Robust Spatial Network Analysis

Tim Evans

Theoretical Physics
and Complexity & Networks groups

doi: [10.6084/m9.figshare.834936](https://doi.org/10.6084/m9.figshare.834936)

Page 1

© Imperial College London

31 Oct – 2 Nov 2013 U.Toronto

Networks of maritime
connectivity in the ancient
Mediterranean

- General Approach to Modelling in Archaeology
- Geography and Interactions
- Comparing Networks
- Robustness

Modelling Scales

Site-Site Interactions

- Archaeology can be “Site Centric”
 - Regional and global interactions hard to consider
- Networks emphasise interactions

Island Archipelagos as an Ideal Network

Vertices = Major Population or Resource Sites

- Sea isolates communities
→ **Natural Vertices**
- Coastal Sites often isolated like islands due to geography and difficulty of ancient land travel

Deducing Interactions

- Artefact counts
 - Measure similarity of sites through counts [Terrell 2010; Sindbæk 2007]
- Texts
 - Appearance of sites in text [Isaksen 2006; “Anskar’s Vita” Sindbæk 2008]
- Geography
 - Direct from geography [Terrell 1977; Irwin 1983; Hage & Harary 1991; Broodbank 2000; Collar 2007; Bevan 2010]

Different Spaces

- I will consider two-dimensional geographical space
- All ideas can be applied to artefact spaces e.g. word frequency space for text similarities

Different Distances

- Physical Distances

- As the crow flies
- Shortest route in km
- Quickest time
- Lowest costs
- ...

- Ranked distances

- Nearest neighbour, second nearest neighbour, etc

Ranked distances

- Used in **PPA** (Proximal Point Analysis)
 - a more sophisticated version in the **Intervening Opportunities model** [Stouffer 1940]
 - Connect to potential targets in the order of proximity irrespective of physical distance
 - closest first,
next closest second,
etc
- e.g. Will prefer to visit nearest hospital in an emergency

Winds and Currents

- Time of travel can be asymmetric to represent prevailing winds, currents
- Frequency of prevailing wind encoded in *cost* or *deterrence* functions

- General Approach to Modelling in Archaeology
- Geography and Interactions
 - Models to define the edges given the sites
- Comparing Models
- Robustness

Island Archipelagos as an Ideal Network

Vertices = Major Population or Resource Sites

Edges = Exchange between sites

- physical trade of goods or transmission of culture
- direct contact or island hopping links

- Interactions controlled by physical limitations of ancient sea travel
→ Simple Links

*The 'Minoa',
see A.Simandiraki (2005) for
commentary on this reconstruction*

Geography and Zones of Control

- Thiessen Polygons (Voronoi Diagrams)
 - equal site sizes
- XTent model
 - [Renfrew and Level 1979]
 - Thiessen with variable site sizes
- Rihll & Wilson model
 - [1987, 1991]

Models of Spatial Extent

→ **Spatial Clustering Methods**

Geography and Interactions

Models of interaction (rather than control) can be classified by the way they treat different aspects:-

- Distance
- Costs of Travel
- Input/Output constraints

Network Description

- d_{ij} Fixed distance from site i to site j
- S_i Fixed site size or capacity
(perhaps also representing hinterland)
- F_{ij} Variable flow from site i to site j ,
the edge weight in a network

Example: Gravity Models

- Models of modern transport systems
 - “Applications to traffic engineering of the law of retail gravitation” [Casey 1955]
- Cost-Benefit viewpoint
 - All trips equally likely subject to constraints on total **“cost”**
 - Maximum Entropy [Wilson, 1967]

Cost constraint – Simple Gravity Models

Not usually actual financial cost.

