented each option indiscriminately.



A. B.

Figure 2 – The three network scenarios as applied to a single unit, the house at IX.6.5 (outlined and highlighted pink). **A)** The closest fountains to its two external doors. Also serves as its 30-second network. **B)** Fountains within one additional minute of the absolute closest fountain (n.40).

In the case of unit IX.6.5, a mid-sized house with two external doors, there is no single closest fountain as the nearest in terms of walking time varied by door (fig. 2a). Expanding the threshold by 30 seconds from the absolute closest (fountain 40) does not connect any additional fountains, but an additional minute of travel time would find these inhabitants visiting five other fountains (fig. 2b).

Following this process across all 1249 inhabited units in the city creates three different social networks with varying levels of interconnection. Although the larger project subjects these networks to a variety of global and local metrics to explore their characteristics, this paper solely focuses upon graph partitioning (table 1). To obtain a sense of the location and size of the neighbourhoods formed by water collection, each network had to be partitioned into communities. With its basis in the density of the social network, this should constitute a more robust approach to Pompeii's fountain neighbourhoods than mere spatial distance alone, while also allowing for human agency with multiple fountain options for individual units.

Table 1 – Global metrics of the three Pompeian fountain social networks.

Threshold	Edges	Graph Density	Network Diameter	Average Path Length	Transitivity
Closest Fountain to Door	27,267	3.5%	15	5.6	0.908
30 Second	50,240	6.4%	9	3.8	0.705
One Minute	83,411	10.7%	9	3.25	0.667

Modularity Optimization and Stochastic Results

Many community detection methods, such as the Louvain algorithm and the newer Leiden algorithm, rely upon modularity (Q), which is a measure of the quality of cluster partitioning (Newman & Girvan, 2004):

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \gamma \frac{k_i k_j}{2m} \right) \delta(c_i, c_j) \qquad (1)$$

where m is the total number of edges in the network, A_{ij} is the adjacency matrix for nodes i and j, k_i and k_j are the degree of nodes i and j, respectively, γ is a resolution parameter, and $\delta(c_i, c_j)$ is the Kronecker delta function, which checks whether nodes i and j belong to the same community (a value of 1) or not (a value of 0). Modularity compares the total weight of within community edges (A_{ij} summed over every node in the community) to the expected weight in a random null model with the same degree distribution, the so-called configuration model ($\frac{k_i k_j}{2m}$). The configuration model rewires the network randomly while preserving each node's degree. The denser the community in comparison to the null model, the higher its modularity score, and theoretically, the better the partition. However, modularity maximization of all but the smallest networks is NP complete if done brute force by testing every possible network configuration, making it impossible to compute absolutely. As a result, several less computationally demanding heuristic algorithms have been developed, Louvain and Leiden included, to approximate optimal partitions. Nonetheless, these suffer from limitations. Most produce stochastic results when run multiple times. They output a range of near-optimal partitions that are close to the true maximum, but often vary widely in composition. Furthermore, there is no objective way to choose between varying solutions, which renders interpretation confusing. Modularity also has a resolution limit in which smaller communities are absorbed into larger ones, making it unable to detect communities smaller than a certain threshold depending on the size of the network.

The Louvain algorithm is among the most popular community detection methods used in network science (Blondel et al., 2008). It is found in many network packages such as Gephi and R and Python igraph.[1] Importantly, Louvain utilizes an adjustable resolution parameter (γ) that controls the number of returned communities. This was intended to help users work around the resolution limit. By multiplying the null model against γ, greater values lead to dense, small, and more numerous communities. Small γ values have the opposite effect, leading to fewer loose, large communities. The resolution parameter, however, was not always available in every package, or was hidden within user guides, leading many people to use the default γ value of 1.

The Leiden algorithm improved the speed and quality of Louvain. It corrected an issue in which Louvain could create communities that were not internally connected (Traag et al., 2019). It also contains an optional alternate quality function called the Constant Potts Model (CPM) that avoids the resolution limit (Traag et al., 2011). In this version, the algorithm measures the quality of a partition not by comparing it against a null model, but by directly filtering out connections below a certain density threshold (the so-called constant, also called γ). Like Louvain, Leiden is found in many of the most popular network packages.[2] Due to its efficiency and accuracy, even with extremely large networks, Leiden is used exclusively in this paper as a community detection method for archaeological networks. Other community detection formulas can produce equally valid results, but these remain outside the focus of this study.

