Imperial College London

Complexity Science

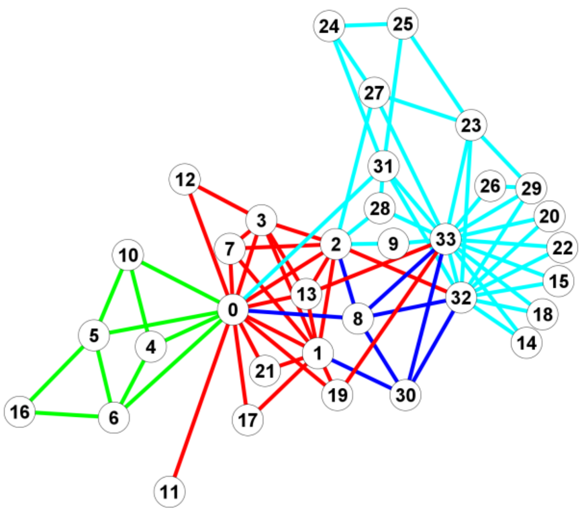
Networks and Space:

The Effects of Space on Network Analysis

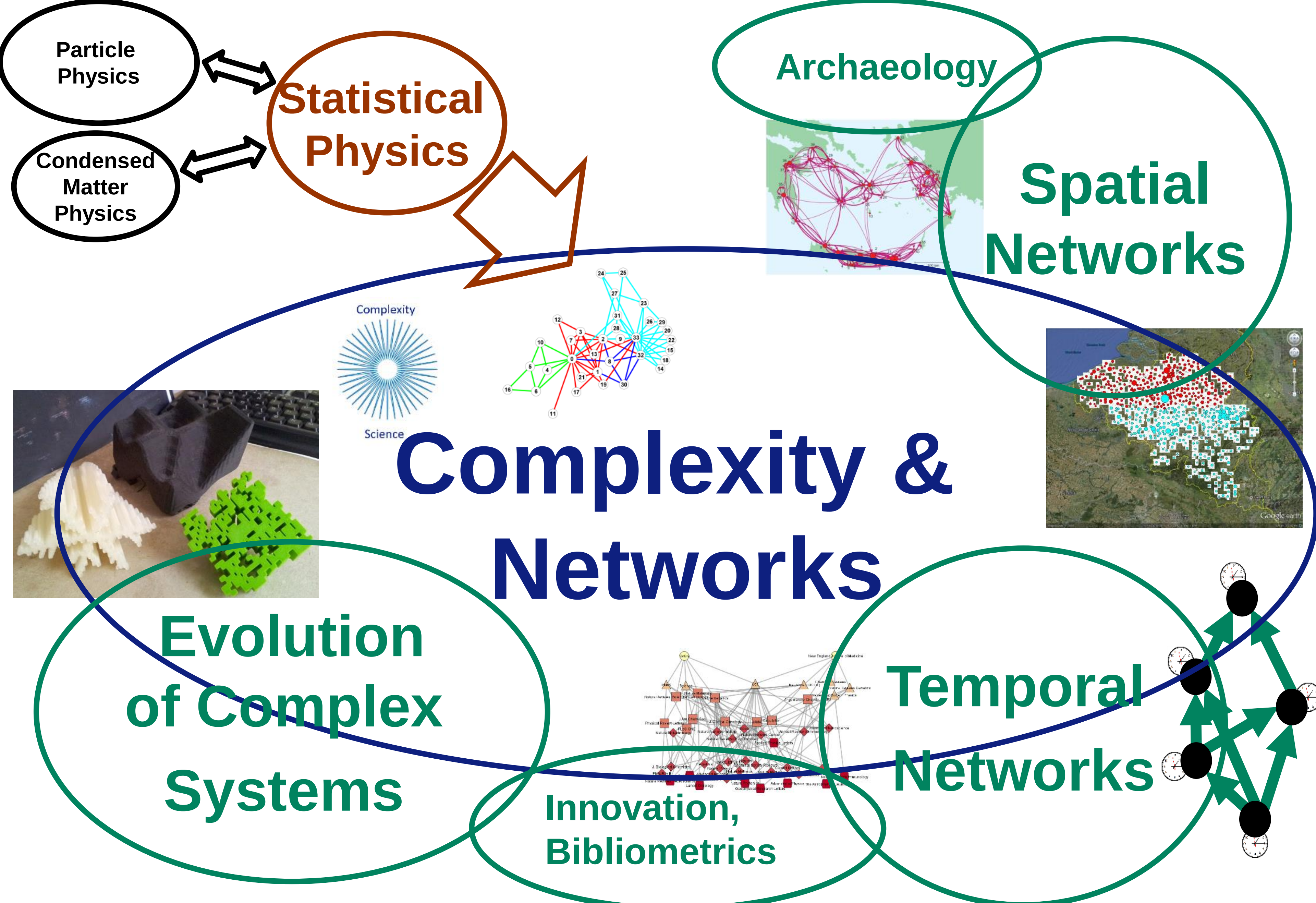
Tim Evans

Centre for Complexity Science

http://netplexity.org



17th April 2015,
SAMM - l’Université Paris 1 Panthéon-Sorbonne



Particle Physics
Condensed Matter Physics
Statistical Physics
Archaeology
Spatial Networks
Complexity Science
Complexity & Networks
Evolution of Complex Systems
Innovation, Bibliometrics
Temporal Networks

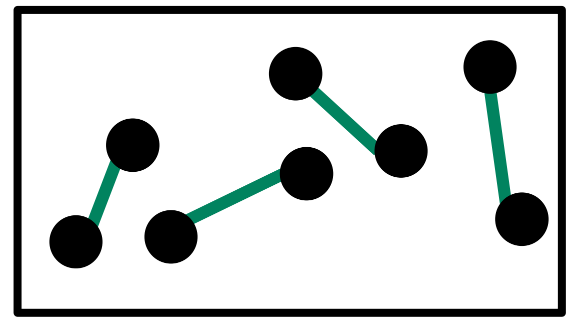
Outline

- INTRODUCTION
- AIMS
- MODELLING
- COMPARING NETWORKS
- SUMMARY

Definition of a Network

Networks are

- A set of nodes ● e.g. people
- A set of edges e.g. friendships
  - Edges describe bilateral relationships between nodes



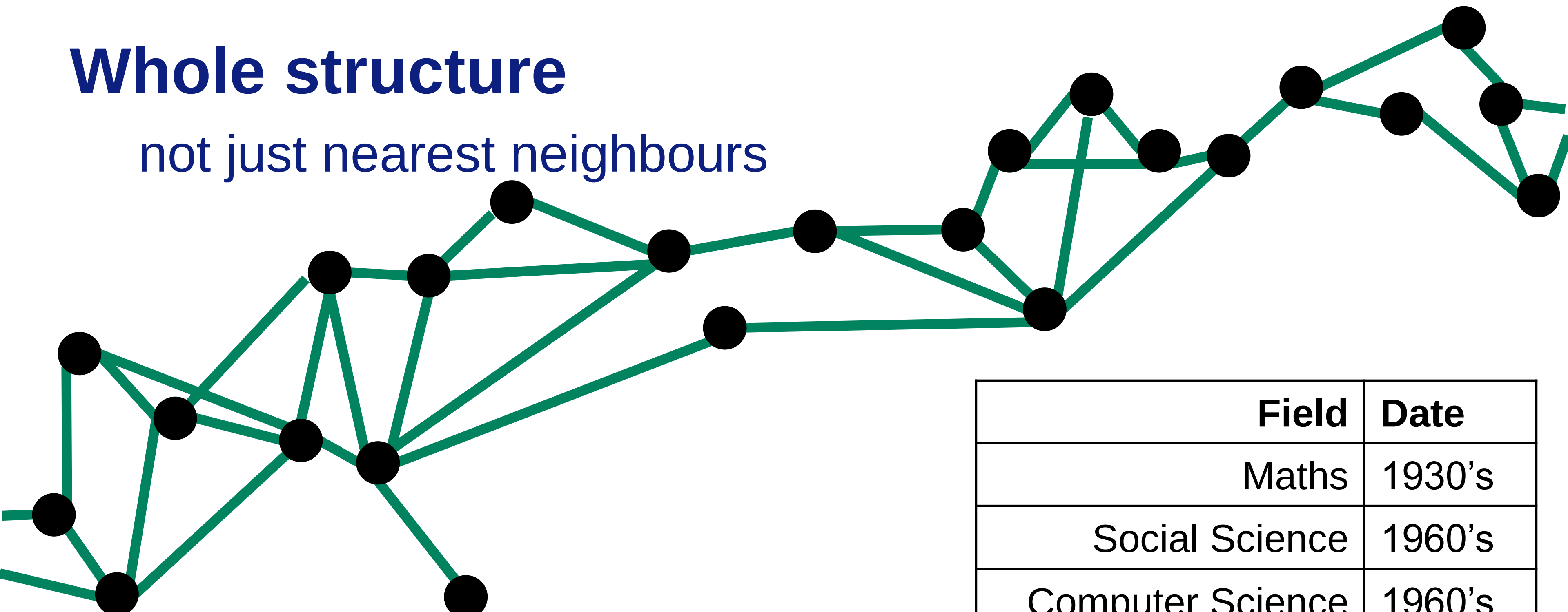
Can analyse statistics of these pairs using usual statistical methods

Network Analysis

Network analysis adds new insights when large scales are relevant

Whole structure

not just nearest neighbours



Field	Date
Maths	1930's
Social Science	1960's
Computer Science	1960's
Physics	2000's

Constraints

Most network analysis considers no constraints on the relationships

- ER random graph all edges equally likely
- Barabasi-Albert model (undirected Price model) all edges possible

Community Detection

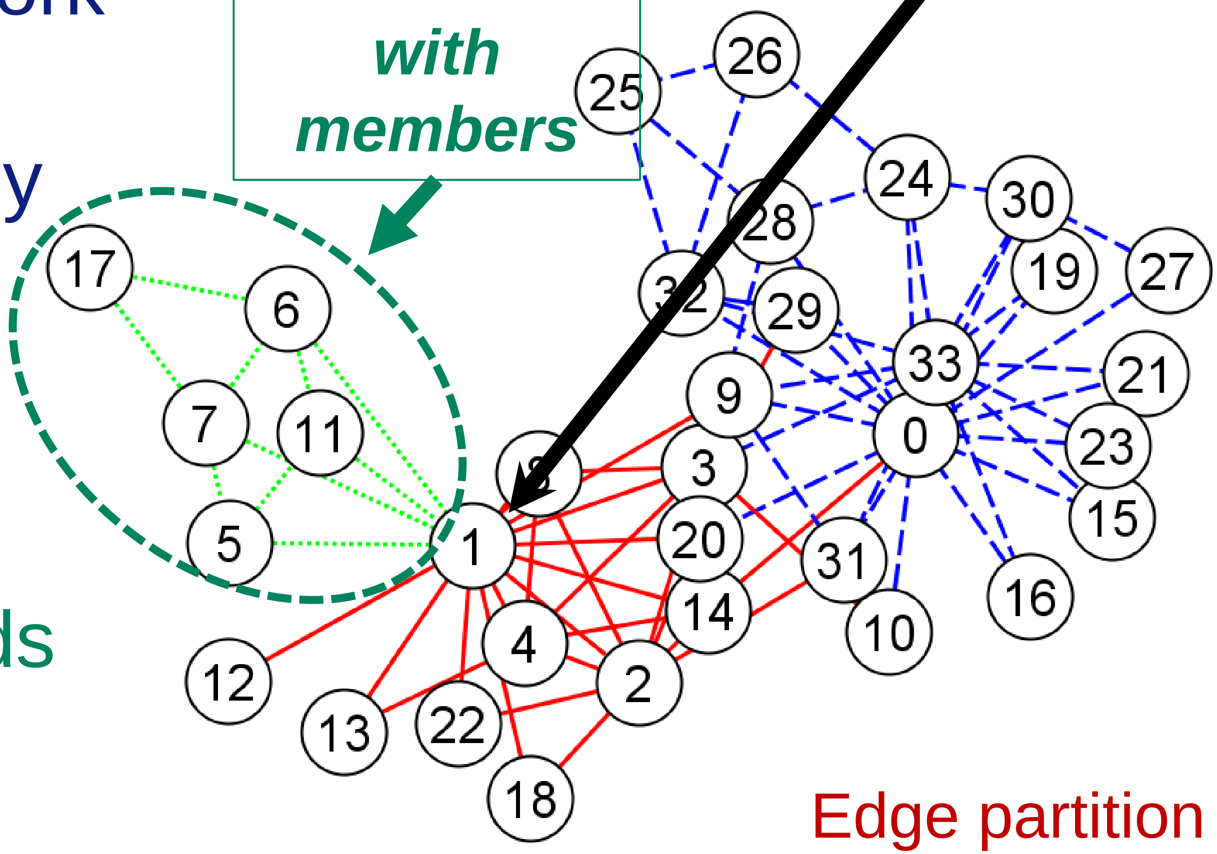
Sometimes network topology reflects constraints clearly

$\Rightarrow$

Standard methods can reveal such features

Junior Instructors not allowed to interact with members

Chief Instructor



Edge partition of Karate club network
[Evans & Lambiotte, 2009, 2010]

Constraints

In reality there are often non-topological constraints coming from other data

- Social
  - Rigid rules e.g. Zachary Karate Club
- Spatial
  - Costs of long distance connections

Focus of this Talk

- Time
  - Short term memory

Networks and Constraints

Many times key features are hidden if you do not understand the constraints

$\Rightarrow$

Need to include effects of constraints in

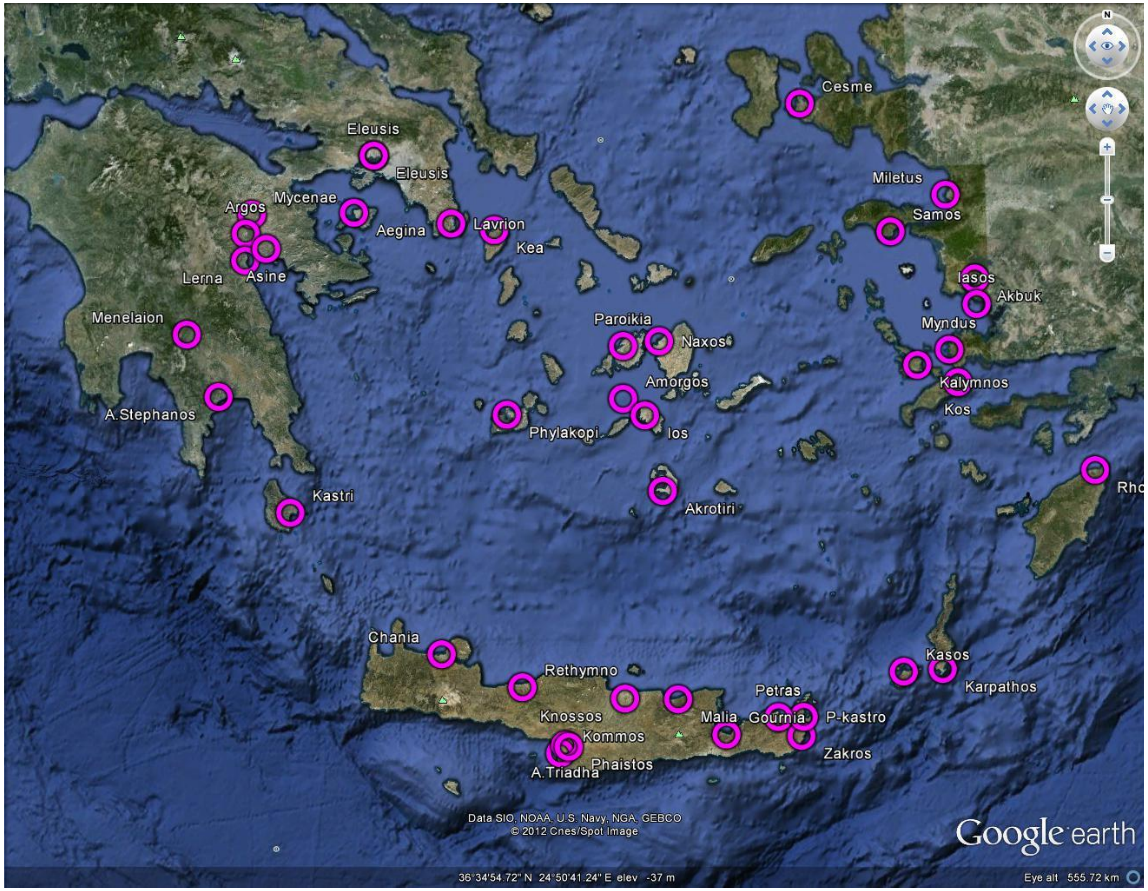
- Models
- Analysis

AIMS

- My First Question
- Applications
- Type of Interactions
- Uses for Answers

My First Question

Given the positions of some sites, what sort of interactions can I expect?



