Data Exploration with GIS Viewsheds and Social Network Analysis

Giles Oatley, Tom Crick and Ray Howell

Department of Computing, Cardiff Metropolitan University, UK
Faculty of Business and Society, University of South Wales, UK

KEYWORDS: Geographic networks, archaeological nodes, viewshed analysis, data mining, social network analysis

Abstract

We present a novel exploratory method that combines line of sight visibility (viewshed analysis) with techniques from social network analysis to investigate archaeological data. At increasing distances different nodes are connected creating a set of networks, which are subsequently described using centrality measures and clustering coefficients. Networks with significant properties are examined in more detail. We use this method to investigate the placement of hillforts (nodes) in the Gwent region of south-east Wales, UK. We are able to determine distances that support significant transitions in network structure that could have archaeological validity.

1 Introduction

The types of technologies utilised in knowledge discovery from databases and data mining develops as opportunities are presented by new datasets. Our study uses both geographical and graph/network structures, and presents an exploratory methodology within which to discover significant distances underlying network creation. In this respect the approach has a more general use than archaeological informatics, for instance neural architectures, transportation networks, and other forms of geographical networks.

We develop connectivity between Iron Age hillforts based on *viewsheds* and an increasing distance threshold. A viewshed is the area of land that is within *line of sight* from a fixed viewing position. We analyse the generated set of networks of connected hillforts using social network analysis, and use the metrics to inform theories of possible use and communication between hillforts.

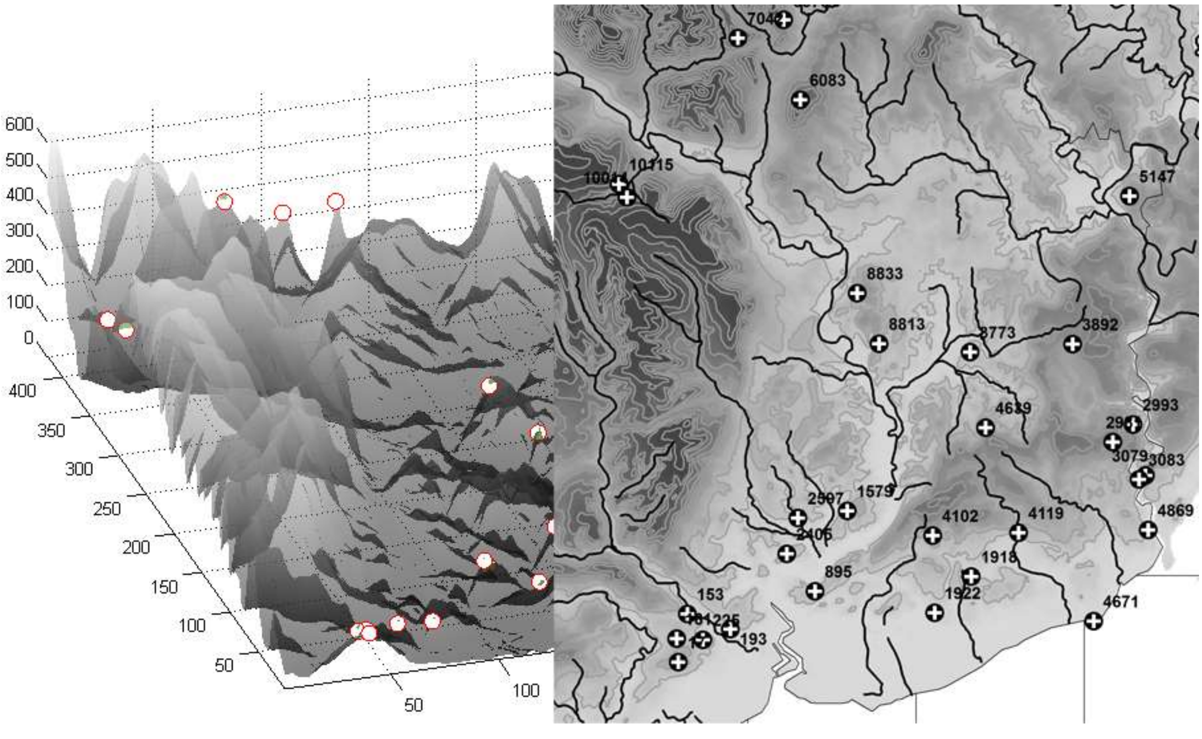

Figure 1 Hillforts in south-east Wales. Hillforts are displayed as white crosses on the front contour display. The same terrain and hillforts (white circles) are displayed behind on a Digital Elevation Model (DEM). The DEM display shows that there are many other sites that could have been used for placement of hillforts.