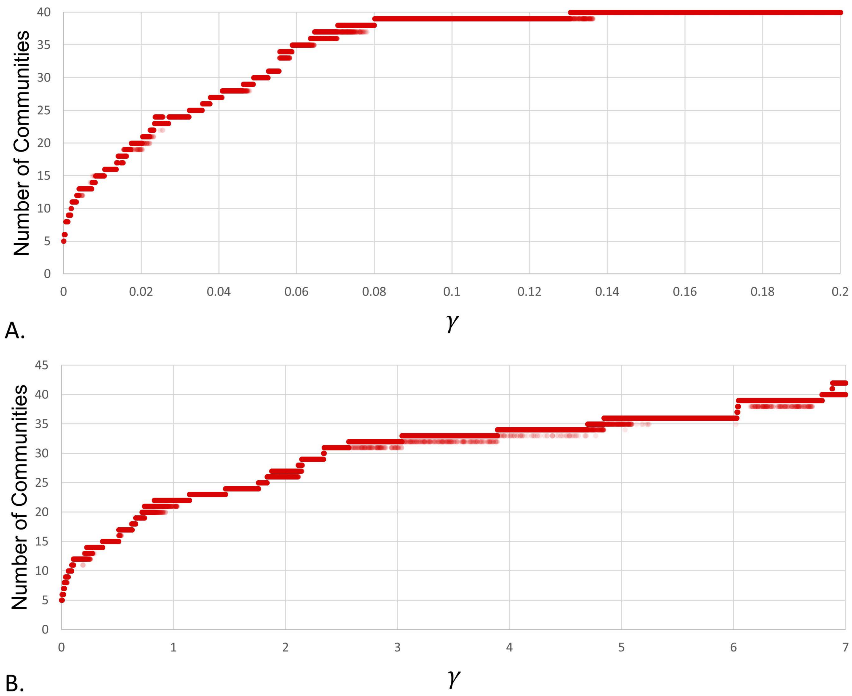


Figure 3 – The number of communities partitioned by 125,000+ runs of the Leiden algorithm on the closest fountain by door network. 5,000 intervals between the minimum and maximum values of γ, with each value of γ run 25 times with 25 iterations per run. **A)** Constant Potts Model. **B)** Modularity.

While resolution parameters were developed to give users more flexibility with which to explore their data, they lead to a bewildering set of choices. For any value of γ using either CPM or modularity, one can obtain numerous partitions. Modifying γ can produce any number of communities. But how can one determine which partition is the best representation of a particular set of data?

[1] https://github.com/vtraag/louvain-igraph; https://r.igraph.org/

[2] https://github.com/vtraag/leidenalg; https://r.igraph.org/; https://gephi.org/plugins/#/plugin/leiden-algorithm-gephi-plugin

Partitions can be explored by iterating the algorithm at multiple values of γ and comparing the number of communities produced. As an example, Leiden was run 25 times, with 20 iterations per run, along 5000 γ intervals within the range that produces fewer than 40 communities – the number of fountains in the network (fig. 3). In the current case study, only communities containing one or more fountains are valid. Leiden efficiently completed these 125,000+ iterations in only a few hours. Observing a plot of the results, one might be tempted to view plateaus of a certain number of communities across a range of γ as evidence of a more robust community structure at these levels. Nonetheless, stochasticity often produces different numbers of communities even at the same γ. This method, however, does not reveal meaningful information about the actual partitions produced. Although there are multiple partitions of, for example, 36 communities, across a range of γ, it does not mean that these all represent the same 36 communities. There could be significant differences between the results, such as communities composed of vastly different units, that are not apparent without further exploration. Yet, with so many solutions, it is impossible to directly compare them.

The Convex Hull of Admissible Modularity Partitions (CHAMP)

Another algorithm, the Convex Hull of Admissible Modularity Partitions (CHAMP), available as a Python package, offers a method to significantly narrow down a subset of partitions that are potentially dominant (Weir et al., 2017).[3] After Leiden has identified hypothetically optimal partitions on a network, CHAMP calculates the modularity score for each partition using a slightly modified equation (1) across a wide range of γ, regardless of the value at which it was originally identified. The result is a linear function of γ versus modularity for each partition (fig. 4). A convex hull of the intersections of these lines is used to identify the upper envelope boundary. This isolates the partitions that produce the greatest modularity score across a specific range of γ, which represent the best quality partition within that span.

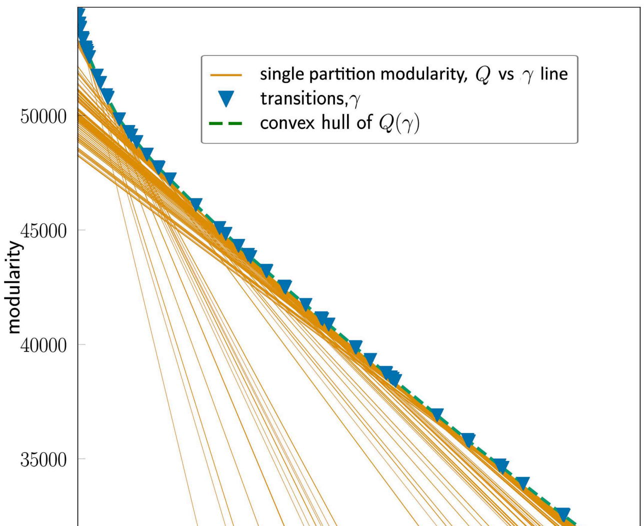


Figure 4 – A graphical representation of the CHAMP algorithm applied to the Closest Fountain to Door network. Leiden was run with 5000 intervals between γ 0 and 7 at 25 iterations each. Each solid orange line represents a single partition identified by Leiden (here reduced to 500 for clarity), but plotted as a linear function of modularity (Q) versus γ. The convex hull of these lines (green dashed line) represents the upper envelope, that is, the partitions that produce the highest modularity score across a given range of γ. The blue triangles mark the intersection of the partition lines along the convex hull, which are the boundaries between optimal partitions, reduced to 61 from the initial 5000 results.

The pruned set of partitions can then be compared with respect to the width of the γ domain in which it is optimal and evaluated against the entire set of unique partitions using a pairwise Adjusted Mutual

[3] https://github.com/wweir827/CHAMP

Information (AMI) heatmap. AMI represents the concordance of partitions to each other on scale of 0 (no alignment) to 1 (complete concordance). Using the heatmap, one can visually assess not only which partition has the widest domain of optimality, but also see how closely it relates to its neighbours in the γ space. This presents a more manageable and powerful method of directly comparing partitions than relying upon the number of communities alone. Importantly, it also eliminates from consideration the vast majority of sub-optimal partitions that result from stochasticity and improper γ selection for a given set of data.