Typical examples:

- Linear

$$c_{ij} = d_{ij}$$

- Log

$$c_{ij} = \log(d_{ij})$$

Cost constraint – Simple Gravity Models

- Cost $\mathbf{c}_{ij} = \mathbf{d}_{ij}$
 $\Rightarrow$ exponential fall off

$$F_{ij} = S_i S_j \exp(-\gamma d_{ij})$$

- Cost $\mathbf{c}_{ij} = \ln(\mathbf{d}_{ij})$
 $\Rightarrow$ power law fall off

$$F_{ij} = \frac{S_i S_j}{(d_{ij})^\gamma}$$

*Similar
to
Newton's
law of
gravity
hence
model's
name*

Input/Output Constraints:-

DCGM

The Doubly Self-Consistent Gravity Model

Flow F_{ij} from site i size S_i to site j size S_j is

$$F_{ij} = \underbrace{a_i b_j}_{\text{input/output constraints}} S_i S_j \underbrace{f(d_{ij})}_{\text{"cost" constraint as deterrence function in terms of distance } d_{ij} \text{ e.g.}}$$

input/output constraints

$$(a_i)^{-1} = \sum_j b_j S_j f(d_{ij})$$

$$(b_j)^{-1} = \sum_i a_i S_i f(d_{ij})$$

“cost” constraint as
deterrence function
in terms of distance

d_{ij} e.g.

$$f(d_{ij}) = \frac{1}{(d_{ij})^\gamma}$$

Optimisation of what? Cost/Benefit Analysis

'Energy', resources

***Favour even
distribution of trips***

***Constrain costs
of interactions***

***Constrain Outputs
to be S_i***

***Constrain Inputs
to be S_j***

$$\begin{aligned} H = & \\ & - \kappa \sum_i F_{ij} (\ln(F_{ij}) - 1) \\ & - \beta \left(C - \sum_{ij} F_{ij} c_{ij} \right) \\ & + \sum_i \alpha_i \left(S_i - \sum_j F_{ij} \right) \\ & + \sum_j \alpha'_j \left(S_j - \sum_i F_{ij} \right) \end{aligned}$$

$$0 \leq \sum_j F_{ij}$$

PPA - Proximal Point Analysis

- Equal sized sites or size ignored
- Sites connect to ***k*** nearest neighbours

***Examples: Terrell 1977; Irwin 1983; Hage & Harary 1991;
Broodbank 2000; Collar 2007, 2013***

MDN – Maximum Distance Network

Equal sized sites, connect if **D** or less apart

Stochastic Model – ariadne

[Evans, Knappett and Rivers 2008- & figshare.com]

- Intrinsic volatility ('temperature')
- Sites and connections vary in size
- Network will give an low value of a 'cost' function

Optimisation of what for ariadne?

`Energy', resources

Isolated sites have optimal size $v_i = 0.5$

Interactions (trade) bring benefits

Increasing 'population' has a cost

Each trade link has a cost

$$\begin{aligned} H = & -\kappa \sum_i 4S_i v_i (1 - v_i) \\ & -\lambda \sum_{i,j} (S_i v_i) \cdot e_{ij} V(d_{ij} / D) \cdot (S_j v_j) \\ & + j \sum_i S_i v_i \\ & + \mu \sum_{i,j} S_i v_i e_{ij} \end{aligned}$$

$$0 \leq \sum_j e_{ij} \leq 1 \quad 0 \leq v_i$$

- General Approach to Modelling in Archaeology
- Geography and Interactions
- **Comparing Models**
 - So many models, how can I find the one I need?
- Robustness

Comparing Networks

- Same arrangement of sites gives different networks
- How can we classify them?
- How can we compare them?

Why Compare Networks?

- If we can create broad categories of types of network model, then we can pick a model from the most appropriate category
 - Models of Zones of Control
 - Models of Mutual Interaction
 - ...
- If we use several models of same type we can see if our results are robust

Network Models

Model	Distance	Input Output Constr.	Site Size	Deterence Func	Network Type
MDN	Physical	No	Equal	Threshold	Simple
SGM	Physical	No	Fixed	Any	W,Dir,Dns
DCGM	Physical	Both	Fixed	Any	W,Dir,Dns
Rihll & Wilson GM	Physical	Both	Variable	Any	W,Dir,Dns
Alonso	Physical	Both	Variable	Any	W,Dir,Dns
PPA	Ranked	Output	Equal	Threshold	Simple
Radiation	Ranked	Output	Fixed	Power Law	W,Dir,Dns
Int. Opp. Model	Ranked	Output	Fixed	Any	W,Dir,Dns
ariadne	Physical	Output	Variable	Any	W,Dir,Dns