Our methodology is applied to the area of the Iron Age tribe known as the Silures in south-east Wales, UK. The Silures have been described as a 'resilient and sophisticated clan based tribal confederation' [6, p.113] and it seems reasonable to explore the extent to which spatial relationships and other landscape factors may help to shed light on the degree to which that description is appropriate. Our preliminary investigation focuses on the Gwent region with a study area which roughly approximates the county as constituted between 1974 and 1996. Figure 1 shows the placement of 30 hillforts in this region. This focus seems appropriate as it relates to the early medieval kingdom of Gwent which derived its name from Caerwent, the *civitas* capital of the Silures, and which may have seen considerable cultural continuity from the Iron Age [6].

The data used includes the Iron Age hillfort data, provided from HMR[1], and a Digital Elevation Model (DEM) based on the Shuttle

[1] Historic Monument Record

Radar Topography Mission data (UK SRTM DEM[2]) with 90m horizontal resolution.

2 Spatial and graph based data

Within data mining exist the fields of graph-based and spatial-based data mining. Graph-based data mining (e.g. [2]) has a close cousin in the long established field of social network analysis (SNA), a set of metrics that operates over graphs (networks) created from links [16]. Metrics include those to find clusters within networks, to find points that have significant properties, for instance how central a point is. Spatial data mining likewise has an extensive history (e.g. [9]), and is the discovery of interesting patterns from spatial datasets.

This research lies in the intersection of these two forms of data, spatial and network/graphs. Related work includes that of the physics literature on geographical networks (e.g. [1]), architectural analysis and the isovist literature including its *visibility graphs* [14,8,15], and the authors own work incorporating a kernel density

---

[2] UK_SRTM_DEM created by Addy Pope. Spatial Reference System - Great Britain National Grid. Available at: http://edina.ac.uk/projects/sharegeo/

estimation weighting factor into the *betweenness* social network metric [10].

We use MATLAB's *los2*/*viewshed* functions and the viewshed functionality in ESRI ArcGIS's *Spatial Analyst* toolbox for our geographical algorithms and our SNA measures include both centrality measures (degree, closeness, betweenness) and clustering coefficients.

Degree centrality is simplest and is a count of the number of links to other nodes in the network. Closeness [12] however is a measure of how close a node is to all other nodes in a network. It is the mean of the shortest paths between a node and all other nodes reachable from it. Betweenness [4] is the extent to which a node lies between other nodes in the network and is equal to the number of shortest paths from all nodes to all others that pass through that node. This measure takes into account the connectivity of the node's neighbors, giving a higher value for nodes which bridge clusters.

The final SNA measure used was a clustering coefficient [17]. This indicates to what extent the nodes in a graph tend to cluster together, with the *local* clustering coefficient quantifying how close a nodes' neighbours are to being a clique (fully connected).

3 Methodology and Results

Viewsheds were generated for each hillfort, in order to determine intervisibility between every hillfort – this is the attribute $\texttt{visible}_{nm}$ in the following algorithm, true if nodes *n* and *m* are intervisible. The distances $\texttt{dist}_{nm}$ between each hillfort were also determined.

We were then able at any given distance threshold to determine which hillforts would be intervisible. We increased the distance value from 0m to 40,000m in steps of 500m.

```
Create all Viewsheds, determine all Intervisibility
Calculate Distances between all nodes
FOR Distance = 0 TO 40000 STEP 500
        FOR EACH node (n)
                FOR EACH node (m)
                        IF (dist_nm < Distance)
                    AND (visible_nm == true)
                                Assert edge_nm
                        END IF
                END FOR
        END FOR
        Create Network_Distance from Edges
        Calculate Network_Distance Measures
```

```
        Analyse Network_Distance Measures
END FOR
```

In this way we can investigate networks of hillforts at different distance values (in this case, 80 networks). Figure 2 shows the number of hillforts (nodes) and visibility ‘connections’ between hillforts (edges) plotted against distance, with Figure 3 showing the clustering coefficient and betweenness measure also plotted against distance.