Results and Discussion

This section presents the results of applying CHAMP across the three Pompeian fountain neighbourhoods to illustrate its utility in choosing partitions. It also offers some preliminary interpretations as they relate to defining urban neighbourhoods in Pompeii. At the time of analysis, CHAMP was unable to correctly work with CPM, so modularity alone was used. As a heuristic for exploring network data, modularity is likely more appropriate as the null model uses the network structure to identify clusters rather than an arbitrary density imposed by the user. The script was also slightly modified to correct minor errors and meet the requirements of this study.[4] CHAMP was run on each of the three networks twice - once with 1000 γ intervals and again with 5000 intervals - to test its reproducibility. The results were far more consistent than a typical Leiden run, with each iteration producing the same top partitions almost each time.

Closest Fountain to Door Network

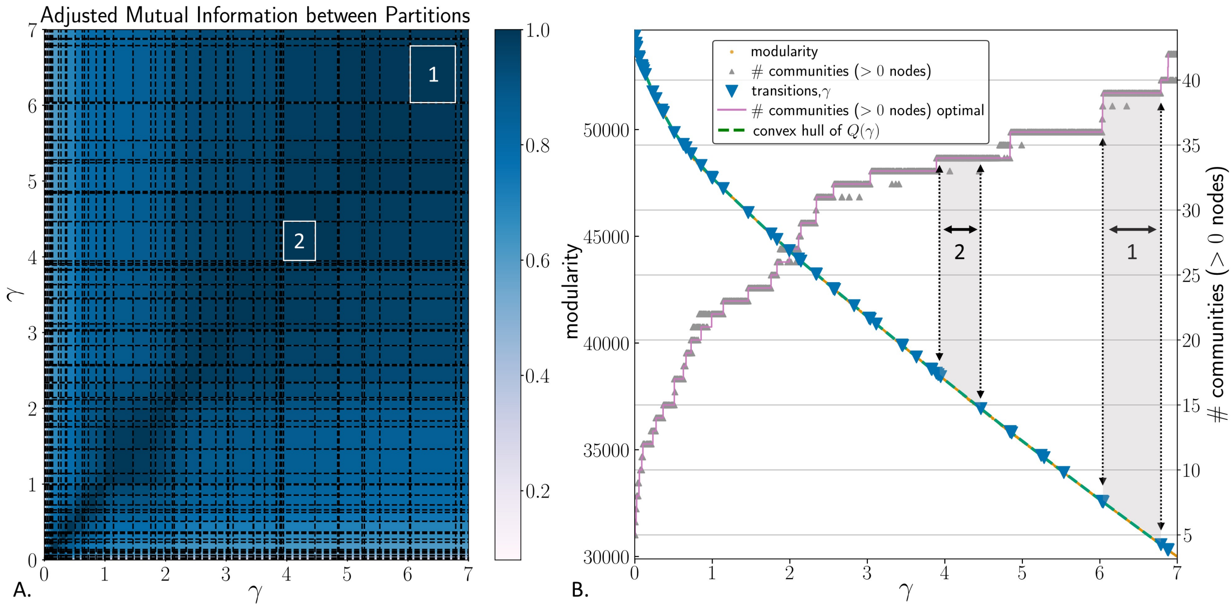


Figure 5 – Closest Fountain to Door Network. Results of CHAMP using Leiden algorithm with modularity. 5000 intervals between γ 0 and 7 at 25 iterations each. **A)** Pairwise adjusted mutual information (AMI) of the subset of partitions identified by CHAMP. Dashed lines represent transitions between ranges of γ. The widest domain (no. 1: γ 6.04 - 6.79, 39 neighbourhoods) and second widest domain (no. 2: γ 3.95 - 4.47, 34 neighbourhoods) are outlined in white. **B)** Modularity versus γ. Grey upward triangles represent a random sampling of 2000 partitions plotted against their number of communities, while the solid pink stepped line shows the number of communities in the optimal partition for the range of γ. The dashed green line is the convex hull. Blue downward triangles show the transition points between various ranges of γ. The grey zones between the dotted lines identify the widest domains (no. 1: 39 neighbourhoods, no. 2: 34 neighbourhoods).

With the network formed by the closest fountain to each unit door, CHAMP pruned 196 unique partitions down to 61, but this still represents a rather volatile community structure (fig. 5). The widest γ

[4] Modified code and data are available online: https://zenodo.org/badge/latestdoi/686189050

domains are mainly found in higher γ values, with 34 or more communities. This is expected since the network exhibits a modular topology with few interconnections between clusters. Many partitions, however, are remarkably similar, as indicated by the uniformity of dark blue in the AMI matrix. In fact, all the optimal partitions between γ 3.9 and 7 have AMI scores > 0.9. The widest domain in both CHAMP runs was a partition of 39 communities. This partition, however, contains three singleton communities - single housing units defined as whole neighbourhoods. For the purposes of this study, this partition must be rejected, underlining the importance of closely inspecting the output of the algorithm within the parameters of specific research questions.