Minoan sites, Middle Bronze Age c2000-1500BC

[Knappett, TSE, Rivers, 2008-2012]

Some Applications

- Transport
  - Traffic flow
- Urban Planning
  - Commuting Patterns
- Economics
  - Migration
- Communications
- Archaeology

Type of Interactions

- Zones of Control
  - Political power, who controls whom?
  - Centralised service provision,
    e.g. Hospitals, shopping malls
- Flows
  - Trade
  - Commuting patterns
  - Migration patterns
  - Information e.g. Potters wheel

Uses for Answers

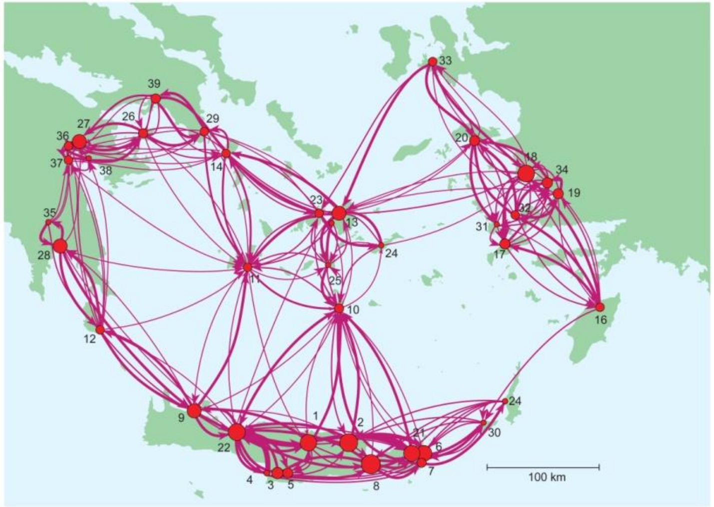
- Fill in missing data
  - Large gaps in archaeological record
- Predictions
  - How does the complex system respond to change?
  - Where should we locate a new service?
- Comparison
  - As null models, to highlight features in real data

Missing Data in Archaeology

- Often records limited or plain missing.
- Where records exist for individual sites, hard to combine for social and scientific reasons.

Example

- Eruption on Thera (modern Santorini)

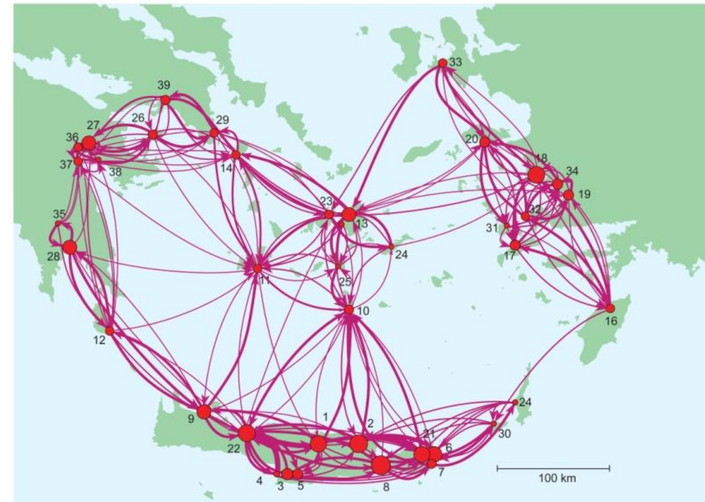


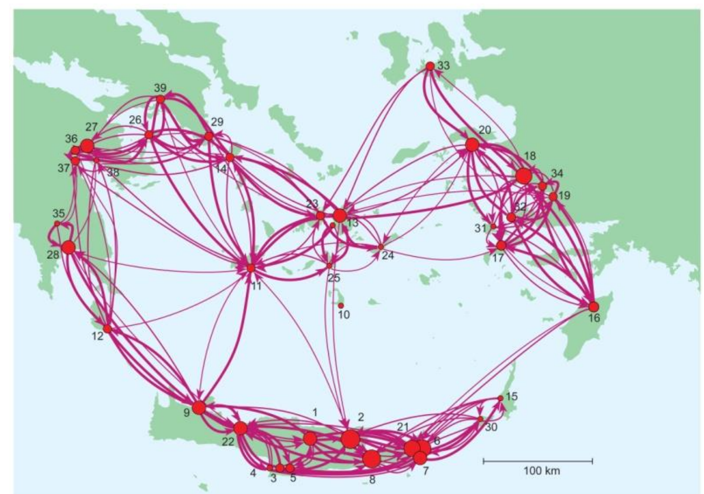
Example – Theran eruption

[Knappett, TSE, Rivers, Antiquity 2012]

- Eruption on Thera (modern Santorini) c1600/1500BC

- Thera at key distance, from Cretan coast, Knossos being one of closest points
- Compare networks before and after

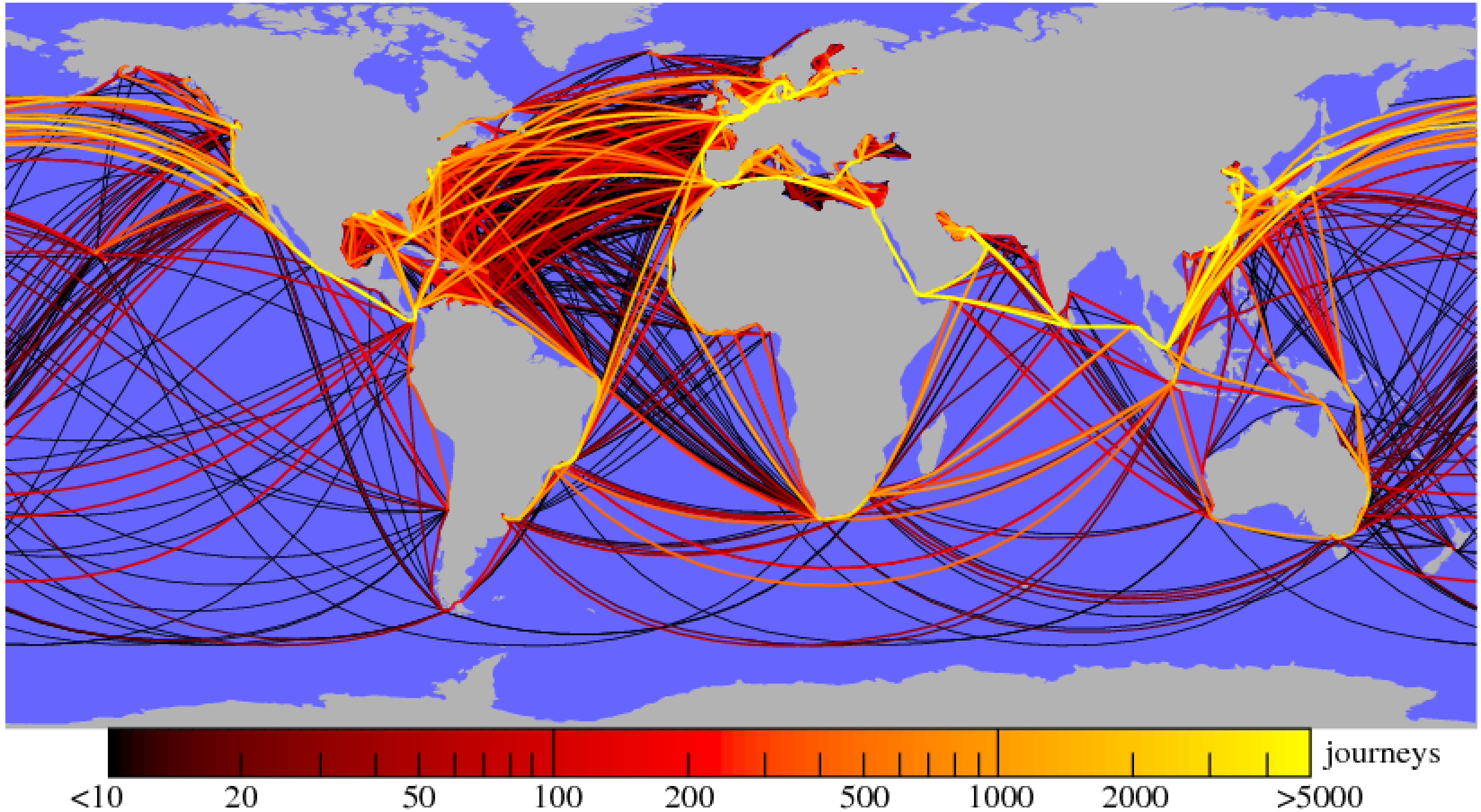




Transport & Ecology

Cargo Ship Movements and Invasive Species [Kaluza et al, 2009]

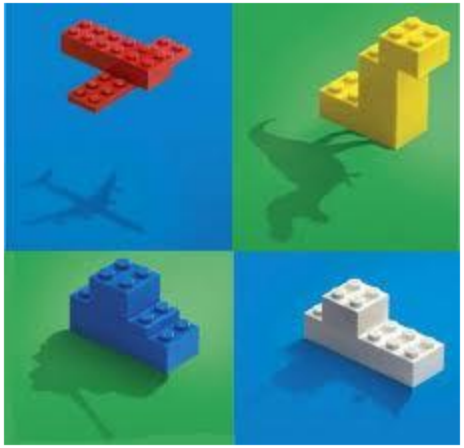
Vertices = Ports, Edges = Trips From/To



Aims of Spatial Modelling

- Many Contexts
- Many different aims

Discussion here will be very generic
with examples drawn from my experiences

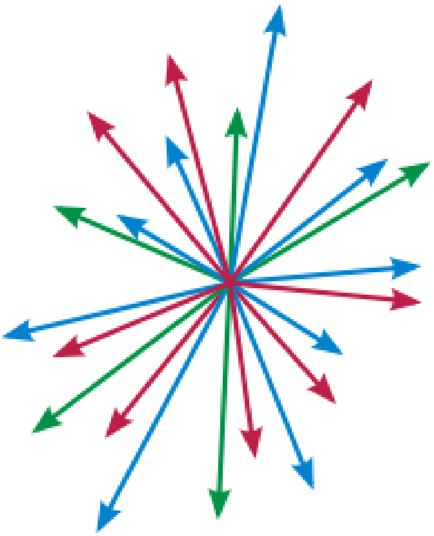


MODELLING

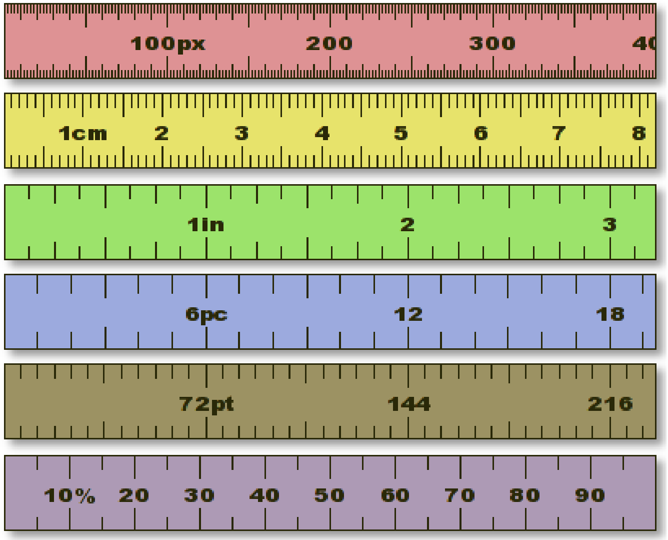
- Choices common to all models
  - Space
  - Distance
- Zone of control models
- Network Models
  - Threshold models
  - Maximum Entropy Models
  - Stochastic Models

MODELLING - Choices Common to All Models

- Which space do we work in?

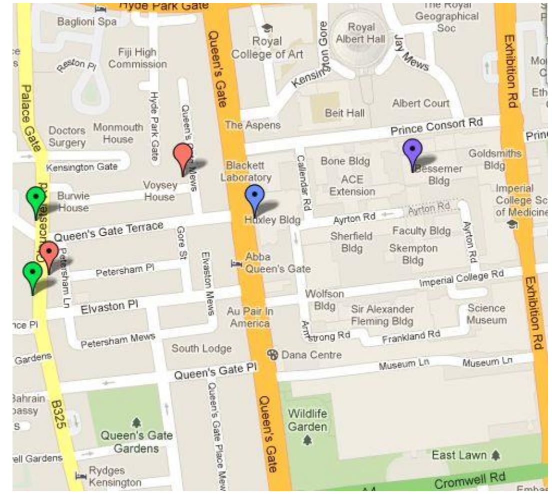


- How do we measure distance?

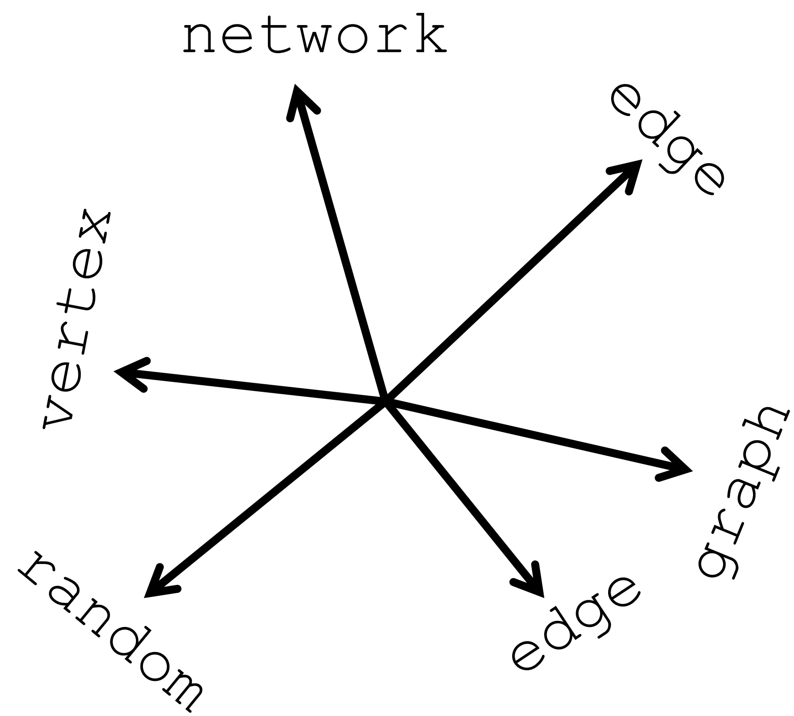


Space

- We will work with two-dimensional space

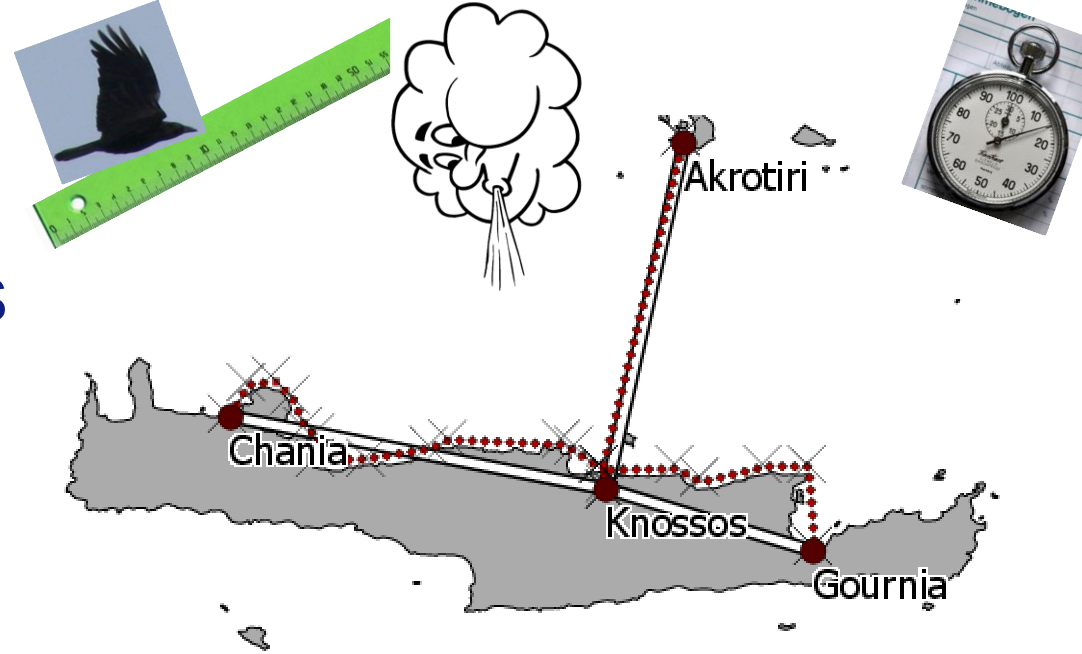


- All ideas can be applied to artefact spaces e.g. document similarity measured in keyword frequency space



