#### **Epidemics on random networks with household structure**

#### Frank Ball

Frank.Ball@nottingham.ac.uk

#### University of Nottingham

Workshop on Networks: Stochastic Models for Populations and Epidemics, ICMS

Edinburgh, 12 September 2011

Joint work with David Sirl (Loughborough University) and Pieter Trapman (Stockholm University)

Research supported by EPSRC

## **Structured-population epidemic models**

- Classical epidemic models usually assume a homogeneously mixing population of homogeneous individuals.
- Important to include individual heterogeneities and social structure so that models better reflect reality.
- Considerable recent interest in models for epidemics among structured populations, which incorporate realistic departures from homogeneous mixing whilst maintaining mathematical tractability.
- Two main classes of structured population models are random network models and household models.
- In this talk we present and investigate a model that incorporates both of these features.

#### **Outline of talk**

- Model
- Threshold behaviour; probability and final outcome of major epidemic
- Numerical illustrations
- Comparison to standard network model
- Limitations/extensions of model
- Concluding comments

### **Population structure**

- Population of N individuals partitioned into m households, of which  $m_n$  are of size n (n = 1, 2, ...).
- Configuration model for network of possible global contacts
  - Assign each individual a number of 'half-edges' independently according to the random variable *D* which describes the degree distribution. (Repeat until total number of half-edges is even.)
  - Pair up half-edges uniformly at random to form global network.
  - Density of imperfections (parallel edges/ self-loops between individuals or households) tends to 0 as  $m \to \infty$  if  $\sigma_D^2 < \infty$  and  $\sigma_H^2 < \infty$ .

(Ball, Sirl and Trapman (2009,2010))

## **Example of (very small) population**



## **Epidemic model**

- **SIR** (susceptible  $\rightarrow$  infective  $\rightarrow$  removed).
- Infectious periods iid according to a random variable I having an arbitrary but specified distribution.
- Whilst infectious, individuals contact each of their local (global) neighbours independently at rate  $\lambda_{\rm L}$  ( $\lambda_{\rm G}$ ).
- Single initial infective chosen uniformly at random from entire population.

(Ball, Sirl and Trapman (2009,2010))

## **Threshold parameter** $R_*$



 $\blacksquare$   $R_* =$  mean number of global neighbours infected by a typical single-household epidemic

$$R_* = \sum_{n=1}^{\infty} \tilde{\rho}_n \mathbf{E}[\tilde{C}^{(n)}],$$

#### where

- $\tilde{\rho}_n = \frac{nm_n}{N} = P(\text{randomly chosen person lives in a household of size } n),$ 
  - $\tilde{C}^{(n)}$  = number of global neighbours infected by a typical [size-n] single-household epidemic with one initial infective.

 $P(major outbreak) > 0 \iff R_* > 1.$ 

# **Calculation of** $E[\tilde{C}^{(n)}]$

- $E[\tilde{C}^{(n)}] = E[C_0] + E[T^{(n)}]E[C_1],$ 
  - $T^{(n)} = \text{final size of single-household epidemic,}$
  - $C_0$  and  $C_1$  are the numbers of global neighbours infected by the primary and typical secondary infective in the household.
- $C_i \sim \operatorname{Bin}(K_i, 1 \exp(-\lambda_{\mathrm{g}}I_i)) \implies \operatorname{E}[C_i] = \operatorname{E}[K_i]p_{\mathrm{g}},$ 
  - $K_0$  and  $K_1$  are the numbers of susceptible global neighbours of primary and typical secondary case,
  - $p_{\rm G} = 1 {\rm E}[\exp(-\lambda_{\rm G}I)]$  is the probability an infective infects a given susceptible global neighbour.
- For initial generation,  $K_0, K_1 \sim D$ . For all subsequent generations,  $K_0 \sim \tilde{D} - 1$ , where  $\tilde{D}$  is the degree of a typical global neighbour  $[P(\tilde{D} = k) = \mu_D^{-1} k P(D = k) \ (k = 1, 2, ...)]]$ , and  $K_1 \sim D$ .