There are several interesting transition points in Figures 2 and 3. For instance: 25/30 nodes (hillforts) are connected at a distance of 5km; 27/30 nodes are connected at 10km distance; and, the maximum number of nodes are connected between 15km and 20km. The clustering coefficient also shows interesting peaks around 15km and 20km. Figure 4 therefore shows clear transition distances of 5-10km, 10-15km and 15-20km. Based on these observations of the 80 generated networks we decided to investigate further the shape of the networks at distances of 5km, 10km, 15km and 20 km, which can be seen in the following Figure 4.
From Figure 4, network 1 (5km) shows localised clusters evident, network 2 (10km) larger regions are connected, with connectivity along the shore and also up waterways. Network 3 (15km) has the greatest diameter of network achieved. Network 4 (20km), at this network distance 5147 has finally joined the network (at 17km).

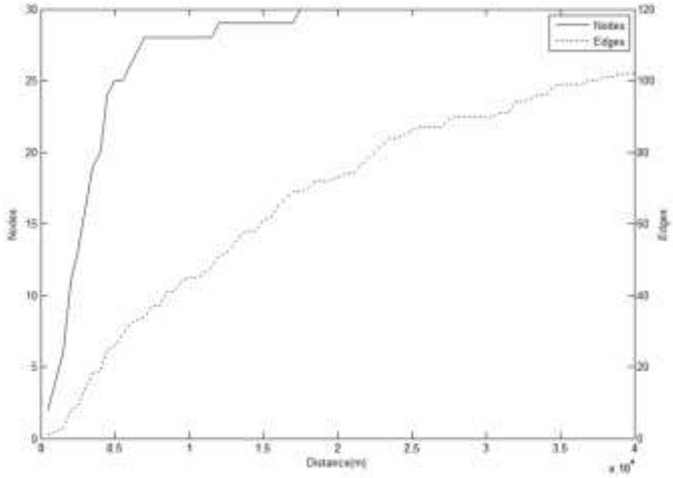

Figure 2 Nodes and edges plotted against distance. These represent how many hillforts are connected to any other hillforts (in any cluster) at the stated distance.

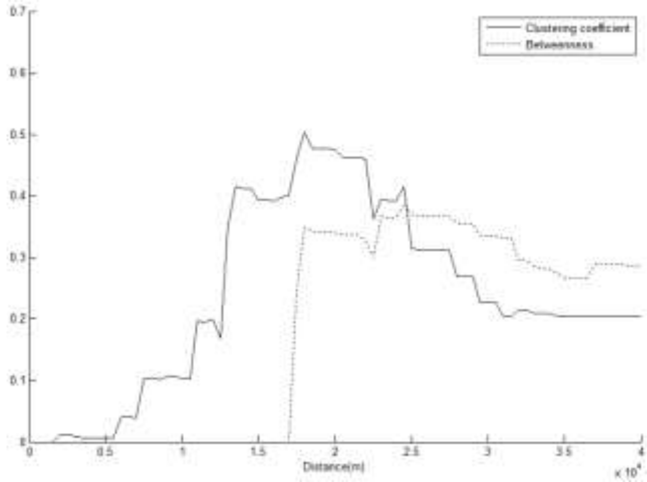

Figure 3 Clustering coefficient and Betweenness.