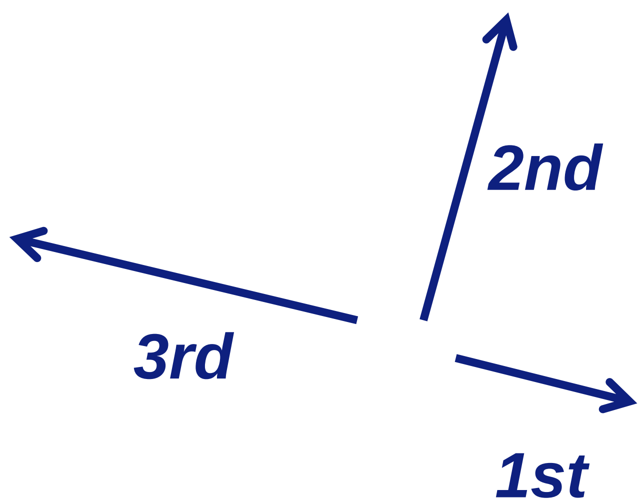
- Co-occurrence in text

Different Distances



- Physical Distances
  - As the crow lies
  - Shortest route in km
  - Quickest time
  - Lowest costs
  - ...

- Ranked distances
  - Nearest neighbour, second nearest neighbour, etc




Ranked distances

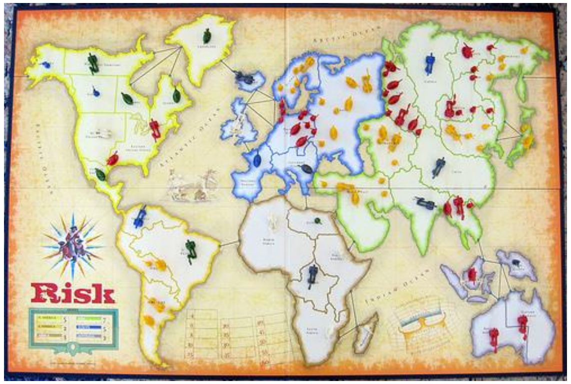
- Used by Stouffer 1940 in Intervening Opportunities model
- Connect to potential targets in the order of proximity irrespective of physical distance
  - closest first,
    next closest second,
    etc

e.g. Will prefer to visit nearest hospital in an emergency and distance to it or to the next nearest is not very relevant

MODELLING - Zones of Control models

- Common border
- Voronoi Tesselation/Thiessen Polygons and Delauney Triangulation
- XTent model

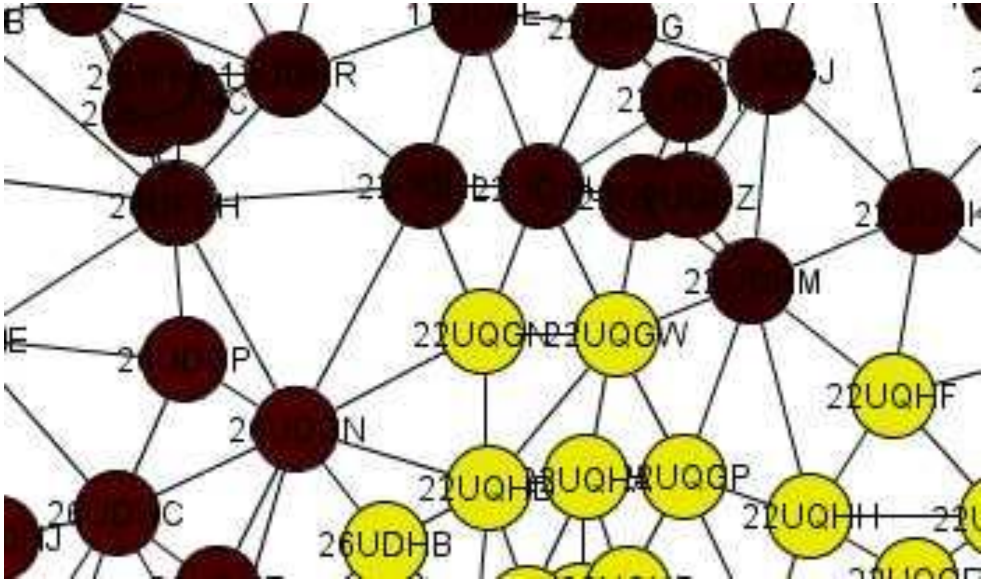
Clustering/Zone of Control models



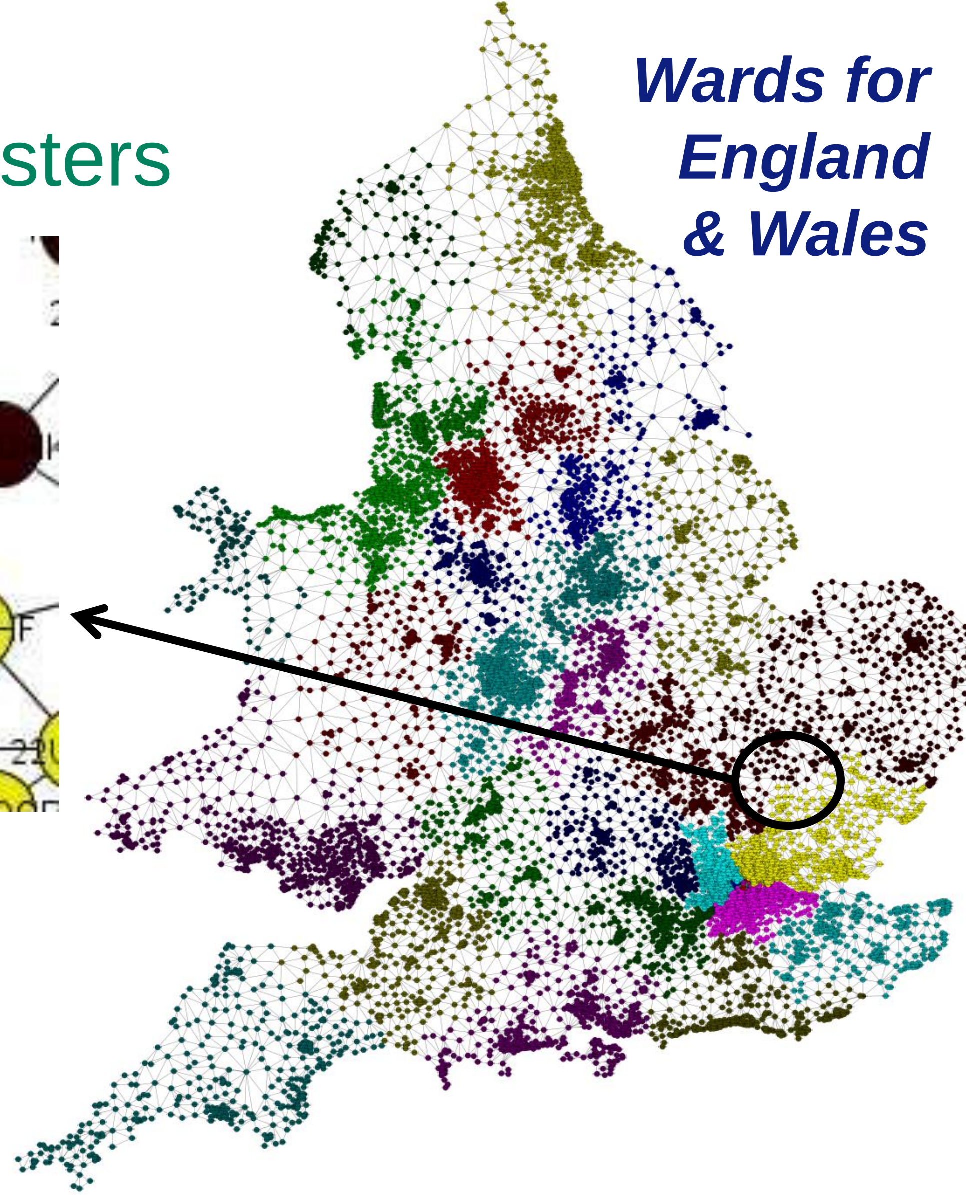
Who controls what?

- Networks simple, just nearest neighbours
- Classic example Delauney Triangulation
  (dual of Voronoi tessellation = Thiessen Polygons)
- Xtent model generalisation
  [Renfrew & Level, 1979; Bevan 2010]

Nearest Neighbours Clusters



Wards for England & Wales



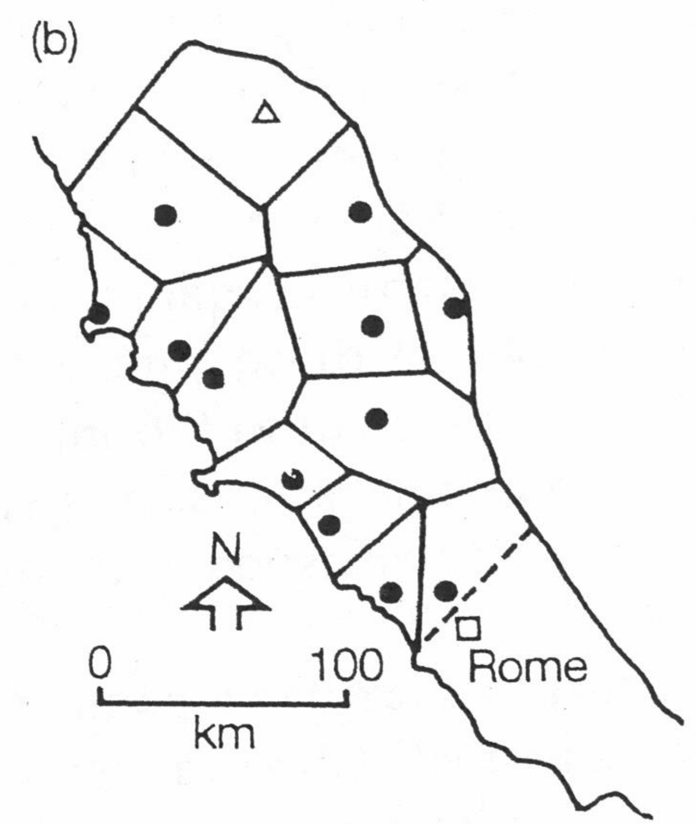
Connect wards if
- share common border
- both are large

⇒ **Spatial Hierarchical Clustering** = Bottom-up city definition [Arcaute et al 2014]

Voronoi Tessellation/Theissen Polygon

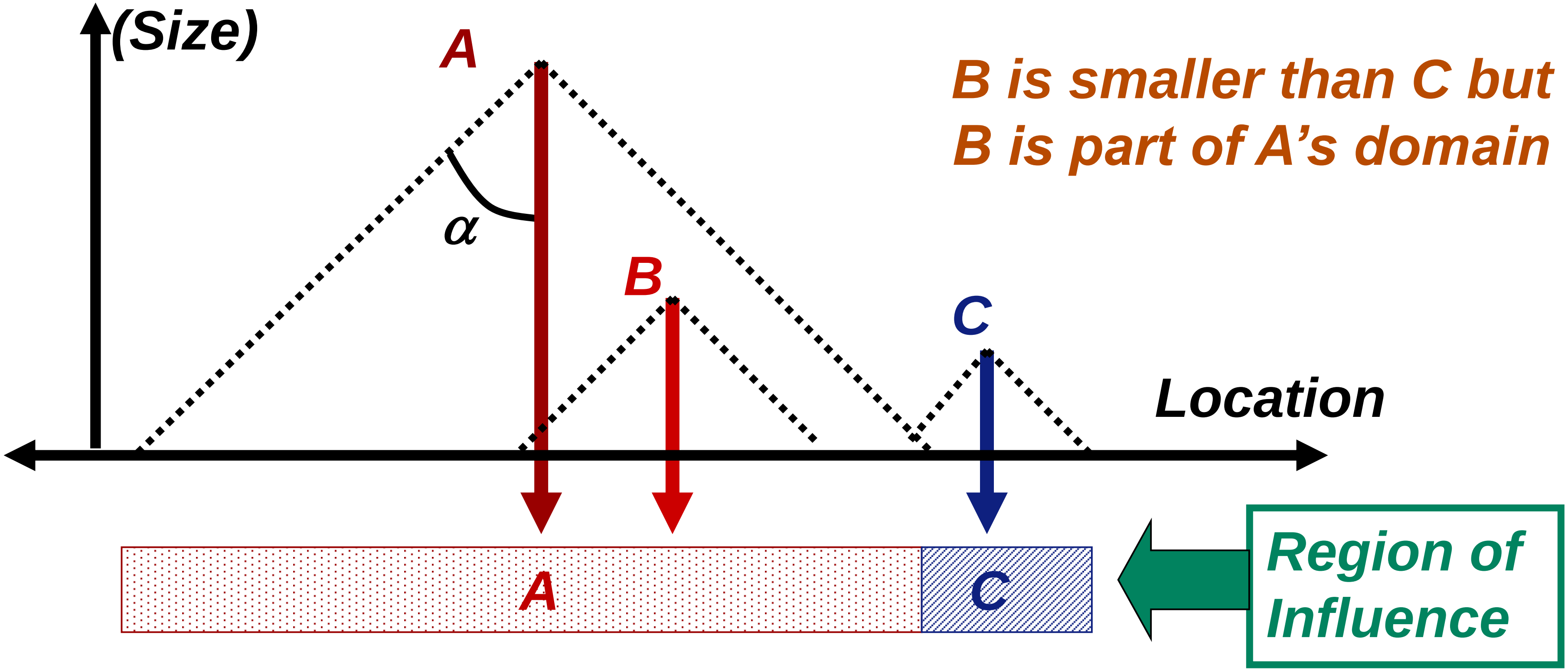
12 Etrurian Cities

[Renfrew 1975]

