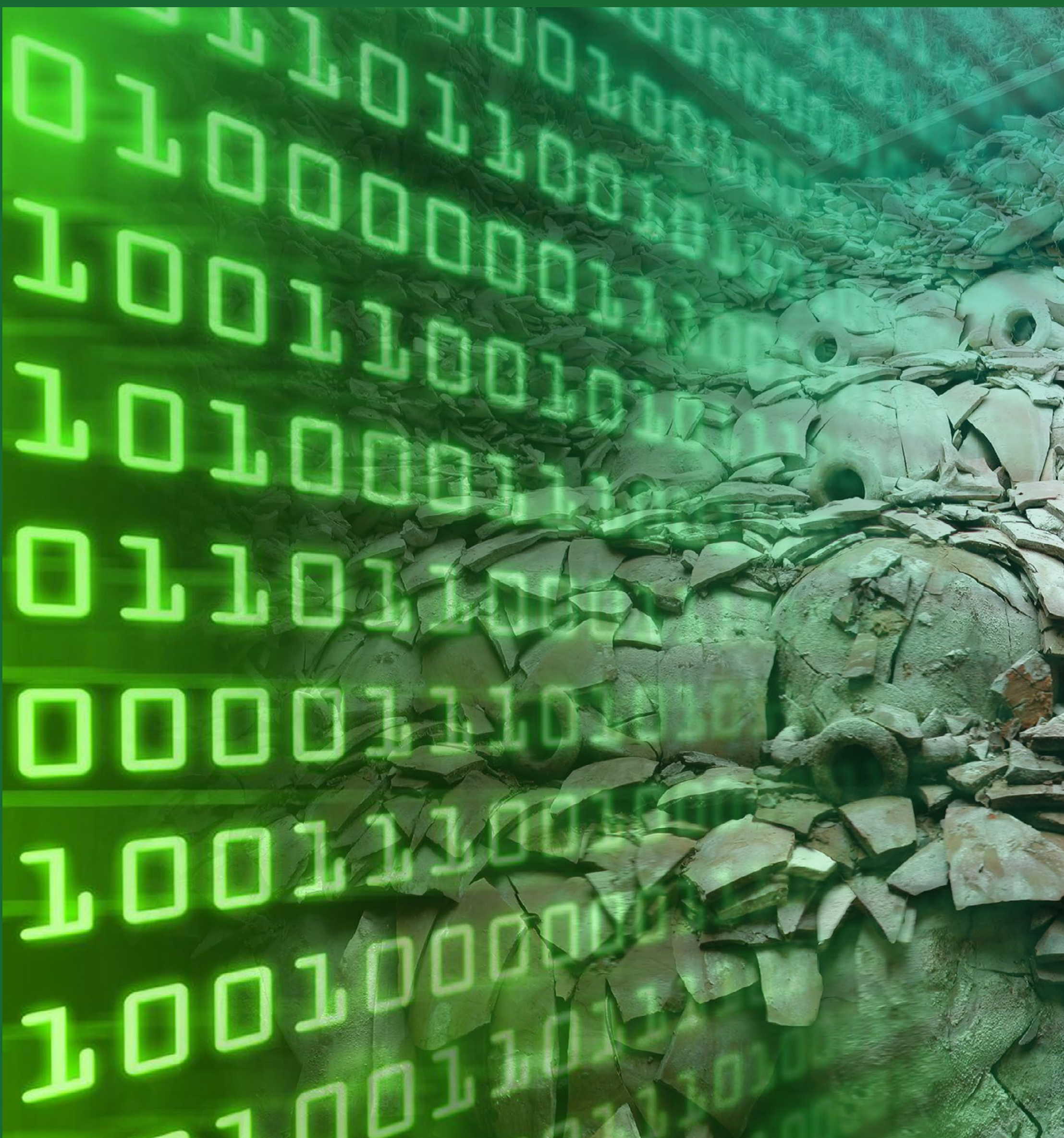
Arqueología y *Téchne*
Métodos formales, nuevos enfoques

Archaeology and *Techne*
Formal methods, new approaches

Editado por
José Remesal Rodríguez
Jordi Pérez González



Access Archaeology

Archaeopress Access Archaeology

About Access Archaeology

Access Archaeology offers a different publishing model for specialist academic material that might traditionally prove commercially unviable, perhaps due to its sheer extent or volume of colour content, or simply due to its relatively niche field of interest. This could apply, for example, to a PhD dissertation or a catalogue of archaeological data.