***W,Dir,Dns =
Weighted, Directed, Dense***

Recipe for Comparing Networks Quantitatively

Look for networks which function in similar way

1. Measure a quantity associated with vertices

- Avoid integer valued quantities or ones defined only for simple networks *e.g. average shortest path*

2. Measure similarity of each pair of vectors

- Pearson correlation coefficient
- Rank values Kendal's tau or Spearman if have outliers

3. Study with Multivariate Analysis methods

- PCA (Principal Component Analysis),
Hierarchical Clustering Methods

Example Networks

Vertex Values (Rank of PageRank)

Vertex	PPA	DPPA	MDN	DCGM	RWGM	MC
1	1.5	1.5	2	2.5	1.5	1.5
2	1.5	1.5	2	2.5	1.5	1.5
3	3	3	2	2.5	3.5	4
4	4	4	4	2.5	3.5	3

Example Networks

Correlation Matrix

(PageRank + Kendal tau method)

	PPA	DPPA	MDN	RWGM	MC
PPA	1	1	0.82	0.94	0.78
DPPA	1	1	0.82	0.94	0.78
MDN	0.82	0.82	1	0.58	0.27
RWGM	0.94	0.94	0.58	1	0.94
MC	0.78	0.78	0.27	0.94	1

Value is close to 1 is two vectors are similar,
low values indicate dissimilarity

Problem:

Every model has some/many parameters

How do we choose values for different networks when we want to make a comparison?

e.g. Do we compare
 $k=2$ PPA

to

$D=100\text{km}$ MDN ?

Different Distances – MDN model, aegean data

Different Distances – PPA model, aegean data

Answer

Look for models with same `physical' characteristic:-

- A “just” fully connected network
 - GCC formation
- Average Distance
 - usual definition does not apply to weighted networks
- Time scale to visit all nodes
 - use random walkers
- Others...?

Average Weighted Distance

Suggest we compare networks with the same **average weighted distance**

i.e. the distance between each pair of sites
multiplied by the fraction of the flow between
those sites

$$AWD = \frac{\sum F_{ij} d_{ij}}{\sum F_{ij}}$$

Hierarchical Clustering Dendrogram

- *aegean39S1L3a*
- *AWD=70km (GCC just formed)*
- *PageRank and Weighted Betweenness*
- *Pearson correlation matrix*
- *Complete clustering*

1	PPA k=2
2	DPPA k=3
3	DCGM D=40km
4	RWGM D=60 $\alpha=1.03$
5	RWGM D=55 $\alpha=1.18$
6	SGM D=75

Principal Component Analysis Scatter Plot

1	PPA k=2
2	DPPA k=3
3	DCGM D=40km
4	RWGM D=60 $\alpha=1.03$
5	RWGM D=55 $\alpha=1.18$
6	SGM D=75

- *aegean39S1L3a*
- *AWD=70km (GCC just formed)*
- *PageRank and Weighted Betweenness*
- *Pearson correlation matrix*
- *Complete clustering*

- General Approach to Modelling in Archaeology
- Geography and Interactions
- Comparing Models
- **Robustness**
 - One way to use all these models

Robustness of Spatial Network Modelling

- Set up sites and distances
- Compare models with same **average weighted distance**
- Measure **function** of each network via vertex properties in each model
- Choose similar models using correlation matrix analysis
- **Use resulting similar models to test ideas**

Acknowledgements

- All work done with
 - **Carl Knappett** (Toronto)
 - **Ray Rivers** (Imperial)
- with help along the way from
 - **Giovanni Brandani** (Imperial)
 - **Edmund Hunt** (Imperial)
 - **Eric Beales** (Toronto)
- netplexity.org figshare.com
or search for “Tim Evans archaeology”
<http://imperial.ac.uk/people/T.Evans>

The Connected Past

- The Web Site <http://connectedpast.soton.ac.uk/>
- The Book
 - T. Brughmans, A. Collar and F. Coward (eds),
The Connected Past: challenging networks in archaeology and history (CUP)
- The Meetings
 - CAA satellite, Paris, week after Easter 2014
 - 2 day meeting, London, first or second week of September, 2014<http://www.complexity.org.uk/events/conpastuk2014/>

