Epidemic networks - modelling the relevant structure

A.P. Ford

Warwick Mathematics Institute
University of Warwick

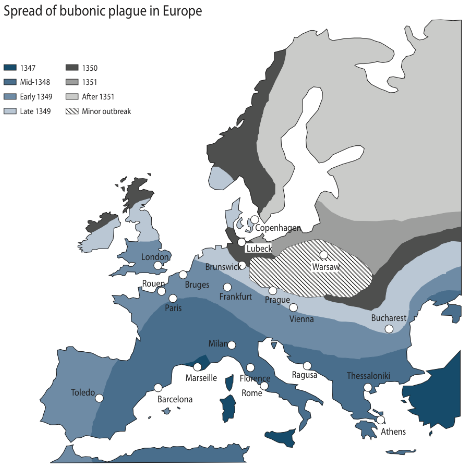
EMOTIONS 2014, Lyon.

Outline

- **Motivation**
- **Random graph models**
  - how do we generate ‘realistic’ networks?
- **Bi-partite graph models**
  - **Hypergraphs**

Motivation

Historical epidemics were wave-like / spatial:

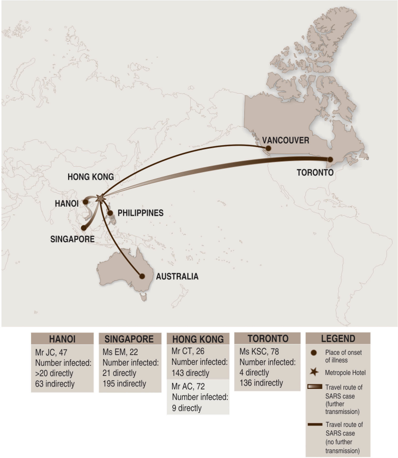


(picture from WHO)

Motivation

Modern epidemics are spatially heterogeneous but transmission rates are not a simple function of Euclidian distance – with all that implies for control.

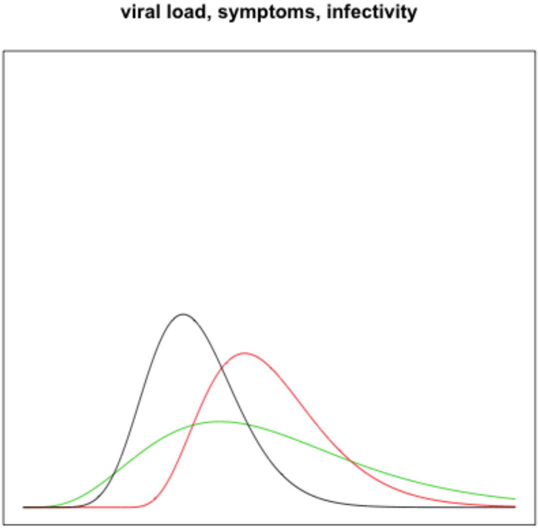
Spread of virus from the Metropole Hotel



(picture from WHO)

Approximations in Models

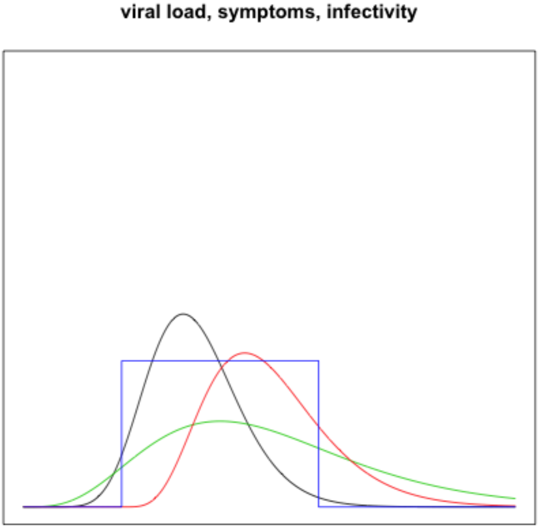
Approximation of infection process in SEIR models.