All *Access Archaeology* publications are available as a free-to-download pdf eBook and in print format. The free pdf download model supports dissemination in areas of the world where budgets are more severely limited, and also allows individual academics from all over the world the opportunity to access the material privately, rather than relying solely on their university or public library. Print copies, nevertheless, remain available to individuals and institutions who need or prefer them.

The material is refereed and/or peer reviewed. Copy-editing takes place prior to submission of the work for publication and is the responsibility of the author. Academics who are able to supply print-ready material are not charged any fee to publish (including making the material available as a free-to-download pdf). In some instances the material is type-set in-house and in these cases a small charge is passed on for layout work.

Our principal effort goes into promoting the material, both the free-to-download pdf and print edition, where *Access Archaeology* books get the same level of attention as all of our publications which are marketed through e-alerts, print catalogues, displays at academic conferences, and are supported by professional distribution worldwide.

The free pdf download allows for greater dissemination of academic work than traditional print models could ever hope to support. It is common for a free-to-download pdf to be downloaded hundreds or sometimes thousands of times when it first appears on our website. Print sales of such specialist material would take years to match this figure, if indeed they ever would.

This model may well evolve over time, but its ambition will always remain to publish archaeological material that would prove commercially unviable in traditional publishing models, without passing the expense on to the academic (author or reader).

Archaeopress Access Archaeology

Arqueología y *Téchne*
Métodos formales, nuevos enfoques

Archaeology and *Techne*
Formal methods, new approaches

Editado por

José Remesal Rodríguez

Jordi Pérez González

Access Archaeology

Archaeopress Access Archaeology

Archaeopress Publishing Ltd
Summertown Pavilion
18-24 Middle Way
Summertown
Oxford OX2 7LG
www.archaeopress.com

ISBN 978-1-80327-181-1
ISBN 978-1-80327-182-8 (e-Pdf)

Cover image: Muro de ánforas de la campaña arqueológica del año 2002 en el Monte Testaccio (Roma). Directores de la excavación: J.M.ª Blázquez Martínez y José Remesal Rodríguez. Fotografía y edición gráfica de la cubierta: Juan Moros Díaz (CEIPAC). *Arqueología y* Téchne*. Métodos formales, nuevos enfoques / Archaeology and* Techne*. Formal methods, new approaches* has been funded within the research projects "Economic and Political Networks (EPNet Project), ERC-2013-ADG-340828 (http://www.roman-ep.net); Relaciones Interprovinciales en el Imperio Romano. Producción y comercio de alimentos hispanos (*Provinciae Baetica et Tarraconensis*) (HAR2017-85635-P; AEI/FEDER, UE); Centro para el Estudio de la Interdependencia Provincial en la Antigüedad Clásica (CEIPAC) (2017 SGR 512); Dinàmiques socioeconòmiques del món rural romà: formes d'hàbitat i cultura material al litoral central català (CLT009/18/00045). Jordi Pérez González funded by Juan de la Cierva-Formación-2019 (Agencia Estatal de Investigación (Ministerio de Ciencia e Innovación) REF: FJC2019-040688-I.

Contents

From Counting Pots to Reconstructing Economy: Computational Tools Developed in the EPNet Project

Iza Romanowska

Simon Carrignon

Maria Coto-Sarmiento

Jean Marc Montanier

Xavier Rubio-Campillo

1. Introduction

Much of the scientific progress can be attributed to new inventions and development of tools and methods for doing science.[1] Regardless of the quality of data, neatness of ideas or creativity of any given researcher little can be achieved without the right tools to do the task. In the past half century much of the new methodological development focused on harnessing the continuously increasing power of computers. Computational methods have long been a staple of many archaeological projects, but in the case of the EPNet projects they became one of the central pivots on which the project aims depended on.[2]

To this end, during the lifetime of the EPNet project the BSC team has focused their efforts on providing research software and bespoke computational tools to enable deep analysis of the archaeological data and to test hypotheses against detected patterns. By foregoing point-and-click tools and developing solutions in computer code (e.g., in R, Python or C++) the EPNet project has produced a substantial library of research software. The shift towards performing computational analysis using programming languages presents two major advantages - scientific transparency and reusability. The scientific transparency term refers to the fact that all research design decisions and every step in data manipulation are documented and open for scholarly scrutiny and replication. The reusability of code means that the initial investment in developing the software tools and pipelines can now benefit the whole archaeological community as code developed for one project can be directly used or modified and applied to other research projects. The research software tools developed as part of the EPNet range from relatively simple data analysis pipelines to complex simulation frameworks optimised for High Performance Computing environment. Here we will provide a concise overview of each tool and the range of potential applications. They will be grouped into broad three categories: tools for applying similarity measures, pipelines for applying Bayesian inference and tools related to agent-based models.