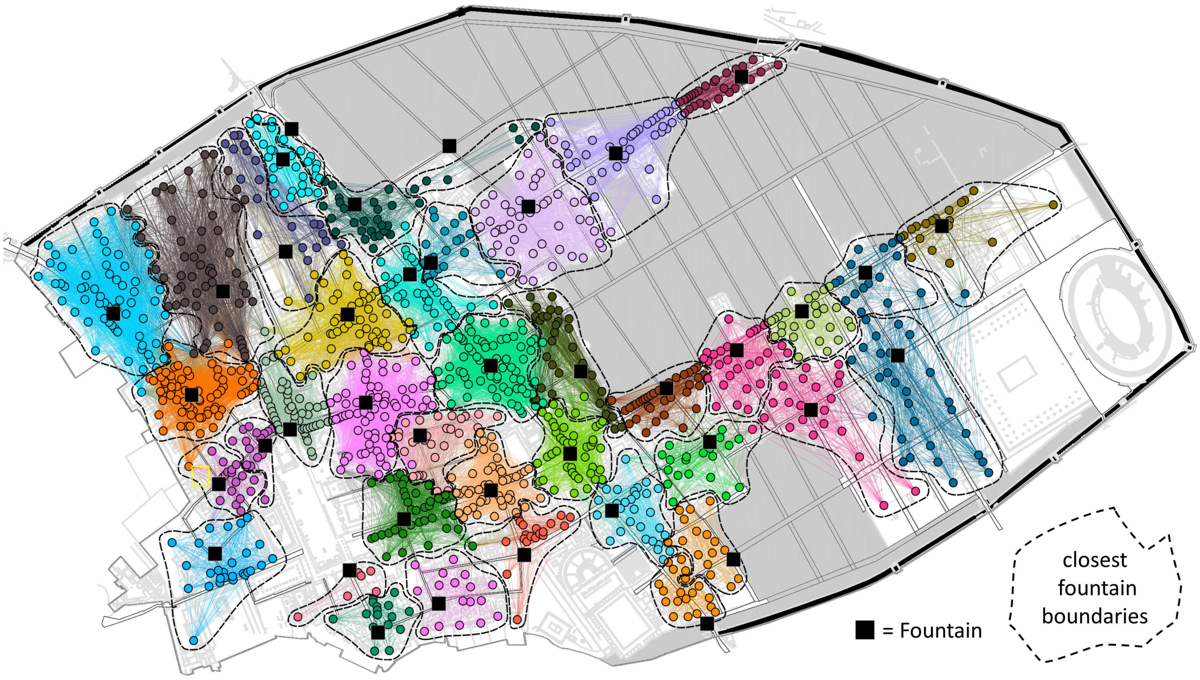


Figure 6 – Closest Fountain to Door Network. Partition into 34 neighbourhoods as identified by the second widest γ domain output by CHAMP. Colours represent neighbourhoods as partitioned by CHAMP. Dashed lines outline zones defined by the single closest fountain to each unit in the spatial network.

The next widest γ domain is a partition into 34 neighbourhoods. Mapping this partition in topographical space and colouring each node by its community permits a visual inspection of the neighbourhoods' spatial characteristics (fig. 6). Dotted lines indicate the boundaries of the closest fountain in the *spatial* network, that is, the single closest fountain in terms of walking time from any external door belonging to a unit, allowing comparison between the two. The detected neighbourhoods largely mirror the closest fountain to each unit in the spatial network, except for 6 communities formed by pairs of adjoining fountain users in the spatial network (e.g., the light blue neighbourhood near the Vesuvian gate). Interestingly, these neighbourhoods must share enough interconnection to constitute single communities, defying the overall pattern. Nonetheless, these results affirm an insular network structure in which a tiny number of units formed vital interconnections between small, tightly clustered neighbourhoods.