ARWICK

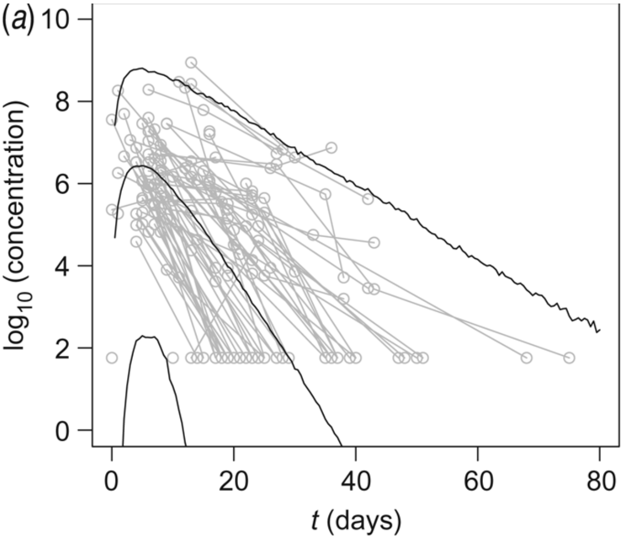
Approximations in Models

Approximation of infection process in SEIR models.



ARWICK

Norovirus Shedding data



†

WARWICK

† Teunis 2014, http://dx.doi.org/10.1017/S095026881400274X

Contact Hetrogeneity

- consider individuals $i, j, \ldots \in \{1, \ldots, N\}$;
- λ_{ij} is the infection rate between i and j
  - when i is infectious and j is susceptible.

WARWICK

Contact Hetrogeneity

- consider individuals $i, j, \ldots \in \{1, \ldots, N\}$;
- λ_{ij} is the infection rate between i and j
  - when i is infectious and j is susceptible.
- $\lambda_{ij} = d(i,j)\rho(i|x_i)\psi(j|x_j)$

WARWICK

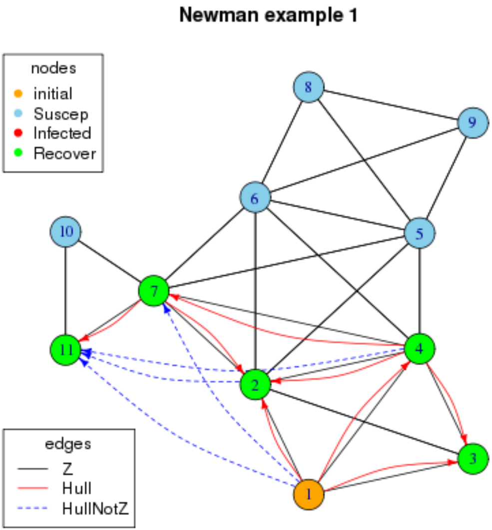
Networks: Definitions

- consider individuals $i, j, \ldots \in \{1, \ldots, N\}$; let $G_{ij} = 1$ if i and j make contact capable of spreading disease.
- $\mathbf{G} = (G_{ij})$ is the *adjacency matrix* of a network / graph $G \in \mathcal{G}$, where $\mathcal{G}$ is the set of all undirected graphs with N nodes.
- $k_i := \sum_j G_{ij}$ is the *degree* of node i
- Individual i is in disease state $X_i(t)$, a discrete random variable; for e.g. network SIR dynamics we have:

$$\begin{aligned} \mathbb{P}[X_i(t+\delta) = I | X_i(t) = S] &= \sum_j \tau G_{ij} \mathbb{I}_{\{X_j(t)=I\}} \delta + o(\delta) \ ; \\ \mathbb{P}[X_i(t+\delta) = R | X_i(t) = I] &= \gamma\delta + o(\delta) \ . \end{aligned} \tag{1}$$

WARWICK

Inference on Network: from observed epidemic



WARWICK

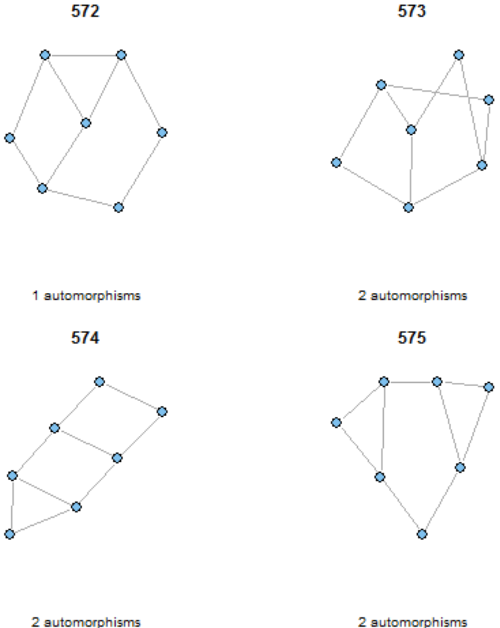
Networks: Models

- Known or random
- Maths or simulation

The degree distribution $P(k)$ the probability a randomly chosen node has degree k is the major factor affecting epidemics on graphs.