#### **Threshold parameter** $R_*$

$$R_* = \sum_{n=1}^{\infty} \tilde{\rho}_n \operatorname{E}[\tilde{C}^{(n)}]$$
  
= 
$$\sum_{n=1}^{\infty} \tilde{\rho}_n \left( \operatorname{E}[C_0] + \operatorname{E}[C_1]\operatorname{E}[T^{(n)}] \right)$$
  
= 
$$\sum_{n=1}^{\infty} \tilde{\rho}_n \left( \mu_{\tilde{D}-1} + \mu_D \mu_{T^{(n)}} \right) p_{\mathrm{G}}$$
  
= 
$$\left( \mu_D (\mu_T + 1) + \frac{\sigma_D^2}{\mu_D} - 1 \right) p_{\mathrm{G}},$$

where  $\mu_T = \sum_{n=1}^{\infty} \tilde{\rho}_n \mu_{T^{(n)}}$  is the size-biased mean withinhousehold final size.  $[\mu_{\tilde{D}} = \mu_D^{-1} \sigma_D^2 + \mu_D]$ 

## **Early stages of epidemic**

Approximate process of infected households by a (forward) branching process  $BP(1, C, \tilde{C})$  having one initial ancestor, and offspring distribution *C* for the initial generation and  $\tilde{C}$  in subsequent generations. (Exact as  $m \to \infty$ ).

- The probability of a major outbreak,  $p_{maj}$  is approximated by the probability  $BP(1, C, \tilde{C})$  avoids extinction.
- P<sub>maj</sub> = 1 − f<sub>C</sub>(σ), where σ is the smallest solution of  $f_{\tilde{C}}(s) = s$  in [0, 1] and e.g.  $f_{\tilde{C}}(s) = E[s^{\tilde{C}}]$ .
- Unless the infectious period *I* is constant, calculation of the PGFs  $f_C$  and  $f_{\tilde{C}}$  is complicated by dependencies between the numbers of global infections made by infectives in the same household but can be done using Ball and O'Neill's (1999) final state random variables.

### **Example network**



## **Digraph of potential infections**



## **Susceptibility set, Example 1**



## **Susceptibility set, Example 2**



## **Final size of major outbreak**

- Expected proportion of population ultimately infected by a major outbreak, z say, is given by the probability that a typical susceptible, i say, is ultimately infected.
- Approximate *i*'s susceptibility set by a households based (backward) branching process  $BP(1, B, \tilde{B})$ .
- $P(i \text{ ultimately infected}) = P(BP(1, B, \tilde{B}) \text{ avoids extinction}).$
- $z = 1 f_B(\xi)$ , where  $\xi$  is the smallest solution of  $f_{\tilde{B}}(s) = s$  in [0, 1].
- Distribution of within-household final size (WHFS) in the event of a major outbreak is available. If  $(z, \lambda_L)$  and the distribution of *I* are fixed, then the distribution of WHFS is invariant to the degree distribution *D*.

#### **Simulations of final size**



Histograms of relative final sizes from 10,000 simulations of the model with  $H \sim (0.2, 0.25, 0.25, 0.25, 0.04, 0.01), \lambda_{\rm L} = 1, \lambda_{\rm G} = 1/10, D \sim \text{Poi}(8)$ and  $I \sim \text{Gamma}(3, 1/3)$  on networks of 50 and 150 households.

## Simulation results for $p_{maj}$ and z



Comparison of simulation-based estimates of probability  $p_{maj}$  and expected relative final size z of a major outbreak for finite populations with asymptotic results (horizontal lines), based on 10,000 simulations for each m.

 $[D \sim \text{Pow}(k_*, a) \text{ means } P(D = k) \propto (\max(k_*, k))^{-a} \ (k = 1, 2, ...).]$ 

## **Effect of degree distribution**, *D*



Dependence of probability of a major outbreak  $p_{maj}$  on degree distribution D. Other parameters are  $H \equiv 3$ ,  $I \equiv 1$ ,  $\lambda_L = 1$ ,  $\lambda_G = 1/10$ .