Bibliography

- Evans, T.S., 2004. *Complex Networks*, Contemporary Physics, 45, 455-474.
- Evans, T.S., Knappett, C., & R. Rivers, 2009. *Using statistical physics to understand relational space: a case study from Mediterranean prehistory*, in D. Lane, S. van der Leeuw, D. Pumain & G. West (eds.), *Complexity perspectives in innovation and social change*: 451-79. Berlin: Springer Methodos Series, [[doi: 10.1007/978-1-4020-9663-1](https://doi.org/10.1007/978-1-4020-9663-1)].
- Evans, T.S., Knappett, C., & R.J. Rivers, 2012. *Interactions in Space for Archaeological Models*, Advances in Complex Systems [[doi: 10.1142/S021952591100327X](https://doi.org/10.1142/S021952591100327X)].
- Evans, T.S., Knappett, C., Rivers, R., 2012. *Thirty Nine Minoan Sites*, Figshare <http://dx.doi.org/10.6084/m9.figshare.97395> .
- Evans, T.S., Knappett, C., Rivers, R., 2012. *ariadne*, Figshare <http://dx.doi.org/10.6084/m9.figshare.97746>
- Evans, T.S, (2014). *Which Network Should I use*, in T. Brughmans, A. Collar and F. Coward (eds), *The Connected Past: challenging networks in archaeology and history*. Cambridge: Cambridge University Press (in press).
- Knappett, C., Evans, T.S. & R. Rivers, 2008. *Modelling maritime interaction in the Aegean Bronze Age*, *Antiquity* 82: 1009-24.
- Knappett, C.; Evans, T.S. & Rivers, R., 2011. *The Theran eruption and Minoan palatial collapse: new interpretations gained from modelling the maritime network*, *Antiquity*, 85, 1008-1023.
- Rivers, R.; Knappett, C. & Evans, T.S., 2013. *Network Models and Archaeological Spaces in Computational Approaches to Archaeological Spaces*, A.Bevan & Lake, M. (Eds.), Left Coast Press.

Bibliography

- Ortúzar, J. d. D. & Willumsen, L. 1994. “Modelling Transport” Wiley.
- Collar, A, 2007. Network Theory and Religious Innovation *Mediterranean Historical Review*, 22, 149-162 .
- Collar, A, 2013. book just published
- Hage, P. & Harary, F. 1991. Exchange in Oceania: a graph theoretic analysis, *Clarendon Press*.
- Irwin 1983. Chieftainship, kula and trade in Massim prehistory in Leach, J. & Leach, E. (Eds.) *The Kula: New Perspectives on Massim Exchange*, Cambridge: CUP, 29-72 .
- Isaksen, L. 2006. Network Analysis of Transport Vectors in Roman Baetica, MSc Thesis, Univ. Southampton.
- Rihll, T.E. & A.G. Wilson, 1987. Spatial interaction and structural models in historical analysis: some possibilities and an example, *Histoire & Mesure* 2: 5-32.
- Rihll, T.E. & A.G. Wilson, 1991. Modelling settlement structures in ancient Greece: new approaches to the polis, in J. Rich & A. Wallace-Hadrill (eds.), *City and country in the ancient world*: 59-95. London: Routledge.
- Sindbæk, S. M. 2007. The Small World of the Vikings: Networks in Early Medieval Communication and Exchange *Norwegian Archaeological Review*, 40, 59-74.
- Simandiraki, A., Minoan Archaeology in the Athens 2004 Olympic Games, *Eur.J.Arch.*, 2005, 8, 157-181
- Sindbæk, S.M. 2007. Networks and nodal points: the emergence of towns in early Viking Age Scandinavia, *Antiquity* 81 , 119–132
- Renfrew & Level, 1979. Exploring dominance: predicting polities from centres, in Renfrew, A. & Cooke, K. (Eds.) *Transformations: Mathematical Approaches to Culture Change*, Academic Press, London, 145-67.
- Stouffer, S. A., 1940. Intervening Opportunities: A Theory Relating to Mobility and Distance, *American Sociological Review* 5, 845–867. [doi:10.2307/2084520](https://doi.org/10.2307/2084520).
- Terrell, J. 1977. Human biogeography in the Solomon Islands, *Fieldiana, Anthropology*, 68 .
- Wilson, A.G. 1967. A statistical theory of spatial distribution models, *Transport Research* 1, 253-269