- Erdös Renyi
- Configuration model
- Small world

Graphs with same degree distribution



Graph - Automorphisms

atlas number	572	573	574	575	576
automorphisms	1	2	2	2	4
number isomorphic	5040	2520	2520	2520	1260
clustering	0.6052	0.6225	0.5762	0.5905	0.5164
transitivity	.2	0	.2	.4	.6

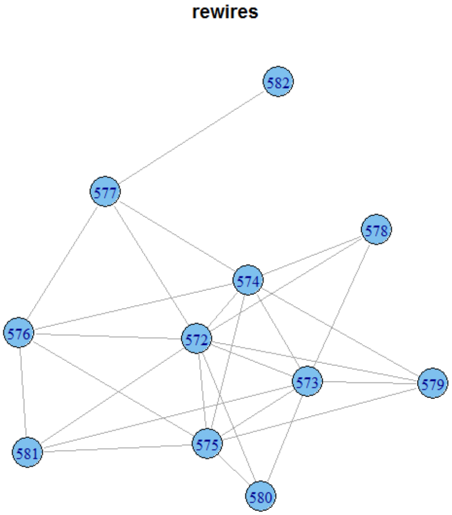
number	577	578	579	580	581	582
auto.	4	4	4	6	6	144
isomor.	1260	1260	1260	840	840	35
cluster	0.5701	0.6082	0.6082	0.5861	0.5861	0.2286
transit	.4	0	.2	.2	0	1.

all graphs with 7 nodes; 3 nodes of degree 2 and 4 nodes degree 3

WARWICK

Rewiring algorithm

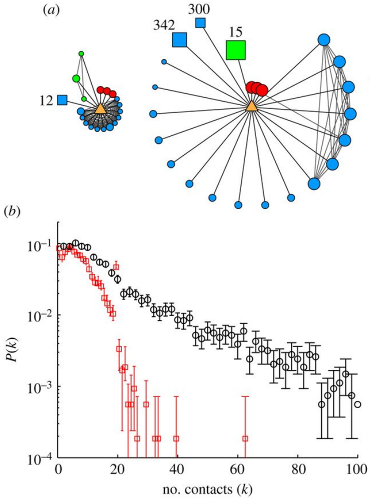
swap pairs of edges



WARWICK

Networks: Heterogeneity

Measurement of direct contact patterns shows complex local structure and significant heterogeneity in degree[†].

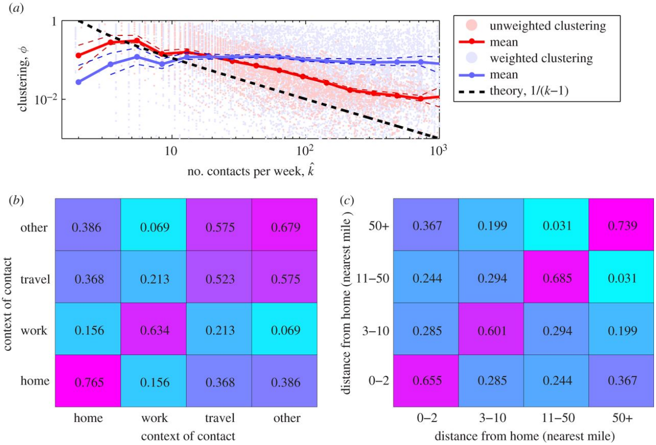


[†] L. Danon, **T. House**, J. M. Read and M. J. Keeling, "Social encounter networks: collective properties and disease transmission," Journal of the Royal Society Interface. **9**:76 (2012) 2826-2833.

WARWICK

Networks: Clustering

Measurement also shows significant transitivity / clustering, generated more by context than space[†].



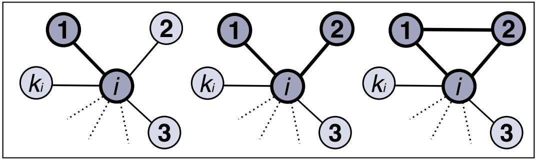
[†] L. Danon, **T. House**, J. M. Read and M. J. Keeling, “Social encounter networks: collective properties and disease transmission," Journal of the Royal Society Interface. **9**:76 (2012) 2826-2833.