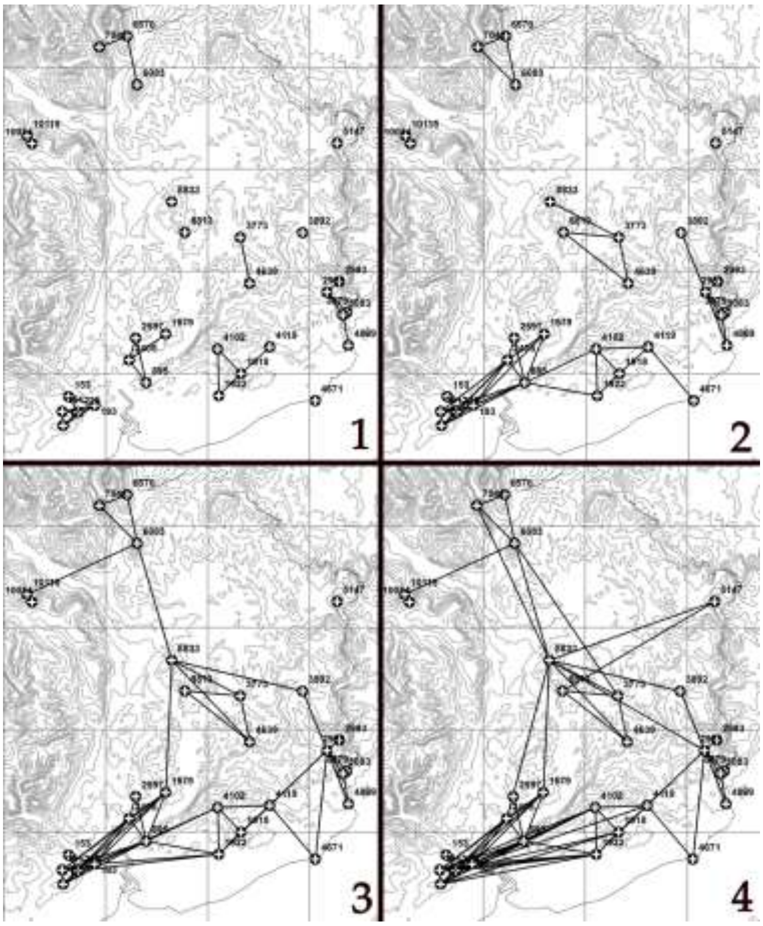

Figure 4 Interesting networks. 1: 5km. 2: 10km. 3: 15km. 4: 20km.

We examine the centrality of individual nodes (hillforts) in these networks with the most significant values are highlighted in Table 1.

Recalling the definitions of these measures, that closeness is a measure of how close a node is to all other nodes in a network, and betweenness is a more useful measure of the node's importance to the network than just connectivity, we make the following observations:

Table 1 Betweenness (Bet), Closeness (Clo), Degree (Deg) for interesting networks.

	1		2		3			4		
Hillfort	Clo	Deg	Clo	Deg	Bet	Clo	Deg	Bet	Clo	Deg
17	0.13	**0.21**	0.24	0.34	0	0.35	0.41	0.01	0.42	**0.59**
161	0.13	**0.21**	0.24	0.34	0	0.35	0.41	0.01	0.41	0.55
193	**0.17**	**0.24**	**0.26**	0.34	0.07	**0.40**	0.45	0.03	0.42	0.45
225	**0.17**	**0.24**	0.25	0.34	0.12	**0.40**	**0.55**	0.09	0.44	**0.62**
895	0.10	0.14	**0.34**	**0.48**	0.09	**0.40**	**0.48**	0.03	0.42	0.48
1579	0.08	0.03	0.22	0.14	0.25	**0.40**	0.41	0.13	0.48	0.41
2405	0.13	0.14	**0.26**	**0.38**	0.01	0.35	0.38	0.10	0.48	0.41
2981	0.13	0.17	0.20	0.28	**0.33**	**0.40**	0.41	**0.37**	**0.52**	0.45
4102	0.10	0.07	**0.26**	0.14	0.06	0.35	0.24	0.01	0.37	0.38
8833	0.00	0.00	0.08	0.03	**0.45**	**0.42**	0.24	**0.51**	**0.55**	0.41

Isolated clique

Nodes join previously separated cliques

Nodes stretching diameter

Still important gatekeeper

Nodes from isolated cliques visible to larger distances

- Consistently #225 has a consistently high centrality value.
- #17, #153, #161, #193 also have high centrality values, and with #225 form a highly connected cluster with small diameter (5km).

- In networks 3 and 4, low degree value and high centrality indicate an important role as ‘gatekeeping’. #8833 and #2981 are nodes with these properties.
- #895 has mid-range high centrality values.

4 Discussion

Among preliminary conclusions arising from this first phase of investigation is that the methodology employed can effectively inform our understanding of Iron Age social structures. For example, viewshed analysis confirms clustering of hillforts in the region and it seems reasonable that this clustering arises, at least in part, from the existence of clans. Moreover, there was clearly extensive line of sight communication, not only within clusters, but also with other hillfort groupings. The model of a clan based confederation with regional emphasis, and possibly variation, but with wider connectivity sufficient to allow the cohesion necessary to have resisted the Roman advance so effectively seems wholly appropriate.

However, the analysis is confounded by the function of hillforts generally. It seems likely that the roles of hillforts have been varied through time and that hillforts could have had defensive, ritual, trade or other functions, e.g. [13] proposed division of hillforts into those for communication, those for watching rivers and land routes, and those for close living.

In order to understand better the value of this approach, which admittedly has been used on a small dataset, we need to consider problems with modern viewshed analysis and its interpretation.

The technique has long been used in archaeological informatics. Its simplest model, a binary viewshed, assumes perfect clarity and perfect visibility and consists solely of 1s and 0s, representing cells which are either visible from the viewing location or not visible from the viewing location. Of course this approach assumes the artificiality of the ‘infinite’ view [18].