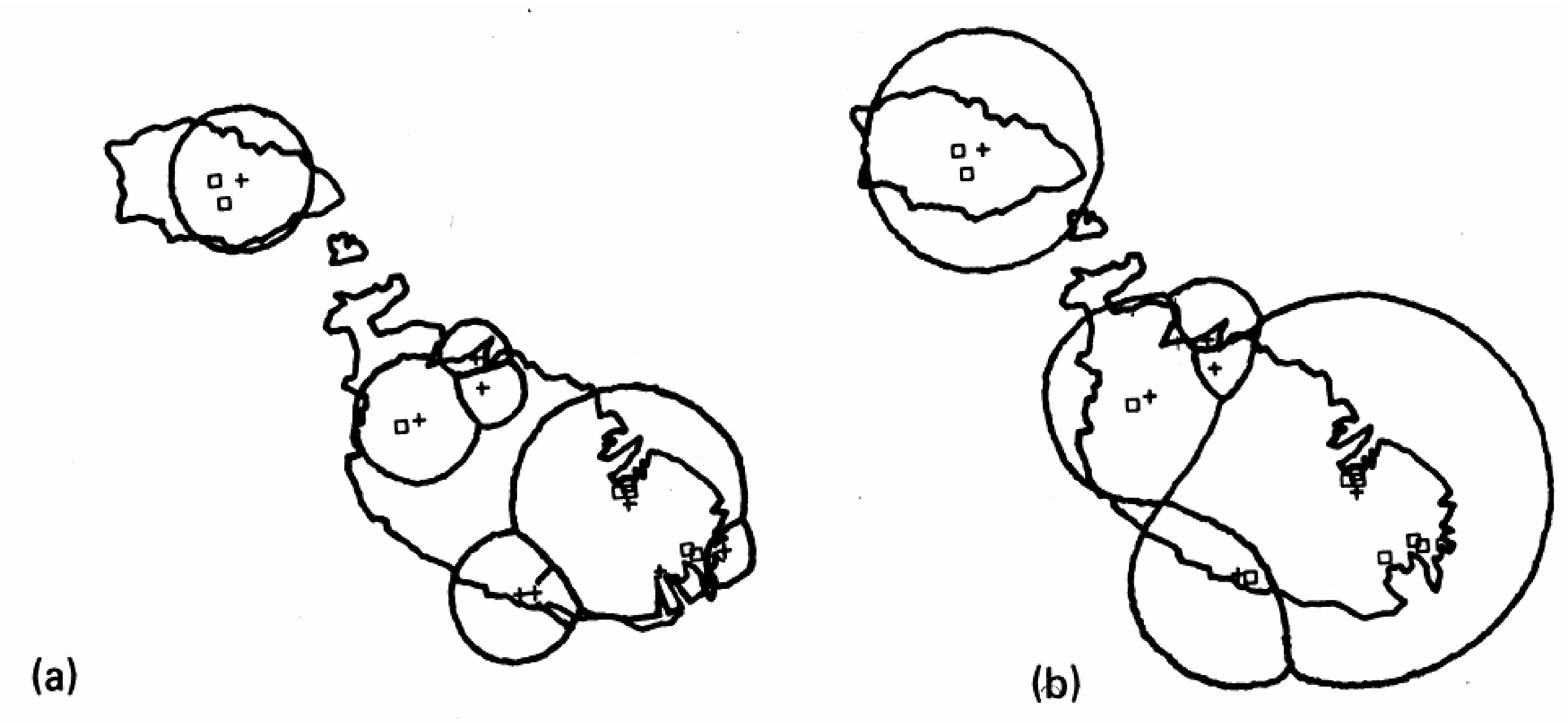


XTent Model [Renfrew & Level, 1979]

- Thiessen polygons for unequal size sites
- Can set influence of site as function of distance to any suitable function



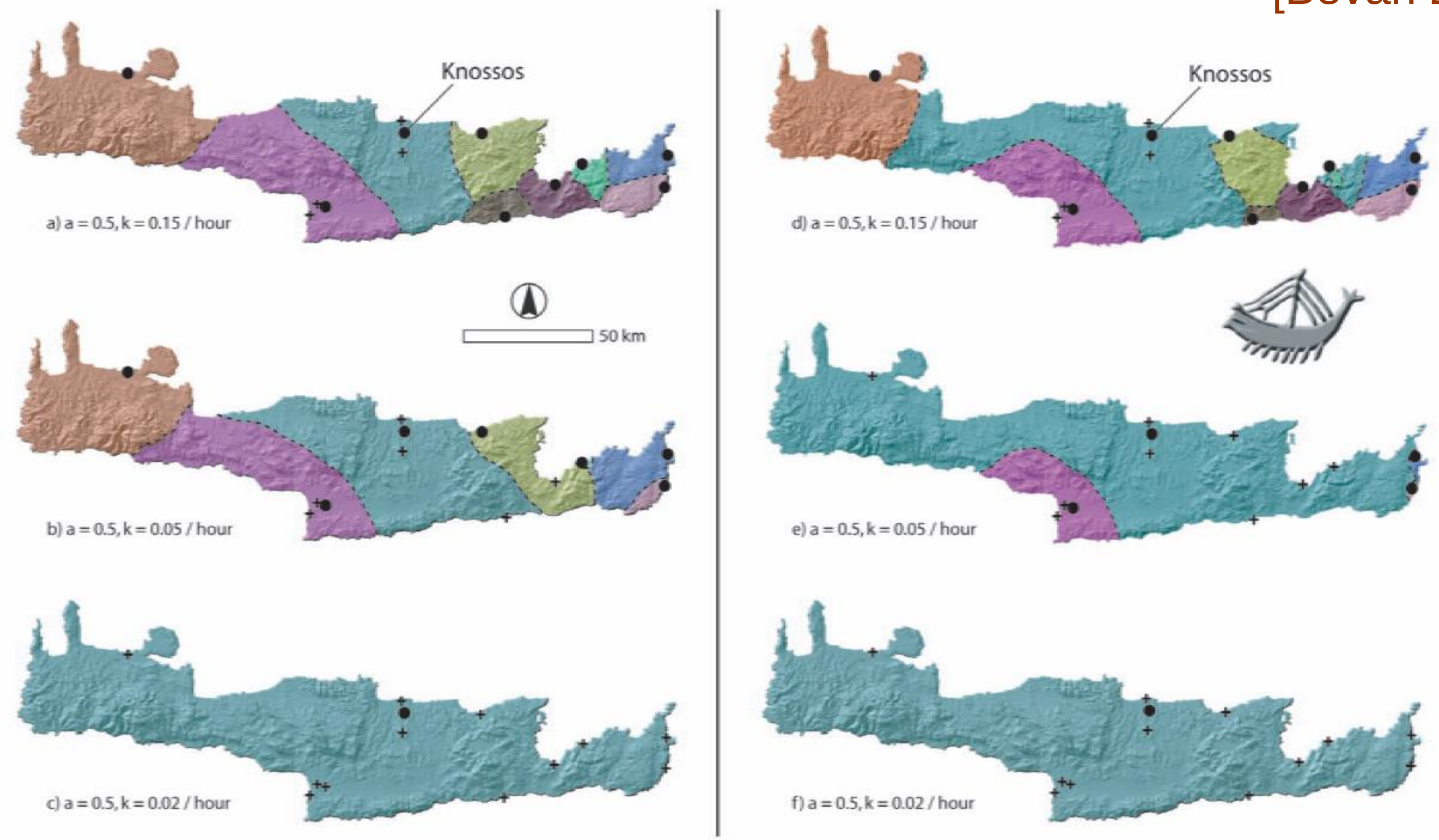
XTent model examples [Renfrew & Level, 1979]



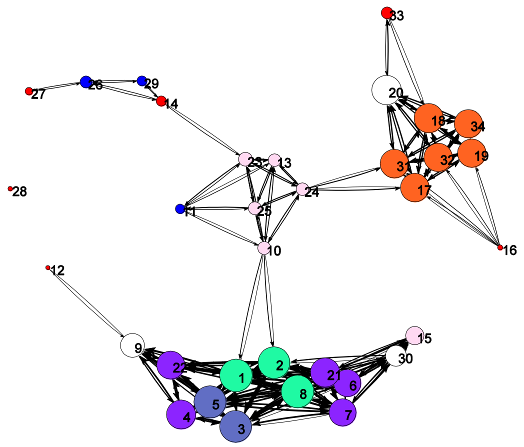
Neolithic Temples of Malta

Xtent Neopalatial Crete (~1750BC - ~1500BC)

[Bevan 2010]



MODELLING - Network Models

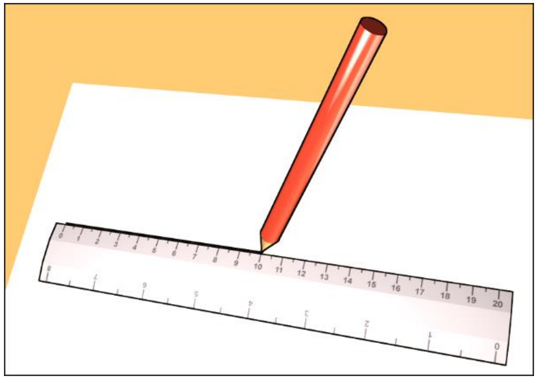


- Threshold models
- Maximum Entropy Models
  - Gravity Models
  - Rhill & Wilson
  - Radiation Model
- Stochastic Models

Threshold models

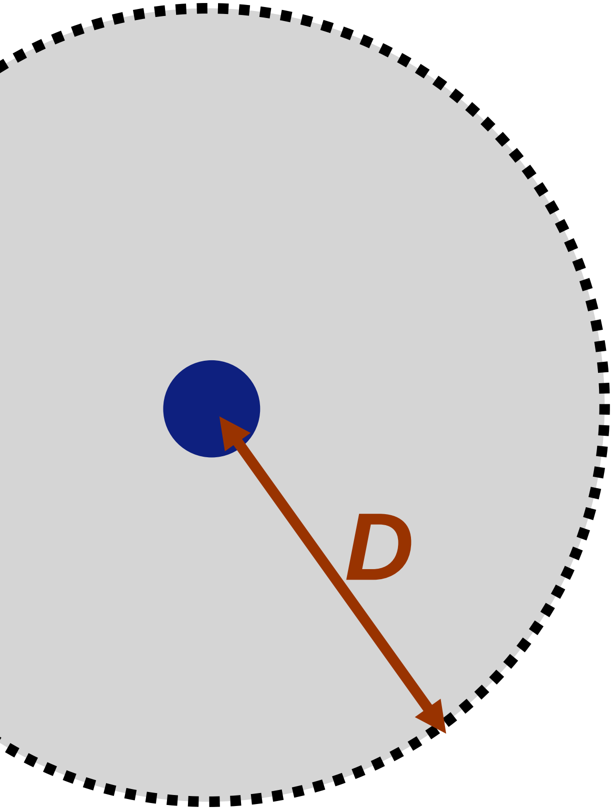
Connect to all sites within distance D

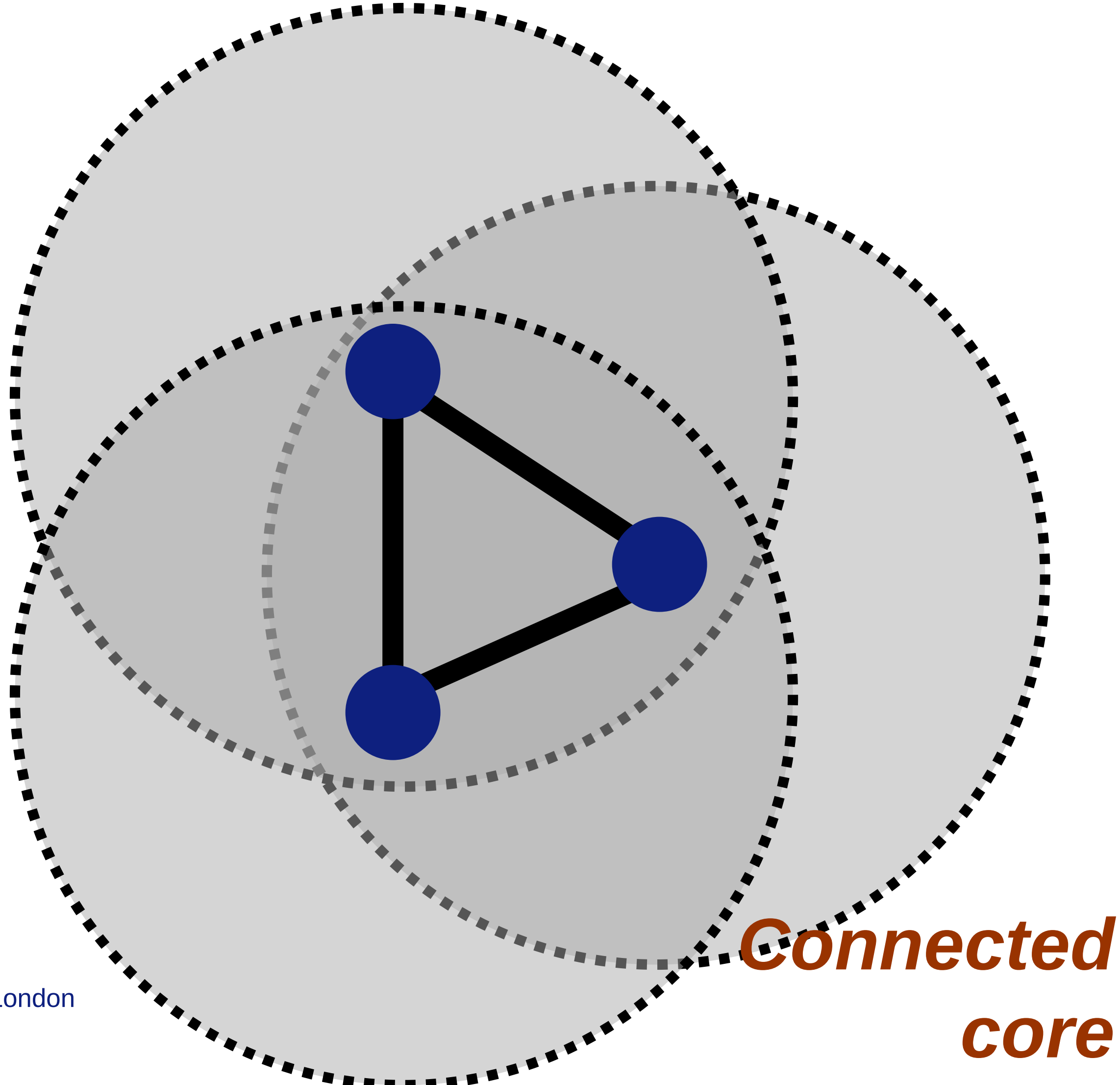
- Using physical distances get Maximum Distance Network model
- Using ranked distances get Proximal Point Analysis
- Simple “pencil and paper” models



MDN – Maximum Distance Network

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





MDN – Maximum Distance Network

Sites distance ***D*** or less apart are connected

- Theoretically tractable, often used with randomly generated site locations
  – Poisson Point Processes
  - ad-hoc wireless models [e.g. Srinivasa & Haenggi 2010]
  - Random Geometric Graphs [e.g.Penrose 2003]