30 Second Network

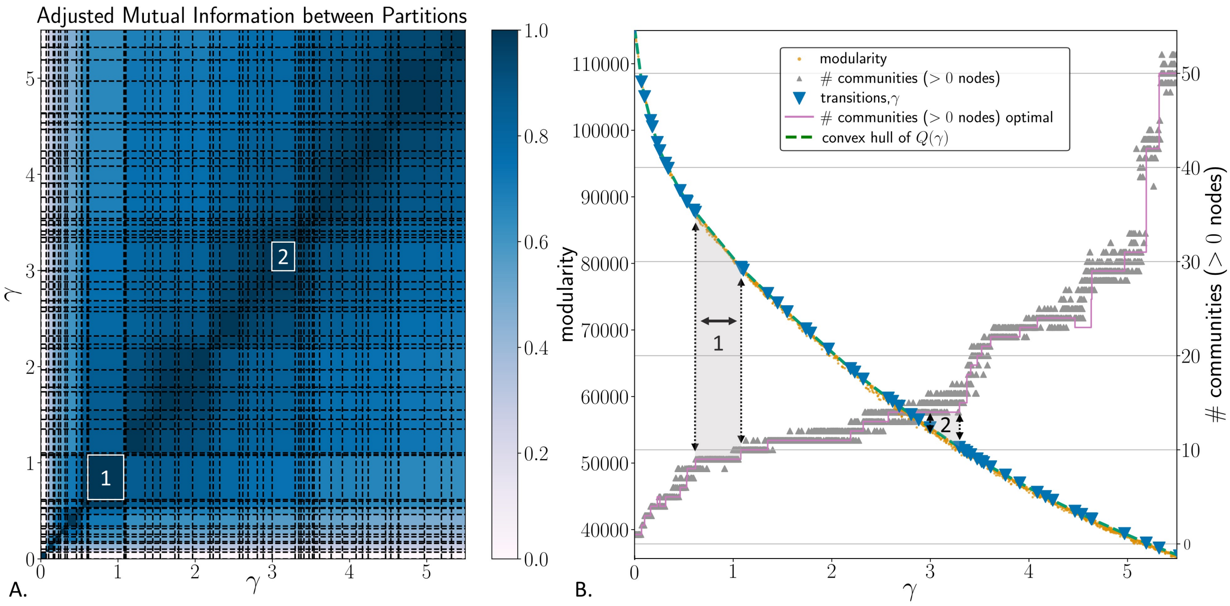


Figure 7 – 30 Second Network. Results of CHAMP using Leiden algorithm with modularity. 5000 intervals between γ 0 and 5.5 at 25 iterations each. **A)** AMI of the subset of partitions identified by CHAMP. The widest domain (no. 1: γ 0.62 – 1.08, 9 neighbourhoods) and third widest domain (no. 2: γ 3.0 – 3.3, 14 neighbourhoods) are outlined in white. **B)** Modularity versus γ. Grey upward triangles represent a random sampling of 2000 partitions plotted against their number of communities, while the solid pink stepped line shows the number of communities in the optimal partition for the range of γ. The dashed green line is the convex hull. Blue downward triangles show the transition points between various ranges of γ. The grey zones between the dotted lines identify the widest domains (no. 1: 9 neighbourhoods, no. 2: 14 neighbourhoods).

The 30 second network, with almost double the number of edges, represents a much more interconnected city, and this is reflected in the CHAMP results (fig. 7). The algorithm pruned more than 2400 unique partitions down to 57. A 9-neighbourhood partition, clearly delineated in the AMI heatmap, was returned as the widest domain twice. However, also visible in the matrix are at least four other "scales" of smaller communities at larger ranges of γ. These are represented by larger dark blue rectangles in the AMI plot that cross several dashed γ transitions. These might represent sub-divisions of the 9 neighbourhoods, or another kind of grouping altogether, a topic further explored below. Comparing the 9-neighbourhood map to the closest fountain divisions, we can see these communities combine users of several fountains, but near the interfaces some units are grouped with more distant fountains (fig. 8a). These internal community characteristics can be intuitively measured using belonging degree (Poorthuis, 2018) (fig. 8b). This is the fraction of a node's intra-community weight over its total weight. Nodes with a belonging degree of 1 only share ties with other community members, while those with lower values share ties with nodes in other neighbourhoods. Not surprisingly, nodes with the lowest belonging degree are found in the centre of the city, adjacent to neighbourhood boundaries. Nodes with low levels of belonging are likely those whose community membership might shift stochastically with each iteration of most community detection algorithms. The hearts of the neighbourhoods though are far more stable.

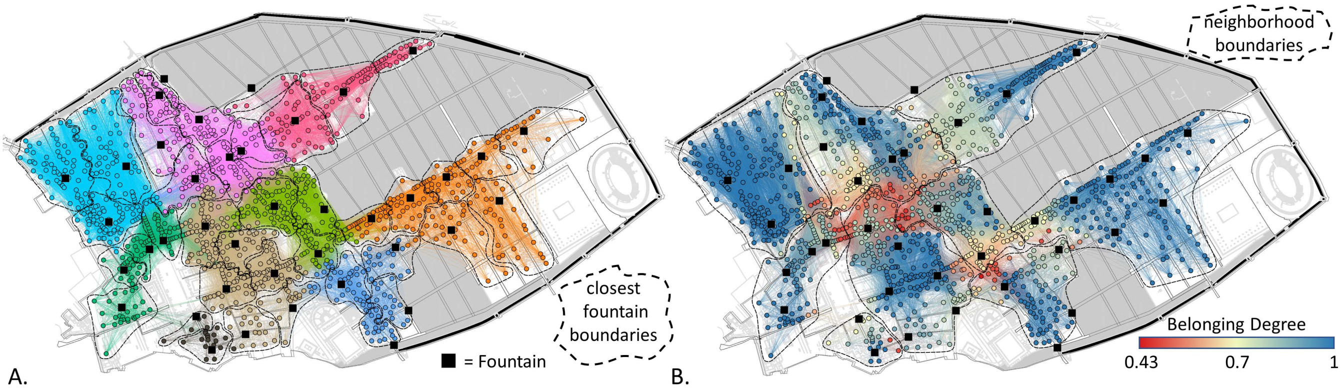


Figure 8 – 30 Second Network. Partition into 9 neighbourhoods as identified by the widest γ domain identified by CHAMP. **A)** Colours represent neighbourhoods as partitioned by CHAMP. Dashed lines outline zones defined by the single closest fountain to each unit in the spatial network. **B)** Colour ramp represents each node's belonging degree. Dashed lines outline the neighbourhood boundaries as partitioned by CHAMP.

The second widest domain returned by CHAMP, with 29 neighbourhoods, was rejected due to singleton communities, but the third widest consists of 14 communities. A contingency table allows us to assess whether these represent smaller subdivisions of the 9-neighbourhood partition (table 2).