2. Similarity measures

Comparing two sets of values is a common task in archaeology and a variety of statistical methods have been employed in the discipline to that end. Refactoring them into data analysis scripts allows for quick comparisons of dataset's elements even for large data compilations.

[1] Galison 1997.
[2] Remesal *et al.* 2014.

Jaccard similarity coefficients is a commonly applied measure of distance between two sets that denotes the ratio of intersection and union of the sets. In simple terms, it is calculated as a percentage of the common elements compared to the total number of unique elements of both sets. For example,

A = [a, b, c, d]
B = [a, b, e, g]
Intersection = [a, b] = 2
Union = [a, b, c, d, e, g] = 6
Jaccard Distance = 2/6 = 0.33333...

Jaccard distance provides the most basic measure of similarity between two sets, in which the order of elements is not important (such as archaeological assemblages), focused on present/absence of common elements. Thus, it shows to what extent the two sets (for example, archaeological assemblages from two sites) resemble each other in terms of the finds. Although there is no one-to-one relationship, similarity of assemblages is usually interpreted as a result of similar socio-economic processes. For example, a similarity between two Roman military camps in terms of pottery stamps found at them would be a strong indication that they both received, say, olive olive as part of the same supply line. The code developed in the EPNet project calculates the Jaccard distance of all pairs of datasets and creates a matrix of similarity. This matrix is then further analysed using Multi-Response Permutation Procedure (MRPP) and a clustering algorithm *neighbour joining.* The code can be found here: https://github.com/xrubio/ecologyStamps. It was used to analyse the amphora stamps found in different Roman provinces.[3] Authors have been able to highlight distinctive trade routes and the importance of spatial closeness, military presence, and Atlantic sea routes.

The Morisita-Horn index is in many ways similar to the Jaccard distance as it compares the proportion of elements that are common for two datasets. However, it also takes into account the frequency of each element thus comparing the full distribution of the datasets. In simple terms, while for Jaccard distance it would not make a difference whether element x is present in dataset A five times and in dataset B 10 times (only the present/absence is considered) the Morisita-Horn index will group datasets with similar frequencies of each type closer together. Thus, for datasets with a coherent spatiotemporal range this method may be more meaningful than the simpler Jaccard Distance. The code developed in the EPNet project calculates the Morisita-Horn index for all pairs of sites and creates a matrix of similarity, which is visualised as a dendrogram. The code can be found here: https://github.com/Mcotsar/Edinburgh_research. It was used to analyse the frequency of distribution of Roman amphora stamps from the olive oil production region of Baetica as well as from military camps and civil settlements in the provinces of Britannia and Germania (publication pending).

Edit distance is a measure of the minimum number of changes that one string (a set of characters, such as a word) needs to undergo in order to become another string. The permitted changes are: deletion, insertion and substitution. For example,

A = 'banana'
B = 'panamas'
Edit distance = [substitution: 'b'-> 'p', substitution: 'n->'m', insertion: 's'] = 3

[3] Rubio-Campillo *et al.* 2018.

Edit distance is a good method for comparing strings of different lengths (or in mathematical parlance partially ordered sets)[4] where part of the original string may be missing (the algorithm aligns strings to minimise the number of changes). Thus, it is uniquely well positioned to enable rapid comparison of large quantities of epigraphic data, such as inscriptions, stamps, *tituli picti*, etc. The code developed in the EPNet project, calculates the edit distance of all pairs of strings, creating a two dimensional matrix visualised as a heatmap. This matrix is then used to calculate clusters and create a dendrogram which groups series of inscriptions according to their similarity. The code can be found here: https://github.com/izaromanowska/edit_distance It was used to analyse *tituli picti* inscriptions on P08 amphorae from the city of Pompeii.[5]

3. Bayesian Inference

Bayesian inference is a family of techniques in which the probability of a hypothesis is adjusted according to each new piece of evidence. It is often contrasted with the frequentist approach commonly used in statistics where the truthfulness of a hypothesis is estimated to a given degree of certainty on the basis of the evidence. Although the difference seems subtle using one or the other framework may have profound consequences on the results obtained. Methods based on the Bayes Theorem have proven to be particularly useful in research where there exist multiple overlapping hypotheses and different sources of data with high levels of uncertainty. Both conditions are typically met in archaeological research and for this reason the approach looks particularly promising.