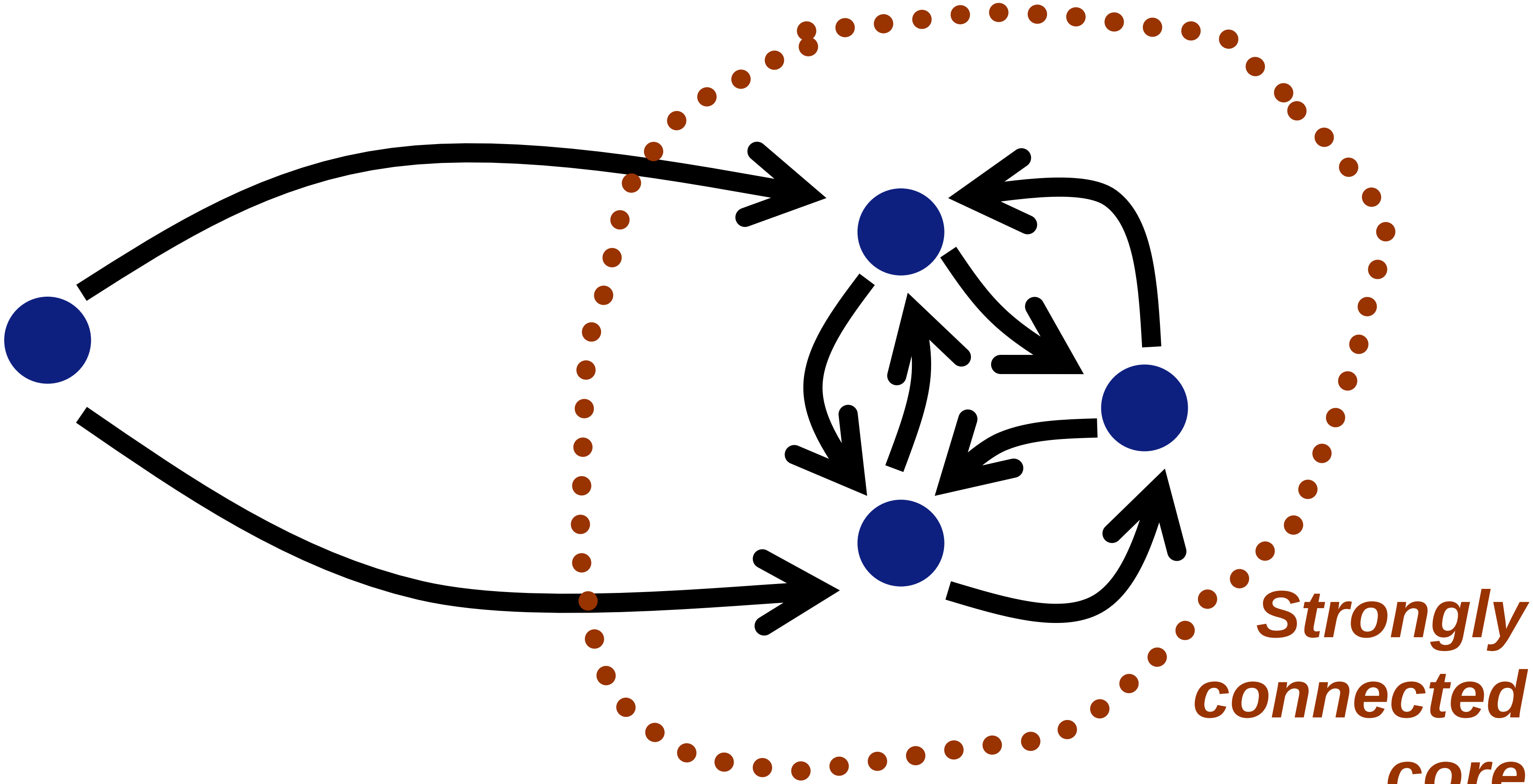
- Not used much with real sites in archaeology

PPA - Proximal Point Analysis

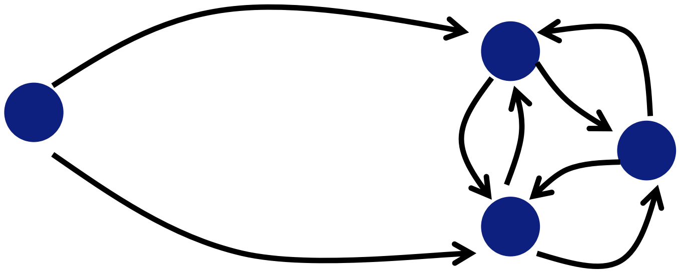
- Sites connected to ***k*** nearest neighbours
  - Maximum Distance Network using Rank distance
- Popular in Archaeology
  - [Terrell 1977; Irwin 1983; Hage & Harary 1991; Broodbank 2000; Collar 2007]
- Simplest example of the use of **Ranked Distance** not *Physical Distance* (Intervening Opportunities Model)

DPPA Example (Directed PPA)

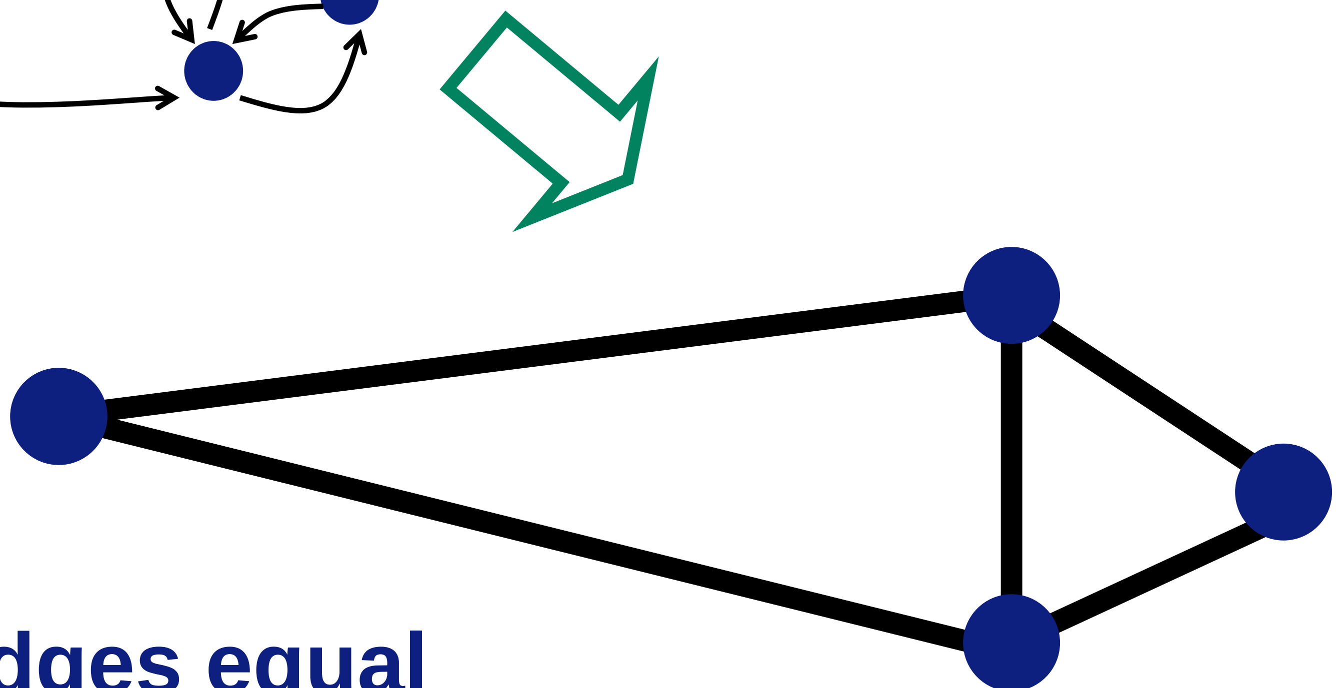
Connect each site to its ***k=2*** nearest neighbours



PPA Example



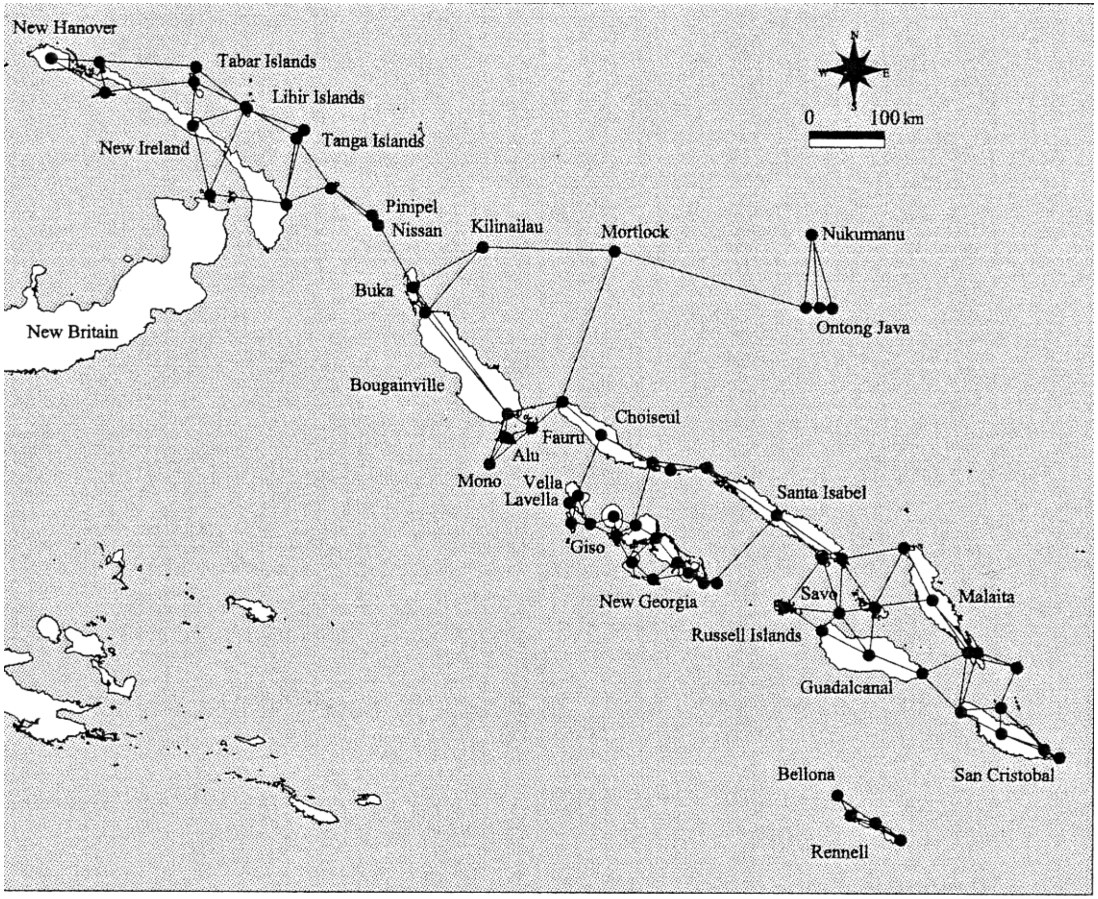
Ignore direction



- **All edges equal**
- **Network now simply connected**

Terrell (1977)

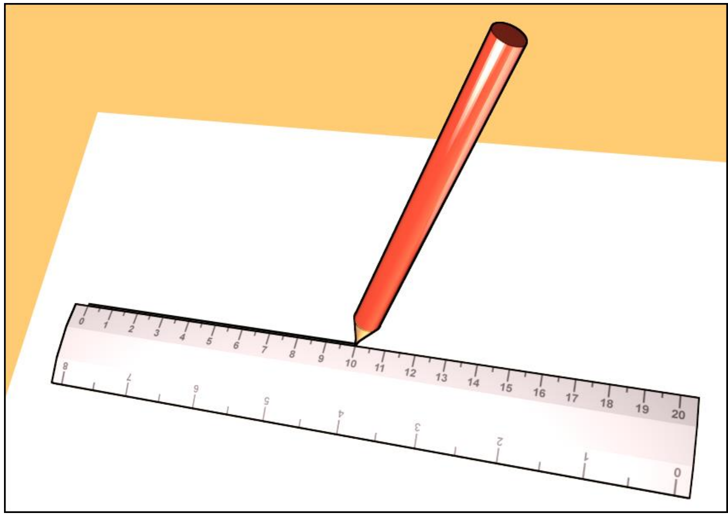
- Solomon Islands (east of Papua New Guinea)
- PPA analysis



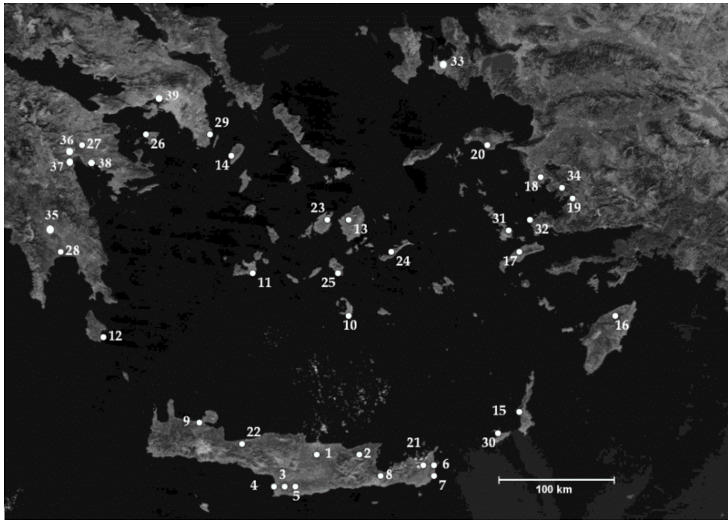
A More Sophisticated Network Description

MDN and PPA are very simple models

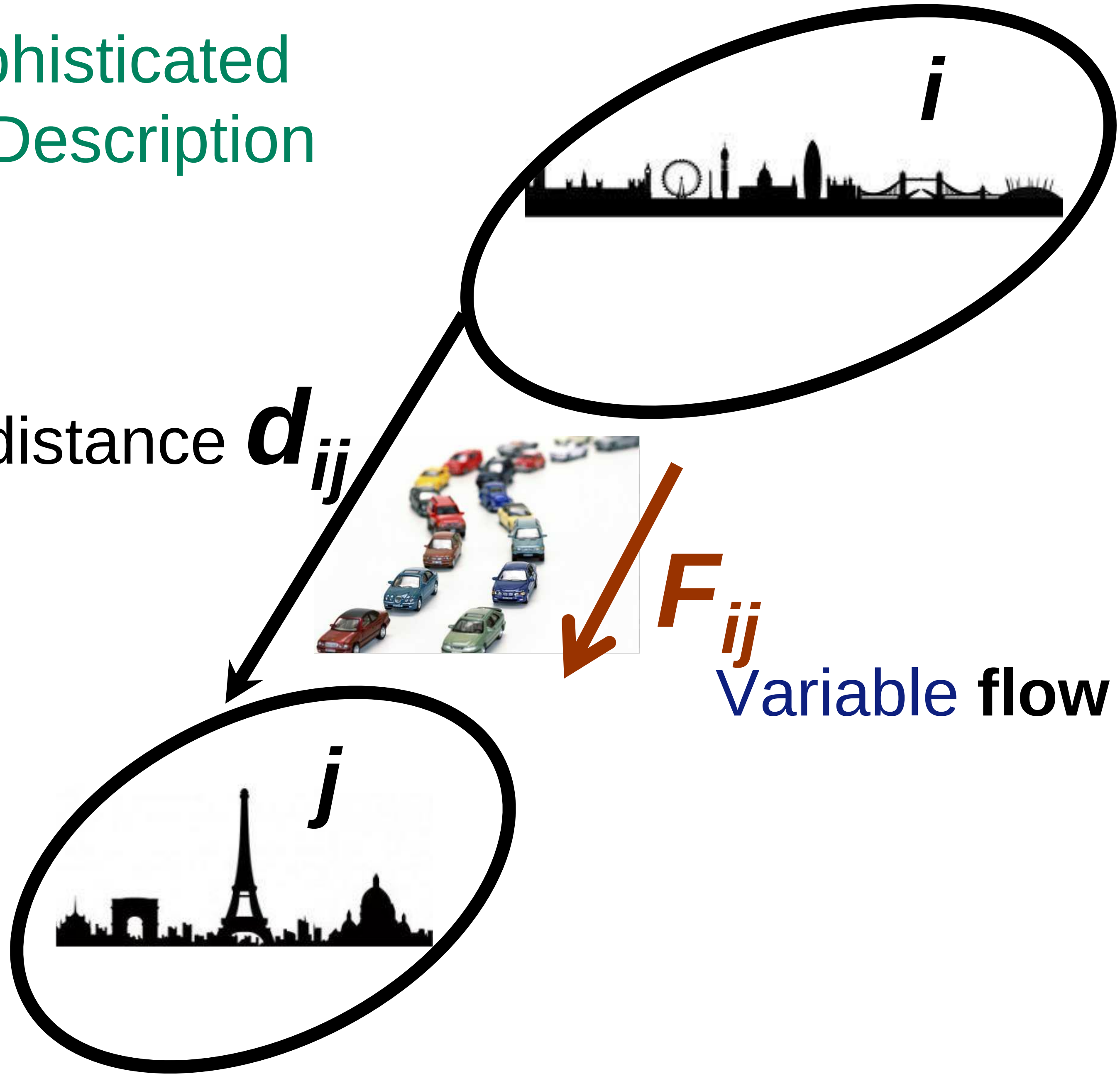
- You can do them with paper and pencil



BUT do they capture all the nuances of a spatial system?