Networks: Heterogeneity and Clustering

Suppose we want to generate full networks with given heterogeneity and clustering; a natural approach is to promote the G_{ij} to random variables. We want to specify:

$$\begin{aligned} M[G] &:= \sum_{i,j} G_{ij} = N\langle k\rangle \quad \text{(total links)}, \\ L[G] &:= \sum_{i,j,k\neq j} G_{ij}G_{ik} = N\langle k(k-1)\rangle \quad \text{(lines)}, \\ T[G] &:= \sum_{i,j,k} G_{ij}G_{jk}G_{ki} = \quad \text{(triangles)}. \end{aligned} \tag{2}$$

Note, these are not independent:



WARWICK

Networks: Measures

We will define a probability measure using a Hamiltonian framework so that the probability given to $G \in \mathcal{G}$ is

$$\pi(G|\boldsymbol{\theta}) = \frac{\mathrm{e}^{-H(G;\boldsymbol{\theta})}}{\mathcal{Z}_{\boldsymbol{\theta}}}\ \text{, where} \quad \mathcal{Z}_{\boldsymbol{\theta}} = \sum_{G\in\mathcal{G}} \mathrm{e}^{-H(G;\boldsymbol{\theta})}\ . \tag{3}$$

We will also write $\tilde{\pi}$ for the ensemble probability of a network taking a value of properties $\boldsymbol{F}[G] = (F_a[G])$

$$\tilde{\pi}(\boldsymbol{f}|\boldsymbol{\theta}) = \sum_{G\in\mathcal{G}} \pi(G|\boldsymbol{\theta}) \prod_a \mathbb{I}_{\{f_a = F_a[G]\}}\ . \tag{4}$$

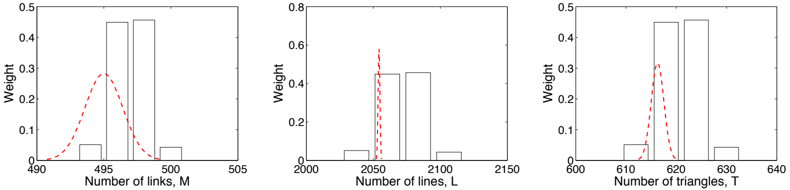
It is easy to sample according to π in a Monte Carlo scheme, but hard to sample according to $\tilde{\pi}$ for combinatorial reasons.

WARWICK

Networks: A heterogeneous clustered measure

The following measure works if parameters are tuned appropriately:†

$$H(G;\boldsymbol{\theta},\boldsymbol{\beta}) = \beta_m \left(M[G] - \theta_m N(N-1)\right)^2$$
$$+\beta_l \left(L[G] - M[G]((\theta_l - 1) + N^{-1}M[G])\right)^2 + \beta_t \left(T[G] - \theta_t L[G]\right)^2 \; .$$



(Red dashed line – target; Black histograms – simulations)

WARWICK

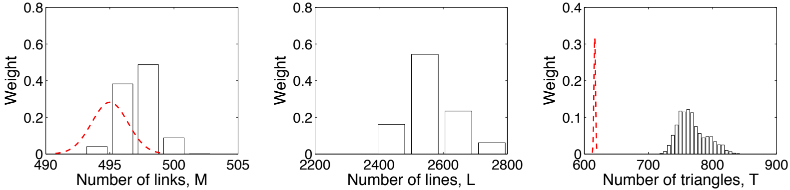
† **T. House**, "Heterogeneous clustered random graphs," Europhysics Letters **105** (2014) 68006.

Networks: A heterogeneous clustered measure

Removing just one term causes failure:

$$H(G;\boldsymbol{\theta},\boldsymbol{\beta}) = \beta_m \left(M[G] - \theta_m N(N-1)\right)^2$$

$$\cancel{+\beta_l \left(L[G] - M[G]((\theta_l - 1) + N^{-1}M[G])\right)^2} + \beta_t \left(T[G] - \theta_t L[G]\right)^2 \; .$$



(Red dashed line – target; Black histograms – simulations)

WARWICK

ERGM and Phase transitions in graphs

The theory explains a host of difficulties encountered by applied workers: many distinct models have essentially the same MLE, rendering the problems "practically" ill-posed. We give the first rigorous proofs of "degeneracy" observed in these models. showing that for many models, the extra sufficient statistics are useless: most realizations look like the results of a simple Erdős-Rényi model. We also find classes of models where the limiting graphs differ from Erdős-Rényi graphs. A limitation... it works only for dense graphs. †

† Chatterjee and Diaconis, 2011, Annals of Statistics

Motivation for bipartite models

- consider individuals $i, j, \ldots \in \{1, \ldots, N\}$;
- λ_{ij} is the infection rate between i and j
  - when i is infectious and j is susceptible.
- $\lambda_{ij} = d(i,j)\rho(i|x_i)\psi(j|x_j)$
- Ebola know the village and household level matter
- 3 level schools, workplace, household