The application of a pure Bayesian hypothesis-testing framework requires advanced mathematical techniques for most case studies and inhibits the use of the framework for very complex models due to the high complexity of a crucial step in the research pipeline: the likelihood function. However, a recent innovation known as ABC (Approximate Bayesian Computation) has allowed to extend the use of Bayesian inference to complex simulations such as Agent-Based Models by using High-Performance Computing.

Bayesian inference can be used to test formally defined hypotheses against the available data - a process commonly referred to as model selection. The posterior distribution of each competing model can be calculated with Markov Chain Monte Carlo Method (MCMC) and assessed in terms of its simplicity and fit to the data with one of the Information Criterion, such as Deviance Information Criterion (DIC). As a result, the researcher obtains the 'score' of how well each model explains the data taking into account both closeness to the data and parsimony (i.e., simpler explanations are favoured over more complicated ones).

The code developed in the EPNet project, formalises four economic hypotheses into statistical models and runs them using MCMC. The code can be found here: https://github.com/xrubio/bayesRome. It was used to quantify the probability of different market structures involved in olive oil trade and to compare each model predictions to a dataset of amphoric stamps.[6]

Approximate Bayesian Computation is another Bayesian inference technique which calculates to what degree any given model predictions are consistent with the available data. Pure Bayesian frameworks

[4] For sets of equal length where the elements are compared pairwise (first with first, second with second...) it is customary to use hamming distance.

[5] Martín-Arroyo Sánchez, Romanowska this volume.

[6] Rubio-Campillo *et al.* 2017.

require the modeller to define the likelihood function as a mathematical function and this step becomes increasingly difficult as a model gets more complex. ABC allows to apply Bayesian framework for these cases by approximating the posterior distribution using computer simulation. Thus, by running thousands of different combination of model parameters it becomes possible to establish which values are most plausible in light of the available evidence, i.e., the archaeological data.

The EPNet project explored the potential of the ABC for inferring historical dynamics with a first simple case study exploring the evolution of Western warfare across 500 years.[7] The lessons from this study were later applied to assess the match between different parameters in an agent-based model of economic interaction. The code developed in the EPNet project distributes different simulation runs on a supercomputer, gather the results and calculate their match to the data and show the posterior distribution of parameters that fit the data best. The code can be found here: https://framagit.org/sc/abc-pandora/. It was used to explore the model described in Carrignon *et al.* (2015) at a theoretical level and then to evaluate the plausibility of different market structures driving craft produce trade in the Roman East.[8]

4. Agent-based Modelling

Agent-based Modelling (ABM) is an increasingly popular simulation technique among social and natural scientists. It has been also gaining ground in archaeology (Lake 2014). Agent-based models consists of individual agents whose simulated interactions with each other and with their environment result in the emergence of population-scale patterns that can be directly compared to archaeological (and other) data.

The major software product that has been developed and maintained by the BSC-CNS team and in particular Xavier Rubio Campillo is **Pandora**[9] - an ABM framework for developing large scale distributed simulations taking advantage of high performance computing facilities such as supercomputers. It enables running much larger and more complex models over a wider parameter space compared to the current level of most archaeological applications. Its functionality extends over GIS integration and simulation data analysis tool - Cassandra. Pandora can be found here: https://github.com/xrubio/pandora.

For anyone who needs to first get in grip with coding agent-based models a series of tutorials was developed by the BSC team. While not assuming any prior knowledge of programming or data science it uses a series of archaeological case studies to gently guide the reader through the first steps in creating agent-based models in popular among social scientists and archaeologists framework: NetLogo. The tutorials and code can be found here: https://github.com/izaromanowska/ABM_tutorials. They can be coupled with the jargon-free introduction to the method and its different aspects published in Romanowska *et al.* 2019, Davies *et al.* 2019 and Crabtree *et al.* 2019.