More Sophisticated Network Description



Fixed distance d_{ij}

F_{ij}
Variable **flow**

Fixed
site size
S_j

More Sophisticated Network Description

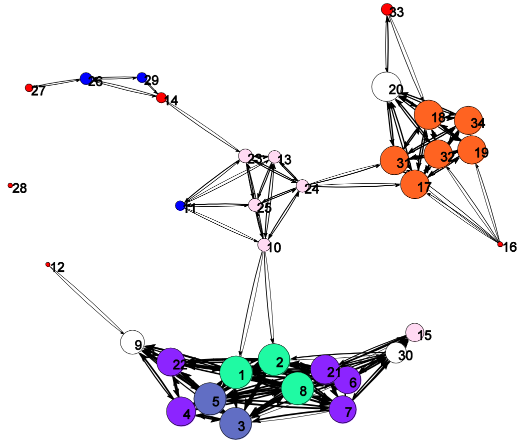
- d_{ij} Fixed distance from site i to site j
- S_i Fixed site size or capacity

    perhaps also representing hinterland

- F_{ij} Variable flow from site i to site j,

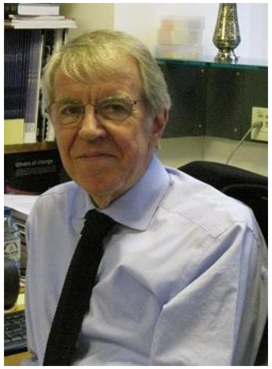
    the edge weight in a network

MODELLING - Network Models

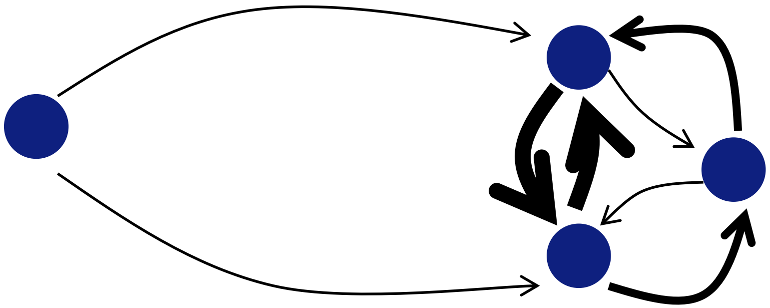


- Threshold models
- Maximum Entropy Models
  - Gravity Models
  - Rhill & Wilson
  - Radiation Model
- Stochastic Models

Maximum Entropy Models



- The approach was pioneered in 1967 by Alan Wilson
- Described in terms of the flow matrix F_{ij} number of trips from site i to site j
- Gives directed weighted dense networks
- In practice many edges have low weights



Maximum Entropy

- Number of trips from ***i*** to ***j*** is flow $\boldsymbol{F_{ij}}$
- Each trip equally likely
- Maximise entropy **S=ln(Ω)** $\Rightarrow$

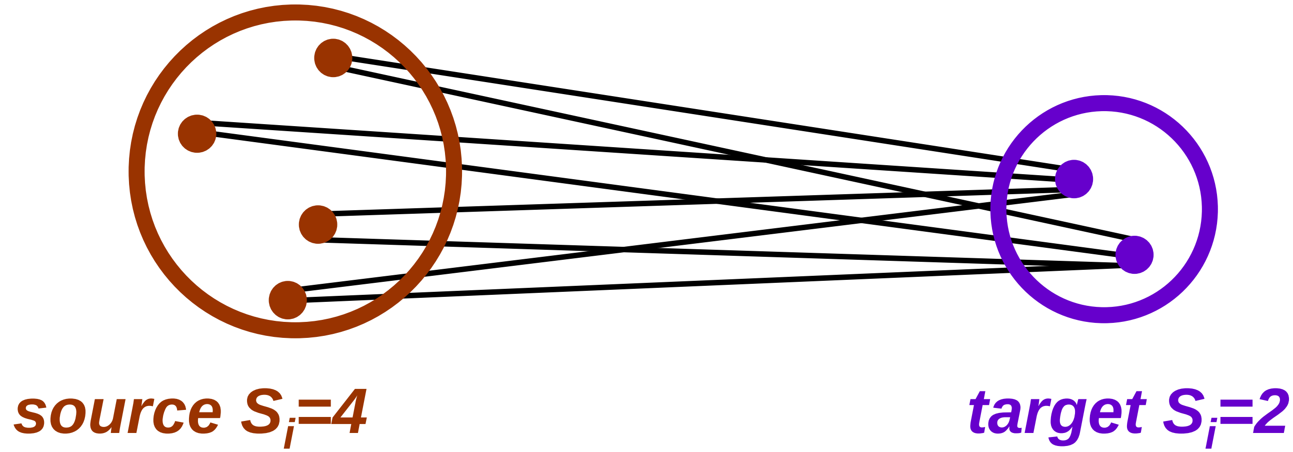
$$S = -\sum_{i,j} F_{ij}\left[\ln\left(F_{ij}\right) - 1\right] + \left\{\text{constraints}\right\}$$

- To fix total flow use $\mu\left(F_{\text{total}} - \sum_{i,j} F_{ij}\right)$

$\Rightarrow$ complete network, $\boldsymbol{F_{ij}}$ const.

Site Sizes

- Number of sources/targets for interaction at each site ***i*** is $\boldsymbol{S_i}$
- If each interaction equally likely,
  $\Rightarrow$ flow will scale with number of interactions $\boldsymbol{S_i S_j}$ between sites ***i*** and ***j***



Site Sizes

- Number of sources/targets for interaction at each site ***i*** is $\boldsymbol{S_i}$ e.g. equals population of site
- If each interaction equally likely then expect flow to scale with number of interactions between sites ***i*** and ***j*** is equal to $\boldsymbol{S_i S_j}$
- So maximise entropy

$$S = -\sum_{i,j} F_{ij} \left[\ln\left(\frac{F_{ij}}{S_i S_j} \right) - 1 \right] + \{\text{constraints}\}$$

- Total flow fixed $\Rightarrow \boldsymbol{F_{ij} \propto S_i\, S_j}$

Cost constraint – Simple Gravity Models

- Each trip from ***i*** to ***j*** costs $\boldsymbol{c_{ij}}$ typically function of distance $\boldsymbol{d_{ij}}$
- Total costs constrained to be ***C***

$$S = -\left\{ \sum_{i,j} F_{ij} \left[\ln\left(\frac{F_{ij}}{S_i S_j} \right) - 1 \right] \right\} + \gamma \left\{ C - \sum_{ij} c_{ij} F_{ij} \right\} + \mu \left(F_{\text{total}} - \sum_{i,j} F_{ij} \right)$$

- Parameter γ is Lagrange multiplier to enforce cost constraint

Deterrence function vs Cost function

Solutions given in terms of function of distance, the **deterrence function** ***f*(d_{ij})**

$$\boxed{F_{ij} = S_i S_j f\left(d_{ij}\right)}$$

Choosing deterrence function is equivalent to specifying total cost ***C***, Lagrange multiplier γ, and actual cost function c_{ij} in entropy function.

(Ignoring overall normalisation for simplicity)

SGM

Cost constraint – Simple Gravity Models

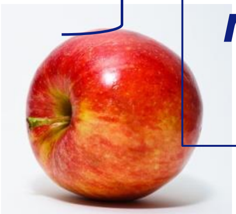
- Cost $c_{ij} = d_{ij}$
  $\Rightarrow$ exponential fall off

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

- Cost $c_{ij} = \ln(d_{ij})$
  $\Rightarrow$ power law fall off

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

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



- Cost $c_{ij} = 0 \quad d_{ij} < D$
  $c_{ij} = -\infty \quad d_{ij} > D$

= Threshold, Maximum Distance Network model

Input and Output Constraints

Common to know about output $\boldsymbol{A_i}$ and/or input $\boldsymbol{B_i}$ for each site $\boldsymbol{i}$

$$S = -\left\{ \sum_{i,j} F_{ij} \left[\ln\left(\frac{F_{ij}}{S_i S_j} \right) - 1 \right] \right\} + \gamma \left\{ C - \sum_{ij} c_{ij} F_{ij} \right\}$$

$$+ \underbrace{\left\{ \sum_i \alpha_i \left(A_i - \sum_j F_{ij} \right) \right\}}_{\textit{output constraints}} + \underbrace{\left\{ \sum_j \beta_j \left(B_j - \sum_i F_{ij} \right) \right\}}_{\textit{input constraints}}$$

Constraints:-
The Doubly Constrained Gravity Model

DCGM

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

$$F_{ij} = \underbrace{a_i b_j}_{\text{input/output constraints}} \underbrace{S_i S_j}_{\text{All pairs equally likely}} \underbrace{f\left(d_{ij}\right)}_{\text{"cost" constraint}}$$

input/output constraints

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

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

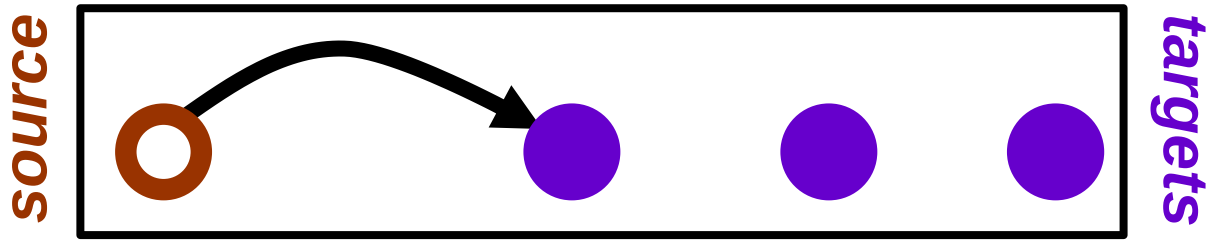
All pairs equally likely

"cost" constraint as **deterrence function** in terms of distance $\boldsymbol{d_{ij}}$ e.g.

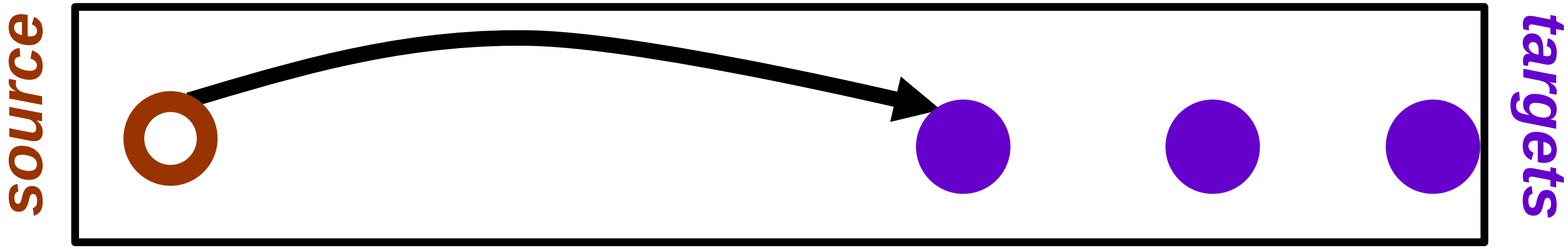
$$f\left(d_{ij}\right) = \frac{1}{\left(d_{ij}\right)^{\gamma}}$$

Intervening Opportunities Models [Stouffer 1940]

Actual physical distance is not important

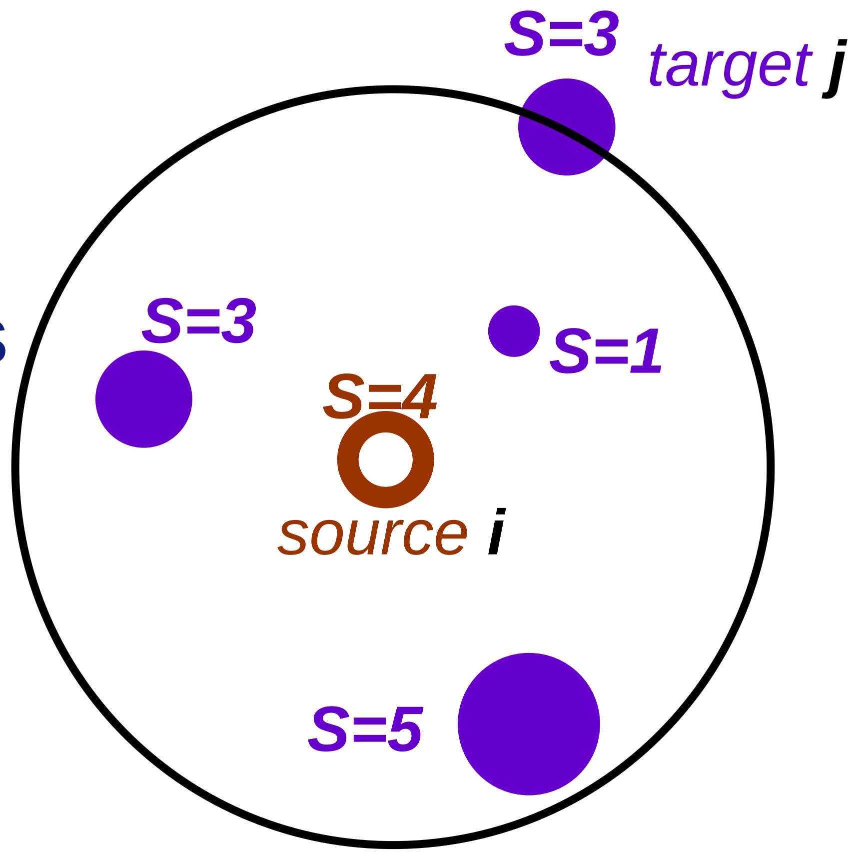


connection strength identical to



Intervening Opportunities

Define $\boldsymbol{D_{ij}}$ as number of all opportunities within circle radius $\boldsymbol{d_{ij}}$



$$\boldsymbol{D_{ij} = 16}$$

Intervening Opportunities as Entropy Maxima

- Just measure distances in terms of rank

Example:

- Number of opportunities at site ***i*** is $\boldsymbol{S_i}$
- Cost from ***i*** to ***j*** is $\boldsymbol{c_{ij} = D_{ij}}$
  = number of opportunities between ***i*** to ***j***
  when starting from ***i***

$$F_{ij} = S_i S_j \exp\left(-\gamma D_{ij}\right)$$

exactly as in simple gravity model

Radiation Model as Maximum Entropy

- $\boldsymbol{S_i}$ = opportunities at site ***i***
  $\boldsymbol{D_{ij}}$ = number of opportunities closer to or as close to ***i*** as ***j*** is
- Cost from ***i*** to ***j*** is $\boldsymbol{c_{ij} = \ln(D_{ij} D_{ij-1} / S_i)}$

$$\Rightarrow \quad F_{ij} = \frac{a_i S_i S_j}{\left(D_{ij} D_{ij-1}\right)^{\gamma}} \qquad \left(a_i\right)^{-1} = \sum_j \frac{S_j}{\left(D_{ij} D_{ij-1}\right)^{\beta}}$$

exactly as in output constrained gravity model

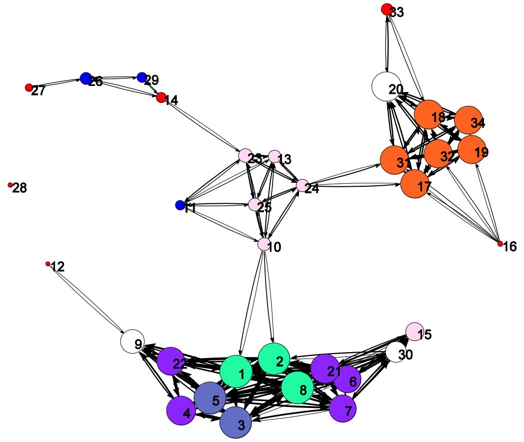
Original Radiation model interpretation

[Simini et al. 2012]

- Original presentation of model used ranked distances with record statistics
- Set γ ***=1*** (fixes total cost ***C***) and there are special algebraic properties leading to

$$F_{ij} = \underbrace{\frac{S_i}{1-\left(S_i / S_{\mathrm{total}}\right)}}_{\textit{1st denominator missing in Simini et al.}} \frac{S_i S_j}{\left(D_{ij} D_{ij-1}\right)}$$

MODELLING - Network Models



- Threshold models
- Maximum Entropy Models
  - Gravity Models
  - Rhill & Wilson
  - Radiation Model
- Stochastic Models

Stochastic Models

All previous models are **deterministic**

- one set of input values, one output model

Feature or Drawback?