WARWICK

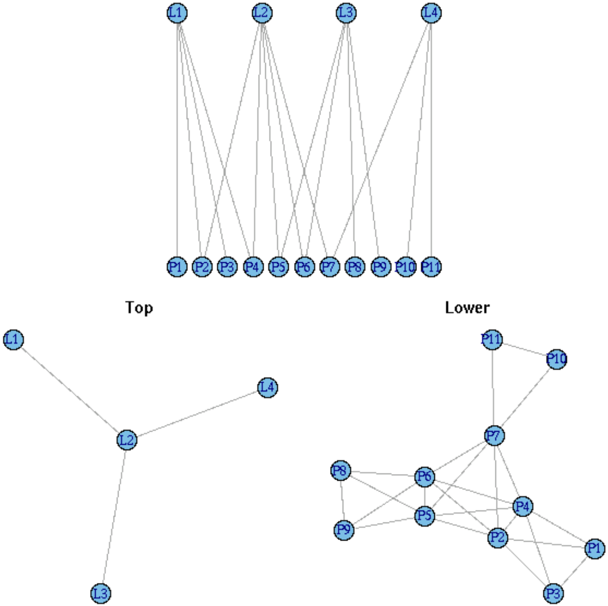
Example bipartite graph

A bipartite graph $G = (U, V, E)$ is two disjoint sets U and V comprising the nodes or vertices[†], and a set of edges E where each edge is a pair of nodes (u, v), $u \in U, v \in V$. In general the two sets U and V can be of the same type so that all vertices in U and V are in some larger set of vertices, here U and V are distinct with U representing individuals who may become infected and V an abstract set of possible contacts, which may include physical premises such as schools, houses or work places and can include a temporal aspect.
A convenient representation is the adjacency matrix $\mathbf{A} = (a_{ij}, i \in U, j \in V)$ where $a_{ij} = 1$ if and only if $(i, j) \in E$ and $a_{ij} = 0$ otherwise.

WARWICK

[†]the terms are used interchangeably

Example bipartite graph



Bipartite representation of a household model

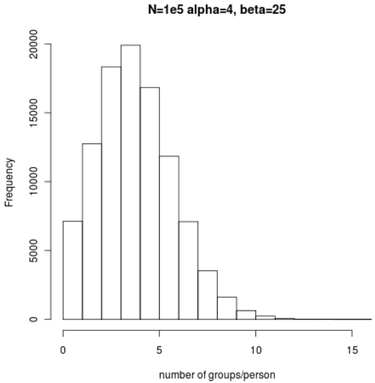
λ_G	λ_H	λ_H	λ_H	λ_H
1	1			
1	1			
1		1		
1		1		
1		1		
1			1	
1			1	
1			1	
1				1
1				1
1				1
1				1

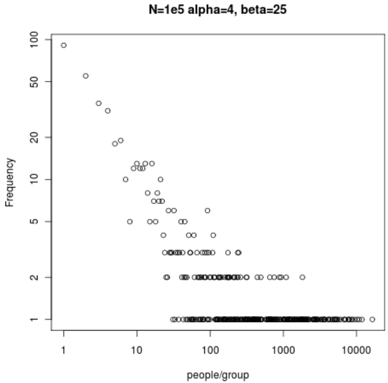
other standard models can also be represented

Random bipartite graphs

- Random Intersection Graphs
  - Theoretical results but not realistic
- Hypergraphs
  - Theoretical results
- Indian Buffet Process
  - Realistic but theory challenging

Example degree distributions of IBP



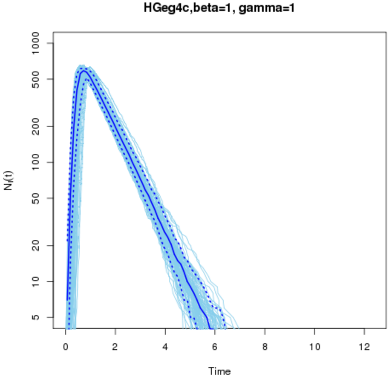


Distribution of marginal sums for an example IBP, $N = 10^5$, $\alpha = 4$, $\beta = 25$.

The left hand plot is a standard histogram, the right hand a count of counts.

WARWICK

Example epidemics on Hypergraphs



WARWICK

Conclusions

- Bipartite Graphs are a useful model
- Lots of interesting problems in distributions of graphs and hypergraphs

WARWICK

Acknowledgements / Thanks

- WIDER (environment)
- EPSRC (my funding)
- The organisers, ECOFECT and your sponsors

WARWICK