Table 2 – Contingency table. 30 second network. Widest domain (9 neighbourhoods, lettered) compared with third widest domain (14 neighbourhoods, numbered).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	All
A	0	129	0	0	0	0	81	0	5	0	0	0	0	15	230
B	0	0	0	5	0	0	0	77	0	0	69	0	0	39	190
C	0	0	125	0	0	0	0	0	4	0	0	0	59	0	188
D	134	0	0	0	0	0	0	0	0	36	0	0	0	0	170
E	0	0	0	0	106	0	0	0	0	37	0	0	0	0	143
F	0	0	0	109	0	0	0	0	0	0	0	0	0	0	109
G	0	0	0	0	0	0	0	0	66	0	0	37	0	0	103
H	0	0	0	0	0	93	0	0	0	0	0	0	0	0	93
I	0	0	0	0	0	0	0	0	0	0	0	23	0	0	23
All	134	129	125	114	106	93	81	77	75	73	69	60	59	54	1249

Some of the 14 districts are equal subdivisions of larger communities. Neighbourhoods 3 and 13 in the 14-neighbourhood partition, for example, mostly bisect neighbourhood C in the 9-district partition. However, other small communities cut across larger ones, taking nodes out of two or three, often those with low belonging degrees. For instance, number 9 in the 14-neighbourhood partition draws in units from neighbourhoods A, C and G, while community 10 is evenly split between neighbourhoods D and E. In terms of Pompeian neighbourhood structures, this might suggest that at smaller scales, the composition of communities would have cut across larger neighbourhood boundaries, rather than forming clean subdivisions of these greater social districts.

One Minute Network

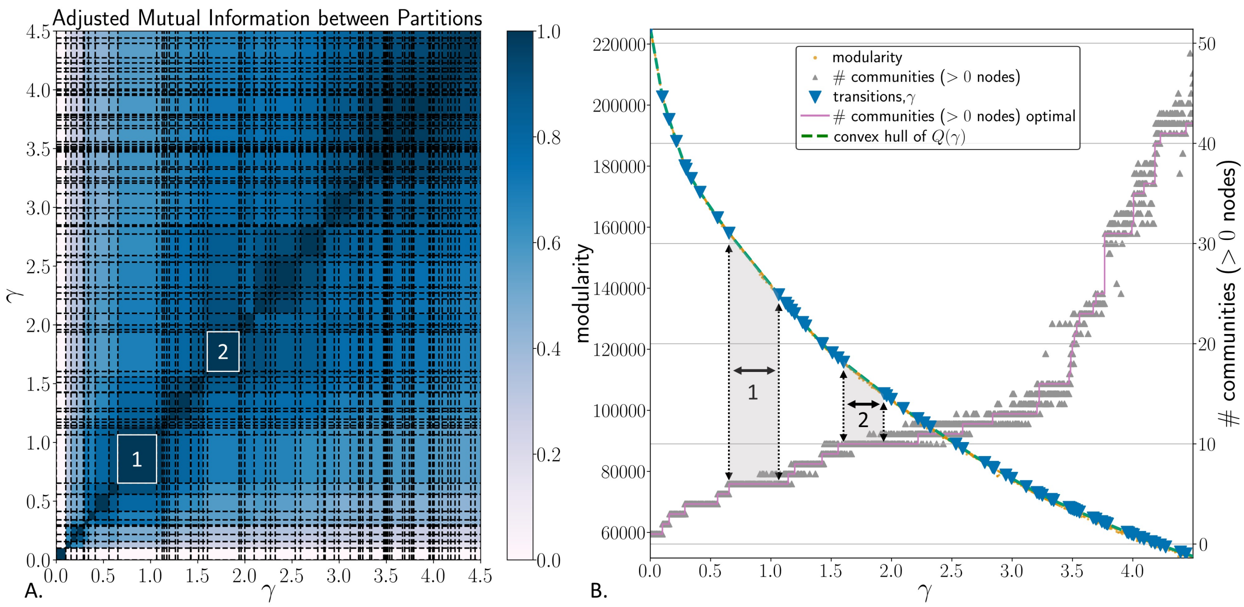


Figure 9 – One Minute Network. Results of CHAMP using Leiden algorithm with modularity. 5000 intervals between γ 0 and 4.5 at 25 iterations each. **A)** AMI of the subset of partitions identified by CHAMP. The widest domain (no. 1: γ 0.65 – 1.06, 6 neighbourhoods) and second widest domain (no. 2: γ 1.6 – 1.94, 10 neighbourhoods) are outlined in white. **B)** Modularity versus γ. Grey upward triangles represent a random sampling of 2000 partitions plotted against their number of communities, while the solid pink stepped line shows the number of communities in the optimal partition for the range of γ. The dashed green line is the convex hull. Blue downward triangles show the transition points between various ranges of γ. The grey zones between the dotted lines identify the widest domains (no. 1: 6 neighbourhoods, no. 2: 10 neighbourhoods).

The one minute network is the densest and most complex of the three. This suffered from a great deal of stochasticity running Leiden independently, but CHAMP returned 73 partitions out of over 1800 unique results. The widest domain is a clearly delineated partition of 6 communities (fig. 9). One would anticipate a greater degree of integration across the city with the longer time threshold, and these results confirm that hypothesis. These 6 divisions largely reflect the topographical divide formed by the forum and the Via del Vesuvio / Stabiana corridor, which bisects the city north to south (fig 10a). The orange neighbourhood, however, constitutes an important bridge connecting east and west. In terms of belonging degree, the lowest scoring units are as usual found near the edges, but the centre of the city again represents the area with the lowest community cohesiveness (fig 10b). Nonetheless, the preponderance of blue nodes (100% belonging) indicates these are durable neighbourhoods in which individuals were unlikely to cross paths with outsiders.