Fisher [3] first proposed the fuzzy viewshed, later expanded upon by [11], whereby a distance decay function is introduced into the standard binary viewshed. This function models the drop in visual clarity that occurs with increasing distance from the viewpoint, incorporating fixed values for the maximum distance from viewpoint of clear visibility, and the distance from viewpoint at which visibility drops to 50%. Ogburn [11] proposes different values for the latter, using 4.75 km, 7.6 km and 16.2 km. Note that these values are decided *a priori*.

Jones [7] investigated the limit of normal human vision, or the distance at which an object could be recognized by some people under very favourable conditions, is reached when an object subtends 1’ of visual arc, or 3440m [11]. However a target 5m wide would subtend this arc at a distance of 17.2 km.

Higuchi [5] also includes this human aspect, developing the concept of three levels of perceptive visibility based upon the visual characteristics of a tree as they related to the distance of that tree from an observer. His short-distance view includes those features which are integral and immediate to the everyday lives of the people. The middle-distance view is the scenic landscape setting, giving context of meaning for a given locale. The long-distance view is the background, where features may be visible but are not readily identifiable. These ranges are again decided *a priori*.

Certainly it is important to determine what it is that is important to be seen, while Higuchi uses trees as his factors, we suspect that people, livestock and smoke plumes might be more significant factors for our study. However the important point to note is that in this study we are not interested in correctly determining *a priori* distances and decay values, although we can draw upon these studies to help us understand our significant networks and distance values discovered through our analysis. We are interested in discovering interesting patterns and clusters and then investigating them *a posteriori* for (archaeological) validity.

Acknowledgements

We are grateful for access to the Historic Monument Record maintained by the Glamorgan Gwent Archaeological Trust.

References

1. ben-Avraham, D. Rozenfeld, A.F. Cohen, R., and Havlin, S., 2003. *Geographical Embedding of Scale-Free Networks*. *Physica A* **330** , 107-116 (2003).

2. Cook D. J. & Holder, L. B., 2006. Mining Graph Data, Wiley-Interscience, ISBN-10: 0471731900, ISBN-13: 978-0471731900

3. Fisher, P. F., 1994. Probable and fuzzy models of the viewshed operation. Innovations in GIS: selected papers from the First National Conference on GIS Research UK. M. F. Worboys. London, UK, Taylor and Francis**:** 161-175.
4. Freeman, L. C., 1977. A set of measures of centrality based on betweenness. *Sociometry* **40**, 35-41.
5. Higuchi, T., 1983. The Visual and Spatial Structure of Landscapes. Cambridge, MA, The MIT Press.
6. Howell, R., 2009, *Searching for the Silures, an Iron Age tribe in south-east Wales*, Stroud, The History Press.
7. Jones L.A., 1920. A method and instrument for the measurement of the visibility of objects. Journal of the Franklin Institute, Volume 190, Issue 2, August 1920, Page 259, doi:10.1016/S0016-0032(20)92055-0.
8. Llobera M, 1996. Exploring the topography of mind: GIS, social space and archaeology. Antiquity 70 612 - 622
9. Lu, W. Han J. and Ooi, B., 1993. Discovery of General Knowledge in Large Spatial Databases, Proc. Far East Workshop on Geographic Information Systems, Singapore, pp. 275-289.
10. Oatley, G.C., McGarry, K. and Ewart, B.W., 2006. Offender network metrics. WSEAS Transactions on Information Science & Applications, 3 (12). pp. 2440-2448.
11. Ogburn, D. E., 2006. Assessing the Level of Visibility of Cultural Objects in Past Landscapes. Journal of Archaeological Science 33: 405-413.
12. Sabidussi, G., 1966. The centrality index of a graph. Psychometrika, 31 (4), 581-603.

13. Stancic Z., & Veljanovski T., 2000. Understanding Roman settlement patterns through multivariate statistics and predictive modelling. In: G. Lock (Ed.) Beyond the Map, IOS Press 2000
14. Steadman P, 1973. Graph theoretic representation of architectural arrangement. Architectural Research and Teaching 2 161 - 172
15. Turner A., Doxa M., O'Sullivan D., & Penn A., 2001. From isovists to visibility graphs: a methodology for the analysis of architectural space. Environment and Planning B: Planning and Design 2001, volume 28, pages 103 – 121
16. Wasserman, S. & Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.

17. Watts, D.J.; Strogatz, S.H. (1998). "Collective dynamics of 'small-world' networks.". *Nature* **393** (6684): 409–10.
18. Wheatley D., 1995. Cumulative viewshed analysis: a GIS-based method for investigating intervisibility, and its archaeological application. In (G. Lock & z. Stancic, Eds) Archaeology and Geographical Information Systems. London: Taylor and Francis, pp. 171-186