Finally, the BSC has developed a number of agent-based models which can be used as a basis of more complex simulations or repurposed for other case studies. In particular, the model of cultural interaction between amphorae workshops can be used to to study the impact of cultural change in crafts

[7] Rubio-Campillo 2016, the code can be found here: https://github.com/xrubio/lanchester.
[8] Carrignon *et al.* submitted
[9] Rubio-Campillo 2014.

production on variability among and between the resulting assemblages, of, for examples, pottery. The code in Python can be found here: https://github.com/simoncarrignon/apemcc. It was used to study horizontal and vertical knowledge transmission between pottery workshops in Roman Baetica.[10]

Another agent-based model developed in the BSC-CNS is the *Trade Distance Model*. This abstract agent-based simulation has been developed to study the impact of distance from the production center on the inception curve of a product. Written in NetLogo[11] the code can be found here: https://github.com/izaromanowska/trade-distance-model. It was used to showcase the way in which simple, abstract agent-based model can be used as heuristics to better understand patterns in data related to Roman trade.[12]

Finally, a generalised ABM framework was developed to study different aspects of the interaction between economic process and cultural transmission among traders (Carrignon et al. 2015). This general framework has been successfully applied in the implementation of two models. The authors also compared the qualitative differences obtained by the two implemented models. The source code used can be found here: https://github.com/montanier/CMR-WSC-CoEvolutionTradeCulture. This model was later modified to study craft trade in the Roman East (code: https://framagit.org/sc/ceeculture[13]).

Conclusions

This brief summary of computational tools shows the breath of applications, topics and methods developed and applied by a small research team in the course of one project - the EPNet. From data analysis of complex datasets to modelling theories and testing them against pattern detected in the data - computational tools facilitate and often enable asking questions that could have never been asked before and tackling some of the complexities of the past. Archaeological practice is teeming with data and theories yet tools for efficient analysis of large datasets and applying them to test hypotheses are still few and far between. This is seriously hampering the discipline's efforts to ensure transparency, replicability and cumulative research. Although the calls for moving away from point-and-click software and towards open methods have been voiced before,[14] the current practice is still far from perfect. One of the major hindrances hampering this effort is the limited exchange between researchers and projects forcing everyone to develop their own solutions instead of building upon earlier work. By performing data analysis and formal modelling exclusively in a scripting environment of R, Python and C++ and publishing them open access the EPNet project members hope to support other researchers in archaeology and beyond in their computational work.

[10] Coto Sarmiento *et al. in press.*
[11] Wilensky 1999.
[12] Romanowska 2018.
[13] Carrignon *et al.* submitted.
[14] Marwick *et al.* 2017.

Literature cited

Carrignon *et al.* submitted = Carrignon, S., T. Brughmans and I. Romanowska. submitted. An agent-based modelling and approximate Bayesian Computation exploration of ceramic tableware data and commercial information flows in the Roman East.

Carrignon *et al.* 2015 = Carrignon, S., J.-M. Montanier and X. Rubio-Campillo. 2015. Modelling the Co-evolution of Trade and Culture in Past Societies, in L. Yilmaz, W.K.V. Chan, I. Moon, T.M.K. Roeder, C. Macal and M.D. Rossetti (eds.) *Proceedings of the 2015 Winter Simulation Conference*: 3949-3960. California.

Coto-Sarmiento *et al. in press* = Coto-Sarmiento, M., S. Carrignon, X. Rubio-Campillo and J. Remesal. J. *in press*. An Agent Based Model to Detect Variations in the Transmission of Potters from Baetica Province, in *The Making of Value and the Value of Making, 19th International Congress of Classical Archaeology, 22- 26 Mayo, Cologne/Bonn, 2018.*

Crabtree *et al.* 2018 = Crabtree, S., K. Harris, B.,Davies and I. Romanowska 2019. Outreach in Archaeology with Agent-based modeling A step-by-step guide for using agent-based modeling in archaeological research (Part III of III). *Advances in Archaeological Practice* 7 (2): 194-202. doi:10.1017/aap.2019.4.