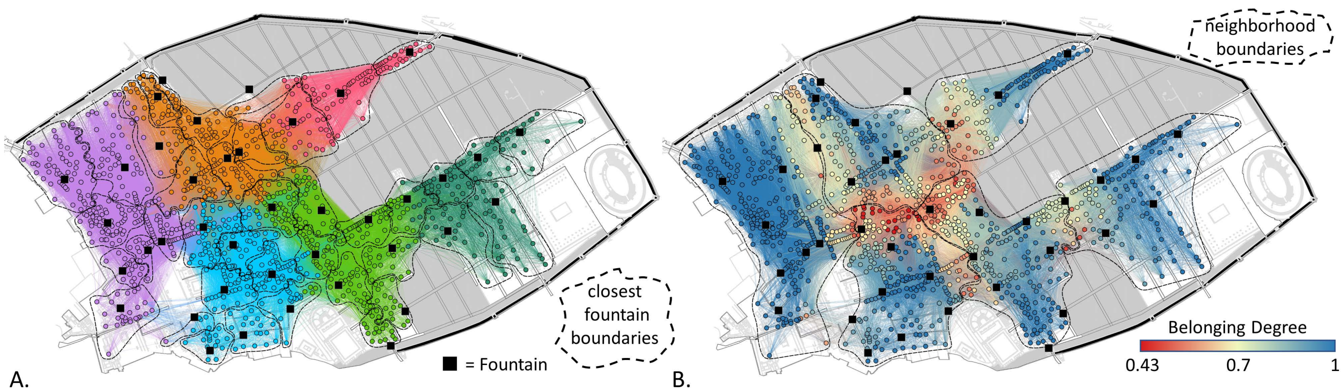


Figure 10 – One Minute Network. Partition into 6 neighbourhoods as identified by the widest γ domain identified by CHAMP. **A)** Colours represent neighbourhoods as partitioned by CHAMP. Dashed lines outline zones defined by the single closest fountain to each unit in the spatial network. **B)** Colour ramp represents each node's belonging degree. Dashed lines outline the neighbourhood boundaries as partitioned by CHAMP.

The second widest domain identified by CHAMP is a partition of 10 communities, still representing rather large subdivisions of the city, but a more modular structure than the top result (table 3).

Table 3 – Contingency table. One Minute Network. Widest domain (6 neighbourhoods, lettered) compared with second widest domain (10 neighbourhoods, numbered).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**All**
A	0	0	144	144	0	0	17	0	0	3	308
B	0	43	0	0	0	119	0	0	90	0	252
C	196	0	0	0	0	10	8	0	1	31	246
D	0	0	0	0	130	0	88	0	0	0	218
E	0	126	0	0	0	0	0	0	0	0	126
F	0	0	0	0	0	0	0	99	0	0	99
All	196	169	144	144	130	129	113	99	91	34	1249

Comparing this against the 6-community structure, we can again see that some smaller neighbourhoods bisect the larger divisions. Here, number 7 in the 10-neighourhood partition is formed by large groups of units from two of the 6-neighbourhood partition communities (A and D), as well as 8 nodes from C. Numbers 3 and 4 in the 10-neighbourhood partition, on the other hand, are entirely subdivisions of the large neighbourhood A. This reminds us that some larger neighbourhoods remained internally closed, even at more individual, face-to-face, scales, while others became mixed at these smaller scales.

Conclusion

This brief discussion only touches upon the variety of analysis possible using community detection to partition archaeological neighbourhoods. By identifying multiple dominant partitions across different ranges of γ, CHAMP offers a multiscalar approach to network community detection within single networks. This case study, however, already considers different network scales using time thresholds. One could also directly compare the top partitions in each of the three networks.

Another useful measure of neighbourhood integration is the intra-cluster density, that is the total density of node ties within a community. Since ties represent units that used a common fountain, the greater the density, the higher the probability that inhabitants met face-to-face. Larger neighbourhoods have lower intra-cluster densities, meaning neighbours were far less likely to meet face-to-face at a fountain even though they shared more social connection in general compared to other regions of the city. Few neighbourhoods in the one minute network's 6-community partition rise above 50%. In such communities, ties between neighbours must have been more dispersed, relying upon mutual contacts. Intra-cluster densities approaching 100%, however, would mean that inhabitants of almost every housing unit had an equal chance of rubbing shoulders at a fountain. These would be the most cohesive and perhaps the most bonded neighbourhoods, with almost daily direct contact between households. 28 out of the 34 neighbourhoods in the closest by door network achieve 100%, but not all. The remaining six range from 52 to 74%.

It should be emphasized that there is no single "correct" community partition. Rather, utilizing the resolution parameter and examining its output in CHAMP offers a multiscalar method of assessing graph partitioning. The choice of scale should be determined with respect to the data and the research questions. CHAMP's utility lies in its ability to narrow down the range of potential options and identify larger patterns across a range of γ, rather than relying upon an arbitrary choice of partition returned by the default algorithm settings or a single iteration.

As a tool for identifying urban neighbourhoods, this approach holds much potential. A more complex Pompeian network that incorporates additional variables of social integration beyond fountains, such as bakeries, bars, communal shrines, and baths, would constitute a more multifaceted social network. Such dense networks would be nearly impossible to partition using topology alone. Thankfully, the algorithms outlined here can efficiently handle extremely large and convoluted networks, returning manageable but statistically solid outputs suitable for detailed analysis and interpretation.