[Heavy(3/2) is Pow( $k_*, 3/2$ ) with varying  $k_*$ . The mass function of HeavyC(3) is  $P(D = k) \propto k^{-3} \exp(-k/\kappa)$  (k = 1, 2, ...) with varying  $\kappa$ .]

## **Comparison to standard network model**

- Standard network model (see e.g. Newman(2002)) is obtained when all households have size 1.
- In network household model (NHM) let *H* denote the household size distribution and *H* denote the size-biased household size distribution.  $[P(\tilde{H} = n) = \mu_H^{-1} n P(H = n) \ (n = 1, 2, ...).]$
- Total degree (local + global) of individual in NHM  $\sim Q \stackrel{\mathcal{D}}{=} \tilde{H} 1 + D.$
- Assume  $\lambda_{\rm L} = \lambda_{\rm G} = \lambda$  and compare critical value of  $\lambda$  with that of standard network model (SNM) with degree distribution Q.

#### **Comparison to standard network model**



Plot of critical infection rate  $\lambda_{\rm L} = \lambda_{\rm G} = \lambda$  for models with  $H \sim {\rm Poi}^+(\mu)$  and  $D \sim {\rm Poi}(10 - \mu)$  (so Q is always  ${\rm Poi}(10)$  and  $\mu = 0$  gives SNM) when  $I \sim {\rm Gamma}(3, 1/3)$ .

## Clustering



Clustering measure

 $\mathcal{C}^{(m)} = \frac{\text{total number of ordered triangles in network}}{\text{total number of ordered triples in network}}$ 

Proportion of triangles not entirely in the same household tends to 0 as number of households  $m \to \infty$ , so

$$\mathcal{C}^{(m)} \approx 1 - \frac{\mathrm{E}[H\{2D(H-1) + D(D-1)\}]}{\mathrm{E}[H(D+H-1)(D+H-2)]}$$

## Effect of clustering when $\lambda_{\rm L} = \lambda_{\rm G}$



Plots of threshold parameters and the probability  $p_{maj}$  and expected relative final size z of a major outbreak for networks with  $H \sim \text{Poi}^+(\mu)$  and  $D \sim \text{Poi}(10 - \mu)$ , so Q is always Poi(10) and  $\mathcal{C}^{(m)} \approx (\mu/10)^2$ . Other parameters are  $I \sim \text{Gamma}(3, 1/3)$  and  $\lambda_{\text{L}} = \lambda_{\text{G}} = 1/5$ .

## **Effect of clustering when** $\lambda_{\rm L} > \lambda_{\rm G}$



Plots of threshold parameters and the probability  $p_{maj}$  and expected relative final size z of a major epidemic for networks with  $H \sim \text{Poi}^+(\mu)$  and  $D \sim \text{Poi}(10 - \mu)$ , so Q is always Poi(10) and  $\mathcal{C}^{(m)} \approx (\mu/10)^2$ . Other parameters are  $I \sim \text{Gamma}(3, 1/3)$ ,  $\lambda_{\text{L}} = 1$  and  $\lambda_{\text{G}} = 1/15$ .

#### **Limitations of the model**

- It does NOT incorporate:
  - (A) clustered global contacts
  - (B) casual global contacts
  - (C) global contacts between > 1 pair of individuals in two given households
  - (D) multiple types of individual (e.g. adults and children)

#### **Limitations of the model**

- It does NOT incorporate:
  - (A) clustered global contacts
  - (B) casual global contacts
  - (C) global contacts between > 1 pair of individuals in two given households
  - (D) multiple types of individual (e.g. adults and children)
- Re (A): network of local and global contacts yields a clustered network, which can be used to model possible global contacts in a network/households model calculations may be difficult.
- Re (B): superimpose homogeneously mixing casual contacts to yield a multi-level mixing model (cf. Ball and Neal (2002, 2008), Kiss *et al.* (2006).