Davies *et al.* 2019 = Davies, B., I. Romanowska, K. Harris and S. Crabtree. 2019. Combining Geographic Information Systems and Agent-Based Models in Archaeology: A step-by-step guide for using agent-based modeling in archaeological research (Part II of III). *Advances in Archaeological Practice* 7 (2): 185-193. doi: 10.1017/aap.2019.5

Galison, P. 1997. *Image and Logic. A Material Culture of Microphysics.* Chicago: University of Chicago Press.

Lake, M.W. 2014. Trends in Archaeological Simulation. *Journal of Archaeological Method and Theory*: 258-287. <https://doi.org/10.1007/s10816-013-9188-1>.

Marwick *et al.* 2017 = Marwick, B., J. d'Alpoim. Guedes, C.M. Barton, L.A. Bates, M. Baxter, A. Bevan, E.A. Bollwerk, R.K. Bocinsky, T. Brughmans, A.K. Carter, C. Conrad, D.A. Contreras, S. Costa, E.R. Crema, A. Daggett, B. Davies, B.L. Drake, T.S. Dye, P. France, R. Fullagar, D. Giusti, S. Graham, M.D. Harris, J. Hawks, S. Heath, D. Huffer, E.C. Kansa, S. Whitcher Kansa, M.E. Madsen, J. Melcher, J., Negre, F.D. Neiman, R. Opitz, D.C. Orton, P. Przystupa, M. Raviele, J. Riel-Salvatore, P. Riris, I. Romanowska, J. Smith, N. Strupler, I.I. Ullah, H.G.V. Vlack, N. VanValkenburgh, E.C. Watrall, C. Webster, J. Wells, J., Winters and C.D. Wren. 2017. Open Science in Archaeology. *The SAA Archaeological Record* 17: 8–14. <https://doi.org/10.7183/2326-3768.4.1.41>.

Remesal *et al.* 2014 = Remesal, J., A. Díaz-Guilera, B. Rondelli, X. Rubio-Campillo, A. Aguilera, D.J. Martín-Arroyo, A. Mosca and G. Rull 2014. The EPNet Project. Production and distribution of food during the Roman Empire: Economics and Political Dynamics, in *Information Technologies for Epigraphy and Cultural Heritage. Proceedings of the First EAGLE International Conference*: 455-464. Sapienza Università Editrice. <https://upcommons.upc.edu/handle/2117/85564>.

Romanowska *et al.* 2019 = Romanowska, I., S. Crabtree, B. Davies and K. Harris, K. 2019. Agent-based Modeling for Archaeologists. A step-by-step guide for using agent-based modeling in archaeological research (Part I of III). *Advances in Archaeological Practice* 7 (2): 178-184. <doi:10.1017/aap.2019.6>.

Romanowska, I. 2018. Using Agent-Based Modelling to Infer Economic Processes in the Past, in J. Remesal Rodríguez, V. Revilla Calvo and J.M. Bermúdez Lorenzo (eds.) *Cuantificar las economías antiguas. Problemas y métodos / Quantifying Ancient Economies. Problems and Methodologies* (Instrumenta 60): 107-118. Barcelona: Publicacions de la Universitat de Barcelona.

Rubio-Campillo, X. 2014. Pandora: A Versatile Agent-Based Modelling Platform for Social Simulation. *Proceedings of the 6th International Conference on Advances in System Simulation*, no. c: 29–34.

Rubio-Campillo, X. 2016. Model selection in historical research using approximate Bayesian computation. *PloS one* 11(1). <https://doi.org/10.1371/journal.pone.0146491>.

Rubio-Campillo *et al.* 2017 = Rubio-Campillo, X., M. Coto-Sarmiento, J. Pérez-Gonzalez and J. Remesal Rodríguez. 2017. Bayesian Analysis and Free Market Trade within the Roman Empire.*Antiquity* 91 (359): 1241–52. <https://doi.org/10.15184/aqy.2017.131>.

Rubio-Campillo *et al.* 2018 = Rubio-Campillo, X, J.-M. Montanier, G. Rull, J.M. Bermúdez Lorenzo, J. Moros Díaz, J. Pérez González and J. Remesal Rodríguez 2018. The Ecology of Roman Trade. Reconstructing Provincial Connectivity with Similarity Measures. *Journal of Archaeological Science* 92 (April): 37–47. <https://doi.org/10.1016/j.jas.2018.02.010>.

Wilensky, U. 1999. NetLogo. <http://ccl.northwestern.edu/netlogo/>. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.