Consider **stochastic** models

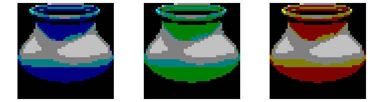
- **ERGM** (Exponential Random Graph Models) – usually network topology based
- **ariadne** **[Evans, Knappett, Rivers 2008+]**



Stochastic Model – ariadne

[Evans, Knappett and Rivers 2008+]

- Has intrinsic volatility set by `temperature' parameter
- Allows sites to vary in size in response to network connections
- Network will give an low value of a `cost' function
  - includes costs for sites and edges sizes and ascribes benefits to interactions



Optimisation of what for **ariadne**?

`Energy', resources

$$H =$$

Isolated sites have optimal size v_i = 0.5

$$-\kappa \sum_i 4 S_i v_i (1 - v_i)$$

Interactions (trade) bring benefits

$$-\lambda \sum_{i,j} (S_i v_i).e_{ij} V(d_{ij}/D).(S_j v_j)$$

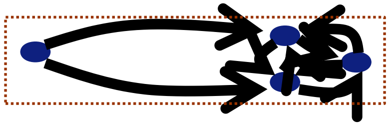
Increasing 'population' has a cost

$$+ j \sum_i S_i v_i$$

Each trade link has a cost

$$+ \mu \sum_{i,j} S_i v_i e_{ij}$$

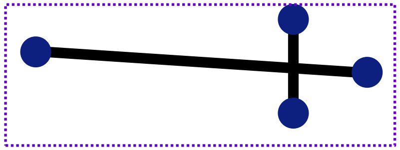
$$0 \leq \sum_j e_{ij} \leq 1 \qquad 0 \leq v_i$$

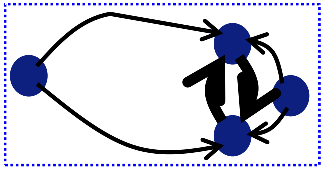


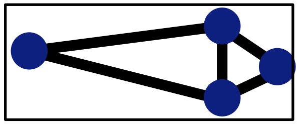
COMPARING NETWORKS

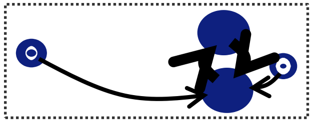
So many models, which do I use?

- Model vs Model
- Model vs Data
- Models from Data



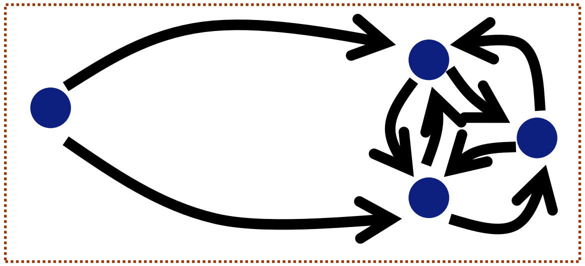


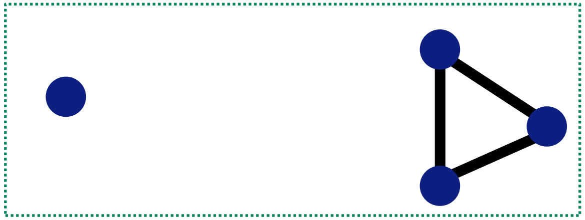


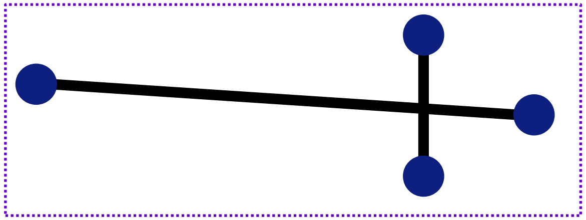


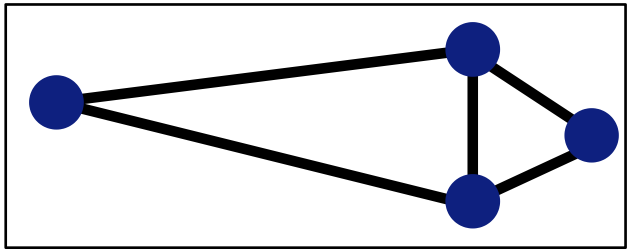
Comparing Network Models

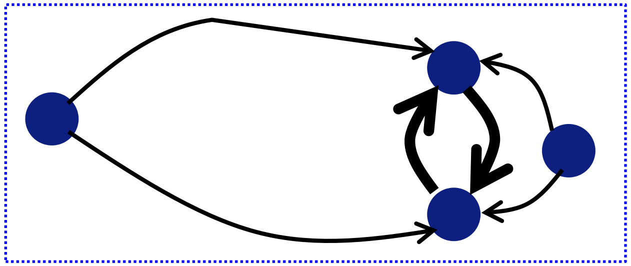
- Same arrangement of sites gives different networks
- How can we classify them?
- How can we compare them?

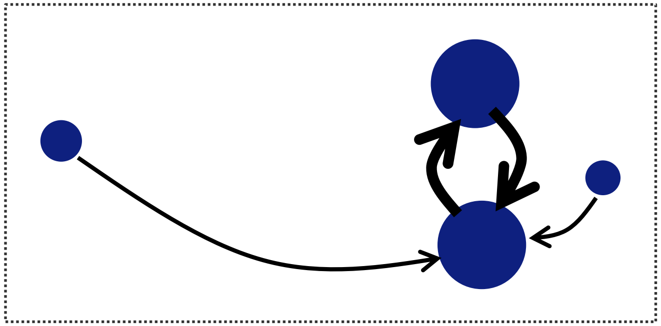












Geography and Interactions

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

- Distance
- Costs of Travel
- Input/Output constraints

[Evans 2014]

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

Average Weighted Distance

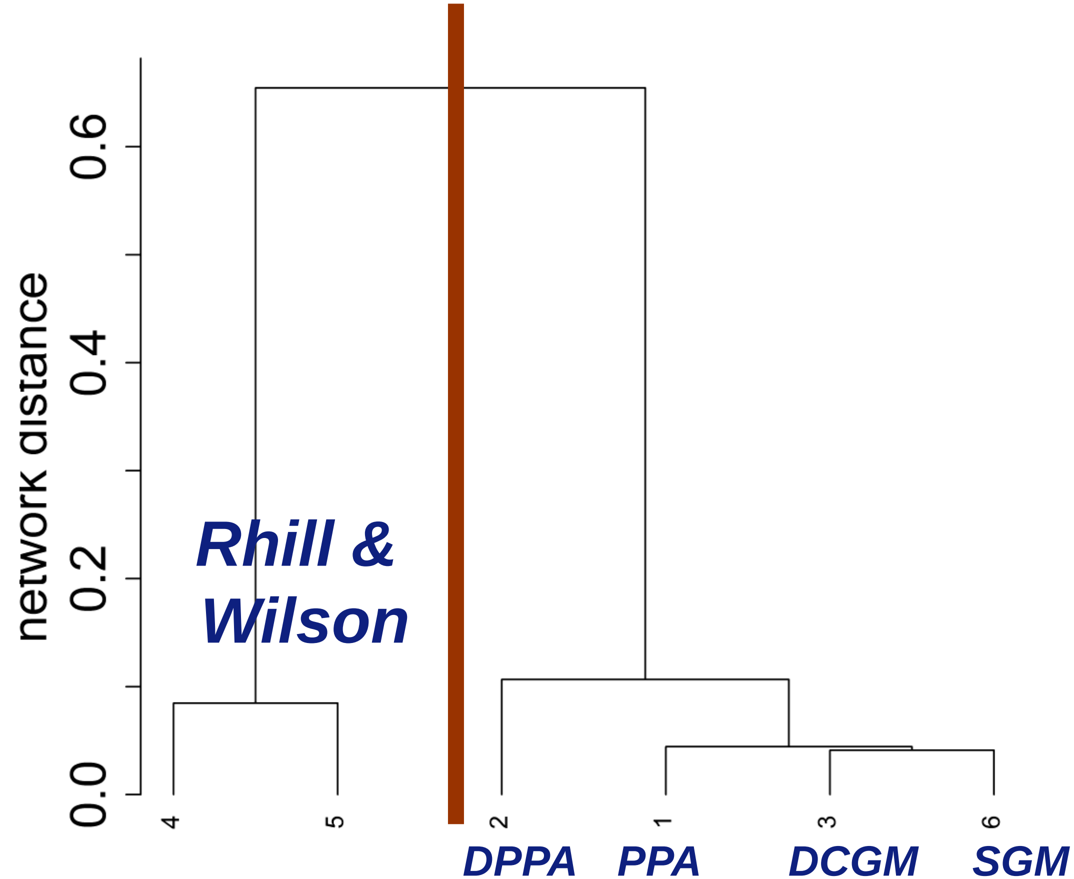
Suggest we compare networks with the same **average weighted distance**

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

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

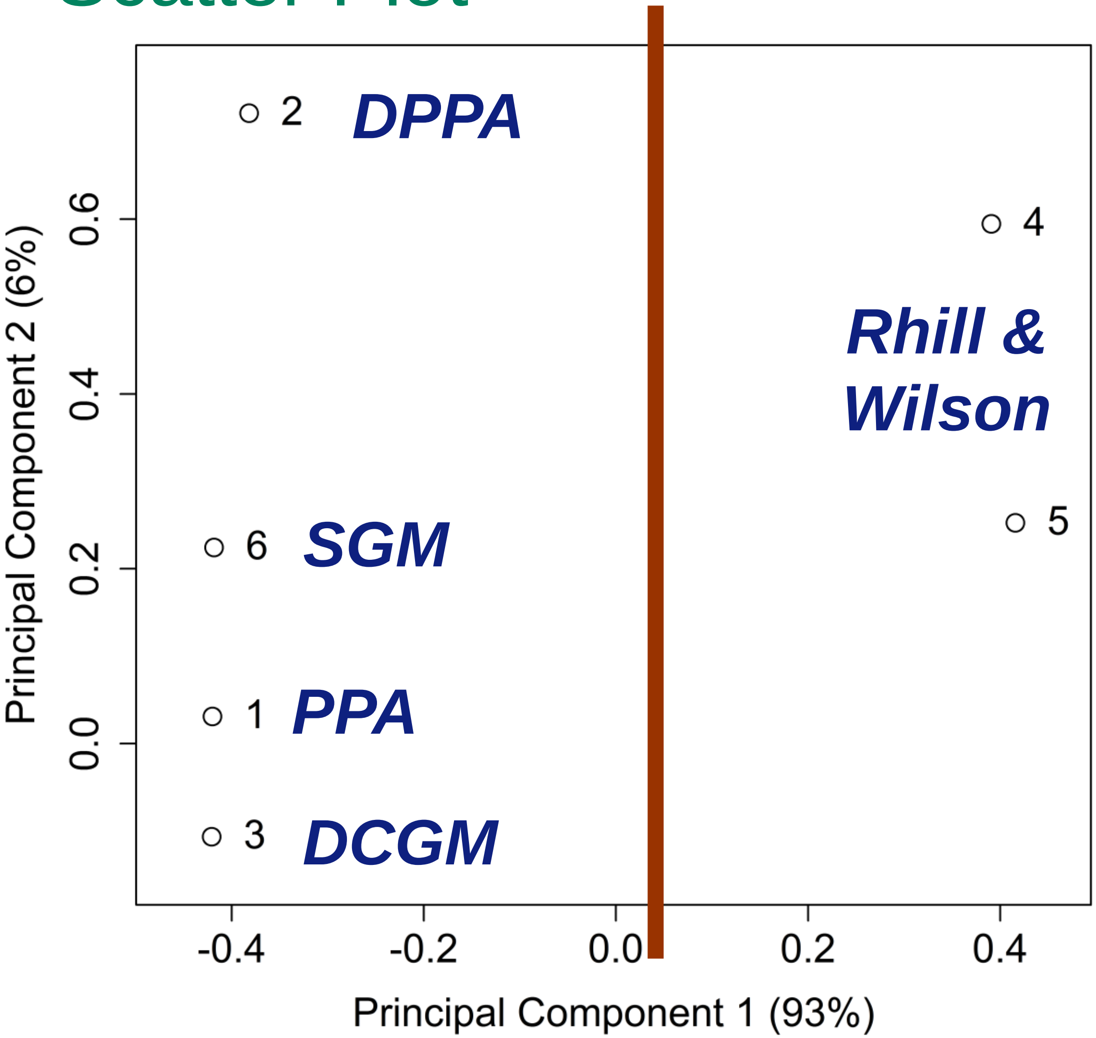
Hierarchical Clustering Dendrogram

- ***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 α=1.03
5	RWGM D=55 α=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 α=1.03
5	RWGM D=55 α=1.18
6	SGM D=75

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

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**

Model vs. Data

May have data on flows

If actual edge values central

- Model statistical fluctuations in flow values
- Use log liklihood to find best fit parameters for model
- Measure quality of fit using R^2, Akaike Information Criterion, etc

[Bamis, 2012; Bamis, TSE, Gastner in prep]

Negative Binomial on NY State commuting data

- Statistical model
  - Flow data $\boldsymbol{W_{ij}}$ is measurement of random variable, with negative binomial distribution of mean given by model $\boldsymbol{F_{ij}(\theta)}$ value, variance $\boldsymbol{F_{ij}(1+\phi F_{ij})}$

$$L\left(\vartheta \mid \{W_{ij}\}, \{F_{ij}\}\right) = \sum_{i,j} \frac{\Gamma\left(W + \phi^{-1}\right)}{\Gamma\left(W+1\right)\Gamma\left(\phi^{-1}\right)} \left(\frac{1}{1+\phi F}\right)^{\phi^{-1}} \left(\frac{\phi F}{1+\phi F}\right)^{W}$$

- Choose model parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$,to maximise log likelihood ***ln(L)***

Negative Binomial Statistical Model Results

Constraint	X^2 i	X^2 j	G^2	R^2	$R^2_{e^2}$	AIC	BIC
			Gravity				
Total Flow	4333 (0)	4157(0)	13658 (0)	0.80	0.22	3692196	3692231
Production	5170 (0)	4972(0)	10073 (0)	0.86	0.48	2474138	2474519
Attraction	3918 (0)	3674(0)	11360 (0)	0.91	0.71	1628281	1628662
Doubly	8670 (0)	8305 (0)	9684 (0)	0.95	0.96	751918	752640
			Radiation				
Total Flow	33510 (0)	33298(0)	16966 (0)	0.68	0.24	5919430	5919436
Production	24909 (0)	24617(0)	15741 (0)	0.75	0.20	4523422	4523774
Attraction	42114 (0)	41610(0)	28607 (0)	0.86	0.84	2559127	2559480
Doubly	16666 (0)	16369(0)	18191 (0)	0.91	0.89	1524763	1525462

Table 5.7: Statistics for goodness-of-fit and relative comparisons of gravity and radiation models on the state of N.Y. The X^2 test number of cells for i binning are 1722, for j 1482 and p-values from χ^2 are in parentheses

[Bamis, 2012]

Negative Binomial Statistical Model Results

- Adding constraints to models improves measures of fits to data
  - even after taking account of extra parameters
- Radiation model fits flows consistently worse than gravity model with similar constraints (say 5-10%)