## **Example of (very small) population**



#### **Households on a network**

- Population of *m* households located on the vertices of a network constructed by the configuration model.
- Consider an SIR epidemic in which a typical infective infects independently any given local neighbour (i.e. individual in his/her own household) at rate  $\lambda_{\rm L}$  and any given global neighbour (i.e. individual in a neighbouring household) at rate  $\lambda_{\rm G}$ .
- Analysis is complicated because even in the early stages of an epidemic previously infected households can be reinfected.
- May be able to make progress by considering an embedded branching process of fully-infected households.

## **Multitype NHM: population structure**

- J classes of individuals, labelled  $1, 2, \ldots, J$ .
- Category  $\mathbf{n} = (n_1, n_2, \dots, n_J)$  household contains  $n_j$  individuals of class j ( $j = 1, 2, \dots, J$ ).
- For  $n ∈ N = \{n : |n| ≥ 1\}$ , population contains  $m_n$  households having category n.
  - total number of households  $m = \sum_{n \in \mathcal{N}} m_n$
  - total number of class-*i* individuals  $N_i = \sum_{n \in \mathcal{N}} n_i m_n$ .
  - total number of individuals  $N = \sum_{n \in \mathcal{N}} |n| m_n < \infty$ .
- Model analysis valid as  $m \to \infty$  with  $m_n/m \to \rho_n$  $(n \in \mathcal{N})$ , where  $\sum_{n \in \mathcal{N}} \rho_n = 1$  and  $\sum_{n \in \mathcal{N}} \rho_n |n| < \infty$ (plus other mild conditions).

(Ball and Sirl (2012))

## **Multitype NHM: population structure**

- The random network is appreciably more complicated (cf. totally assortative global mixing).
- Now  $D^{(i)} = (D_1^{(i)}, D_2^{(i)}, \dots, D_J^{(i)})$  is a vector giving the (random) number of global neighbours of each type of a typical type-*i* individual.
- ▲ Assign to each individual of type *i* a number of '*i* → *j* half-edges' (*j* = 1, 2, ..., *J*) according to a random sample from  $D^{(i)}$ .
- For each i = 1, 2, ..., J, pair up the  $i \rightarrow i$  half-edges as before.
- For each *i* < *j* = 1, 2, ..., *J*, pair up each *i* → *j* half-edge with a *j* → *i* half-edge chosen uniformly at random without replacement.

(Ball and Sirl (2012))

## **Multitype NHM: population structure**

- Must have (roughly) the same number of  $i \rightarrow j$  and  $j \rightarrow i$  half-edges.
- Easily accommodated if degrees are prescribed in advance.
- If degrees are random then we must have

$$\nu_i \mu_j^{(i)} = \nu_j \mu_i^{(j)}$$

for all  $i \neq j$ , where  $\nu_i = \sum_{n \in \mathcal{N}} n_i \rho_n / \sum_{n' \in \mathcal{N}} |n'| \rho_{n'}$  is the proportion of individuals of type *i* in the population and  $\mu_j^{(i)} = \sum_{d \in \mathbb{Z}_+^J} d_j P(\mathbf{D}^{(i)} = d)$  is the mean number of type-*j* global neighbours of a type-*i* individual.

Proportion of unpaired half-edges tends in probability to 0 as  $m \to \infty$ .

## Multitype NHM: epidemic model

- Infection mechanism SIR.
- Infectious period of class-*i* infective  $\sim I_I^{(i)}$  having an arbitrary but specified distribution .
- Infection rates (individual  $\rightarrow$  individual)
  - local (within-household)  $\lambda_{ij}^L \quad \Lambda^L = [\lambda_{ij}^L]$
  - global (between-household)  $\lambda_{ij}^G \quad \Lambda^G = [\lambda_{ij}^G]$
- Single initial infective chosen uniformly at random either from all individuals of specified type or from entire population. (Easily generalised.)