Acknowledgements

I give my sincere thanks to Peter Mucha for discussing aspects of this study, especially the applicability of CHAMP. I also thank Eric Poehler and the Pompeii Bibliography and Mapping Project for the spatial data underlying this project. The three reviewers provided helpful feedback that greatly improved the accessibility of the final draft. Any remaining errors are, of course, my own.

Funding

No additional funding sources were used for this project.

Data, scripts, code, and supplementary information availability

All of the data and scripts used in this article are available online at https://zenodo.org/badge/latestdoi/686189050.

Conflict of interest disclosure

The author declares that he complies with the PCI rule of having no financial conflicts of interest in relation to the content of the article.

References

Bastian M, Heymann S, Jacomy M (2009) *Gephi: an open source software for exploring and manipulating networks.* International AAAI Conference on Weblogs and Social Media. https://doi.org/10.1609/icwsm.v3i1.13937

Blondel V, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, **10**, P10008.

Csárdi G, Nepusz T, Traag V, Horvát S, Zanini F, Noom D, Müller K (2023). *igraph: Network Analysis and Visualization in R*. https://doi.org/10.5281/zenodo.7682609, R package version 1.5.0, https://CRAN.R-project.org/package=igraph.

Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D (2012). qgraph: Network Visualizations of Relationships in Psychometric Data. *Journal of Statistical Software*, **48**, 4, 1–18. https://doi.org/10.18637/jss.v048.i04

Haug, A, Hielscher, A, Krüger, A-L (2023) *Neighbourhoods and City Quarters in Antiquity*. de Gruyter.

Laurence R (2007) *Roman Pompeii: Space and Society*. 2nd ed. Routledge.

Mazzucato C (2019) Socio-Material Archaeological Networks at Çatalhöyük a Community Detection Approach. *Frontiers in Digital Humanities*, **6**, 8. https://doi.org/10.3389/fdigh.2019.00008

Newman M.E.J, Girvan M. (2004) Finding and evaluating community structure in networks. *Physical Review E* ***69***, 026113. https://doi.org/10.1103/PhysRevE.69.026113

Notarian M (2023) A Spatial Network Analysis of Water Distribution from Public Fountains in Pompeii. *American Journal of Archaeology*, **127**, 1, 85-118. https://doi.org/10.1086/722233

Pacifico D, Truex L, eds. (2019a) Special Issue: Excavating Neighborhoods: A Cross-Cultural Exploration. *Archaeological Papers of the American Anthropological Association*, **30**, 1 https://doi.org/10.1111/apaa.12108

Pacifico D, Truex L (2019b) Why Neighborhoods? The Neighborhood in Archaeological Theory and Practice. *Archeological Papers of the American Anthropological Association*, **30**, 1, 5-19. https://doi.org/10.1111/apaa.12110

Peeples M, Bischoff R. (2023) Archaeological networks, community detection, and critical scales of interaction in the U.S. Southwest/Mexican Northwest. *Journal of Anthropological Archaeology*, **70**, 101511. https://doi.org/10.1016/j.jaa.2023.101511

Poorthuis A. (2018) How to Draw a Neighborhood? The Potential of Big Data, Regionalization, and Community Detection for Understanding the Heterogeneous Nature of Urban Neighborhoods. *Geographical Analysis*, **50**, 182-203. https://doi.org/10.1111/gean.12143

Scholnick JB, Munson JL, Macri MJ (2013) Positioning Power in a Multi-relational Framework: A Social Network Analysis of Classical Maya Political Rhetoric. In *Network Analysis in Archaeology: New Approaches to Regional Interaction*, edited by Knappett C, Oxford University Press, 95-124.

Shai S, Stanley N, Granell C, Taylor D, Mucha P. (2020) Case Studies in Network Community Detection. In *The Oxford Handbook of Social Network*, edited by Light R, Moody J, Oxford University Press, 311-333. https://doi.org/10.1093/oxfordhb/9780190251765.013.16

Smith M (2010) The archaeological study of neighborhoods and districts in ancient cities. *Journal of Anthropological Archaeology*, **29**, 137-54. https://doi.org/10.1016/j.jaa.2010.01.001

Thompson AE, Walden JP, Chase ASZ, Hutson SR, Marken DB, Cap B, et al. (2022) Ancient Lowland Maya neighborhoods: Average Nearest Neighbor analysis and kernel density models, environments, and urban scale. *PLoS ONE* **17**, 11, e0275916. https://doi.org/10.1371/journal.pone.0275916

Traag V, Waltman L, van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports* **9**. https://doi.org/10.1038/s41598-019-41695-z

Traag V, Van Dooren P, Nesterov Y (2011). Narrow scope for resolution-limit-free community detection. *Physical Review E*, **84**, 016114. https://doi.org/10.1103/PhysRevE.84.016114

Visser, RM (2021) Dendrochronological Provenance Patterns. Network Analysis of Tree-Ring Material Reveals Spatial and Economic Relations of Roman Timber in the Continental North-Western Provinces. *Journal of Computer Applications in Archaeology,* **4**, 1, 230–253. https://doi.org/10.5334/jcaa.79

Weir W, Emmons S, Gibson S, Taylor D, Mucha PJ (2017) Post-Processing Partitions to Identify Domains of Modularity Optimization, *Algorithms* **10**, 3, 93. https://doi.org/10.3390/a10030093