Null Model from Data

#Region	Crete	Crete	Crete	Crete	Crete	Crete	Crete	Crete	Crete	Cyc
#Betweer	0.205633	0.618571	0.001354	0.029048	0.001982	0.424345	0.342305	0.546614	0.103939	0.2
#Betweer	20	4	32	28	31	11	14	7	23	
#Influence	1.67222	4.043742	0.473438	0.513033	0.47502	3.811917	3.032127	3.797388	0.863757	1.7
#Influence	20	7	31	29	30	9	13	10	25	
#Betweer	0.242486	0.786505	0.002722	0.078141	0.004266	0.596695	0.512073	1	0.205258	0.1
#Betweer	14	1	24	20	22	3	4	0	15	
#NBetwee	3.342369	7.281057	0.133503	1.176985	0.181936	3.092669	3.125422	7.03305	3.668492	1.9
#NBetwee	12	1	32	24	31	14	13	3	11	
#NBetwee	2.789859	5.636137	0.233714	2.071243	0.324462	2.843842	2.767707	9.853891	6.590892	1.1
#NBetwee	20	5	31	23	29	18	21	1	3	
#EDGE val	---	---	---	---	---	---	---	---	---	---
From/to	Knossos	Malia	Phaistos	Kommos	A.Triadha	P-kastro	Zakros	Gournia	Chania	Akr
Knossos	0	0.175	0	0	0	0.066	0.019	0.265	0.008	
Malia	0.212	0	0	0	0	0.15	0.096	0.15	0.001	
Phaistos	0	0	0	0	0	0.054	0.114	0.522	0	
Kommos	0	0	0	0	0	0.106	0.12	0.347	0	
A.Triadha	0	0	0	0.194	0	0.053	0.126	0.395	0	
P-kastro	0.016	0.259	0	0	0	0	0.231	0.167	0	
Zakros	0.009	0.53	0	0	0	0.151	0	0.083	0	
Gournia	0.051	0.181	0	0	0	0.278	0.13	0	0	
Chania	0.175	0.152	0	0	0	0	0	0.002	0	
Akrotiri	0.016	0.142	0	0	0	0.027	0.001	0.019	0	
Phylakopi	0	0.004	0	0	0	0	0	0	0.007	
Kastri	0	0	0	0	0	0	0	0	0.297	
Naxos	0	0	0	0	0	0	0	0	0	
Koa	0	0	0	0	0	0	0	0	0	

Can use data to provide a realistic deterrence function for Gravity models

$$\boxed{F_{ij} = S_i S_j f\left(d_{ij}\right)}$$

e.g. Set ***f(d)*** to match *average* flow **<*F*>** between sites distance ***(d ± Δ)*** apart

[Expert et al, 2011]

Null model to reveal hidden features

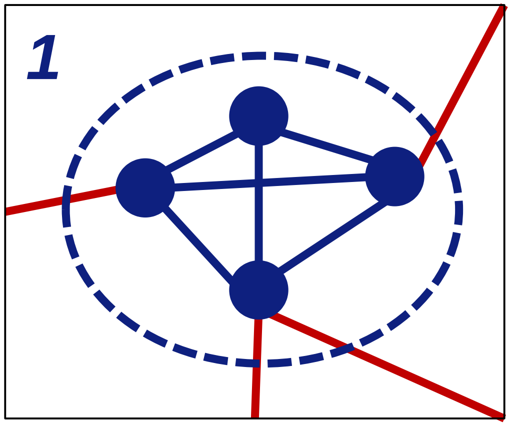
e.g. modularity finds clusters in data by comparing

1) expected number of links between sites within a community

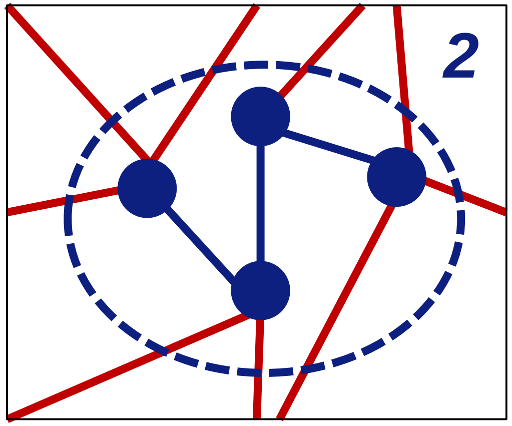
2) expected number of links between same sites in a **null model**

[Girvan & Newman 2002]

Original graph



vs



Random Graph degrees unchanged

Modularity and Null Models

All connections equally likely given site size

$$Q = \sum_{C} \sum_{i,j \in C} \left(\frac{F_{ij}}{W} - \frac{S_i S_j}{\underline{WW}} \right)$$

becomes

$$Q = \sum_{C} \sum_{i,j \in C} \left(\frac{F_{ij}}{W} - \underline{\frac{S_i S_j f(d_{ij})}{Z}} \right)$$

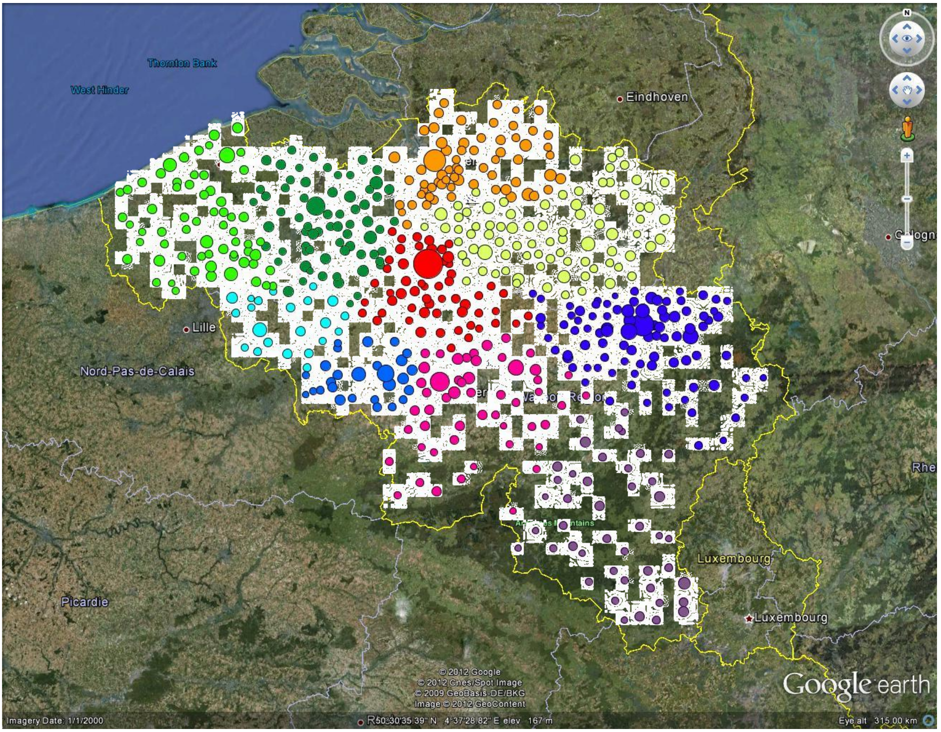
$$\left(\boldsymbol{W = \sum_{i,j} F_{ij}} \right)$$

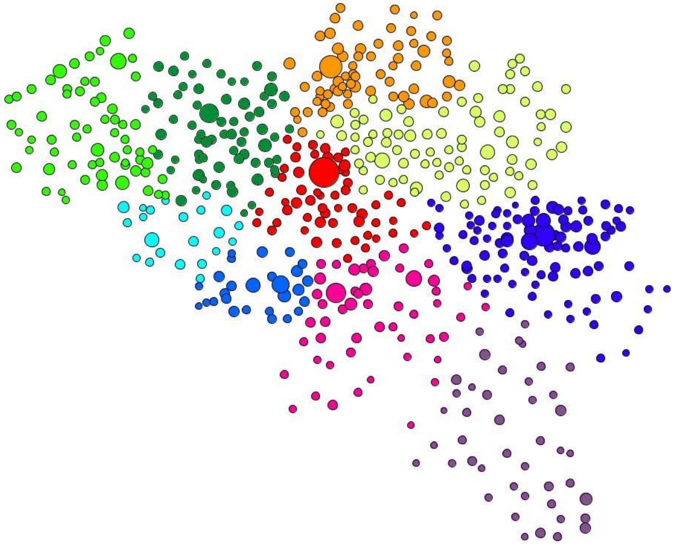
All connections equally likely given site sizes ***and*** separation

Null model to reveal hidden features

[Expert et al, 2011]

e.g if search for cluster in Belgian mobile phone data find most calls are local, clusters centred on main cities.

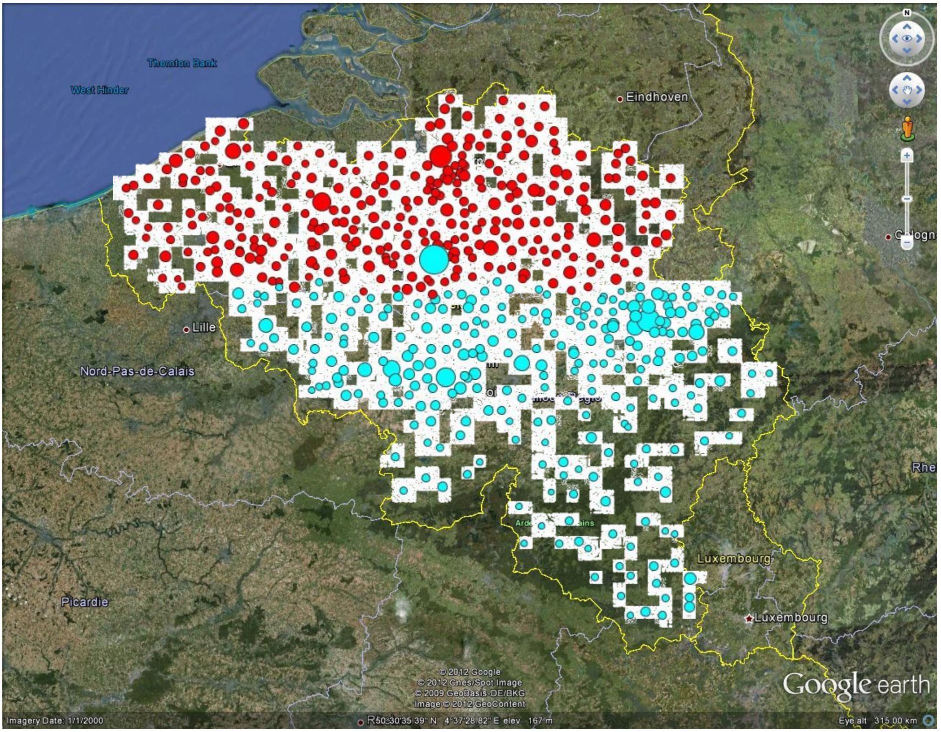


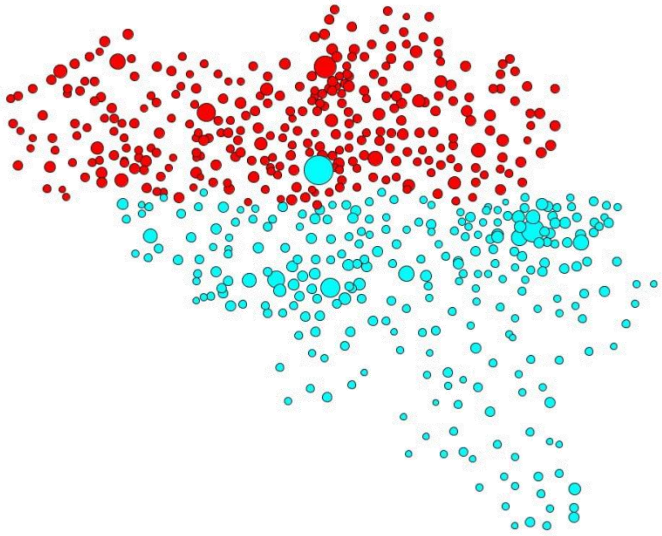


Null model to reveal hidden features

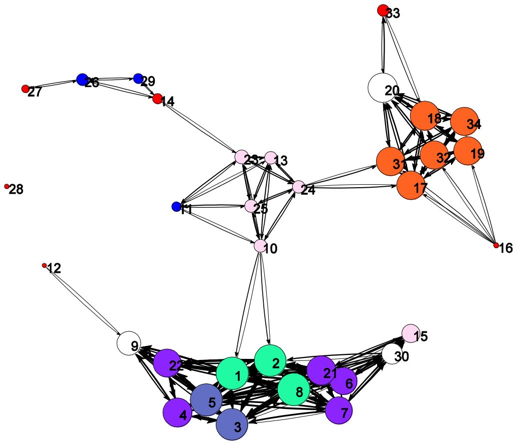
[Expert et al, 2011]

If compare call frequency against expectations from spatial model, clusters found match language divide





SUMMARY


33
27
26
29
14
20
18
34
31
32
19
23
13
28
24
17
11
25
10
16
12
15
9
30
22
2
21
1
6
8
5
4
7
3

Summary

- A few key choices and features common to all models
- Maximum entropy provides unifying if not always best description
- Wide range of models – choose to fit needs
- There are quantitative ways to compare different models

THANKS

Acknowledgements

Different parts done in collaboration with

- **Carl Knappett (Toronto)**
- **Ray Rivers (Imperial)**
- **Elias Bamis (Imperial)**
- **Michael Gastner (Budapest)**
- **Paul Expert (Kings)**
- **Renaud Lambiotte (Namur)**
- **Vincent Blondel (Louvain)**

Search for "Tim Evans networks" or look at

http://netplexity.org **@netplexity**

Bibliography

- Arcaute, E.; Hatna, E.; Ferguson, P.; Youn, H.; Johansson, A. & Batty, M. 2014. *City boundaries and the universality of scaling laws* Journal of The Royal Society Interface, 12, 20140745-20140745 http://dx.doi.org/10.1098/rsif.2014.0745 http://arxiv.org/abs/1301.1674
- Ilias Bamis, Constrained Gravity Models for Network Flows, MSc thesis, Imperial College, 2012
- Bevan, A.,Political Geography and Palatial Crete, *Journal of Mediterranean Archaeology,* **2010***, 23* .
- Barthélemy, M., Spatial Networks, *Physics Reports,* **2011***, 499*, 1-101
- Girvan, M. & Newman, M. E. J., Community structure in social and biological networks, PNAS*,* **2002***, 99*, 7821-7826
- P. Kaluza, A. Kölzsch, M.T. Gastner and B. Blasius, The complex network of global cargo ship movements, *J. R. Soc. Interface,* **2010,** 7, 1093-1103 [doi: 10.1098/rsif.2009.0495]
- KNAPPETT, C., EVANS, T. & R. RIVERS, 2008. Modelling maritime interaction in the Aegean Bronze Age, *Antiquity* 82: 1009-24.
- Knappett, C.; Evans, T. & Rivers, R., The Theran eruption and Minoan palatial collapse: new interpretations gained from modelling the maritime network, *Antiquity,* **2011***, 85*, 1008-1023
- Penrose, M., *Random Geometric Graphs (OUP, 2003).*
- Renfrew, C. & Level, E. Exploring dominance: predicting polities from centres, in *Transformations: Mathematical Approaches to Culture Change,* Renfrew, A. & Cooke, K. *(Eds.), Academic Press, London,* **1979**, 145-67
- 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.
- Simini, F.; Gonzalez, M. C.; Maritan, A. & Barabasi, A.-L., A universal model for mobility and migration patterns, *Nature,* **2012***, 484*, 96-100
- Srinivasa, S. and Haenggi, M., Distance distributions in finite uniformly random networks: Theory and applications, *IEEE Transactions on Vehicular Technology,* **2010**, 59, 940–949.
- Stouffer, Samuel A. , Intervening Opportunities: A Theory Relating to Mobility and Distance, *American Sociological Review* 1940, **5**, 845–867. doi:10.2307/2084520.
- A.G.Wilson, A statistical theory of spatial distribution models, Trans. Res. 1 (1967) 253-269

Bibliography (Others)

- 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. C. F., 2013. Religious Networks in the Roman Empire: the Spread of New Ideas, CUP.
- 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.
- 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

Bibliography (TSE)

- Evans, T.S., 2004. *Complex Networks*, Contemporary Physics*, 45*, 455-474 [doi: 10.1080/00107510412331283531]
- 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].
- Evans, T.S., Knappett, C., & R.J. Rivers, 2012. *Interactions in Space for Archaeological Models*, Advances in Complex Systems [doi: 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: Oxford University Press (in press).
- Expert, P.; Evans, T. S.; Blondel, V. D. & Lambiotte, R. 2011. *Uncovering space-independent communities in spatial networks, PNAS, 108*, 7663-7668 [doi: 10.1073/pnas.1018962108]
- 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*.*