## **Multitype NHM: size biasing**

- Secondary individuals have the 'normal' degree distribution: the same distribution as *D*<sup>(i)</sup> for a type-*i* individual.
- The degree distribution of a primary individual also depends on the type of individual that infected it. A type-*i* individual globally contacted by a type-*j* individual has degree distribution

$$P(\boldsymbol{D}^{(ij)} = \boldsymbol{d}) = d_j P(\boldsymbol{D}^{(i)} = \boldsymbol{d}) / \mu_j^{(i)}$$

In the forward approximating branching process we must type households by the type of the primary infective and the type of the individual who globally infected it. A similar typing is required in the backward approximating branching process.

## **Numerical Examples**

The basic model for our examples is the following MTNHM: (children and adults are type-1 and type-2 individuals)

# children (type 1) •  $D^{(1)} \equiv (15, 4), D^{(2)} \equiv (6, 10) \text{ and } I^{(1)} \stackrel{\mathcal{D}}{=} I^{(2)} \sim \mathsf{Exp}(2)$ •  $\Lambda^{L} = \begin{bmatrix} 2 & 1 \\ \frac{1}{2} & \frac{1}{5} \end{bmatrix}$  and  $\Lambda^{G} = \begin{bmatrix} \frac{1}{5} & \frac{1}{10} \\ \frac{1}{10} & \frac{1}{20} \end{bmatrix}$ This gives  $R_* \approx 2.25$ ,  $p_{\text{mai}} \approx 0.48$  ( $p_{\text{mai}} \approx (0.57, 0.34)$ ),  $z \approx 0.75$  ( $z \approx (0.82, 0.64)$ ).

#### **Numerical results**



Histograms of relative final sizes for 10,000 simulations of the MTNHM on networks of 20 and 200 households.

#### **Numerical results**



Simulation-based estimates of major outbreak probability against number of households, together with asymptotic value, for the MTNHM. Each estimate is based on 10,000 simulations and the plot shows the sample proportions  $\pm$  2SE.

#### **Numerical results**



Effects of vaccination in our MTNHM with non-random vaccine action given by  $a_1 = b_1 = \frac{1}{2}$ with different vaccine allocation regimes on the expected relative final size of a major outbreak. (Note that  $a_2 = b_2 = 1$ , so vaccine has no effect on adults.)

## **Concluding comments**

- General framework for analysing SIR epidemics in a population incorporating household and network structure
  - asymptotic approximations good for moderately-sized populations.
- Household structure, degree distribution and clustering each have a significant impact on disease dynamics and performance of vaccination strategies.

## **Concluding comments**

- Extensions/further work
  - allow  $\Lambda^L, \Lambda^G$  and degree distributions  $D^{(i)}$  to depend on household category n,
  - acquaintance and optimal vaccination strategies,
  - choice of multivariate degree distributions  $D^{(i)}$  that are realistic, include dependencies and amenable to calculation,
  - relax independent degree assumption, so there are fewer leftover half-edges, without changing asymptotic results.
- All (most?) tractable models have an element of homogeneous mixing at some level.

#### **References**

- Ball, F. G. and Neal, P. J. (2002) A general model for stochastic SIR epidemics with two levels of mixing. *Math. Biosci.* **180**, 73–102.
- Ball, F. G. and Neal, P. J. (2008) Network epidemic models with two levels of mixing. *Math. Biosci.* **212**, 69–87.
- Ball, F. G. and O'Neill, P. D. (1999) The distribution of general final state random variables for stochastic epidemic models. *J. Appl. Prob.* **36**, 473–491.
- Ball, F. G. and Sirl, D. J. (2012) An SIR epidemic model on a population with random network and household structure and several types of individuals. *Adv. Appl. Prob.* **44**, to appear.
- Ball, F. G., Sirl, D. J. and Trapman, P. (2009) An SIR epidemic model on a random network with household structure. *Adv. Appl. Prob.* **41**, 765–796.
- Ball, F. G., Sirl, D. J. and Trapman, P. (2010) Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.* **224**, 53–73.
- Kiss, I., Green, D. and Kao, R. (2006), The effect of contact heterogeniety and multiple routes of transmission on final epidemic size. *Math. Biosci.* **203**, 124–136.

Newman, M. (2002) Spread of epidemic disease on networks. Phys. Rev. E 66 016128.