

# HOW

DO **VIRUS** SPREAD  
AMONG COMPUTERS  
AND HUMANS

**2017**

MATHEMATICAL MODELLING

PASZTOR Barna - ASSOR Samuel - POHL Peter - YUEN Natalie

**UCL**

# INTRODUCTION

## **Example:** The Ransomware Attack in 2017

Motivation: Model the diffusion of a virus attack in a network of computers

Nodes: individuals

Edges: possible contacts between individuals

Topology of the WWW:

- Hubs and clusters
- Hubs are connected while clusters are also interconnected



## PROPERTIES

The **Power Law Distribution** (*leptokurtic* distrib.)

- The probability that a randomly chosen node has degree  $d$  is  $P(d) = 2m^2 d^{-3}$   
(Few nodes with high degree (hubs) and plenty of nodes with low degree)

The **Mean** number of a node's degree is given by:

$$\langle d \rangle = \sum_{j=m}^{\infty} jP(j) \approx \int_m^{\infty} 2m^2 j^{-2} dj = 2m^2 \int_m^{\infty} j^{-2} dj = 2m^2 \left[ \frac{-1}{j} \right]_m^{\infty} = 2m$$

The **Variance** of a node's degree blows up to infinity since the second moment  $\langle d^2 \rangle$  diverges also to infinity from the formula above

# GENERATING ALGORITHM

- The **Barabasi-Albert** Growth Model (1999):

Nodes are born over time and indexed by their date of birth

The system starts with a group of  $N_0$  nodes all connected to one another

Each node upon birth forms  $m$  (undirected) edges with pre-existing nodes

**Preferential attachment** (rich get richer): Instead of selecting  $m$  nodes uniformly at random, it attaches to nodes with probabilities proportional to their degrees

The probability that an existing node  $i$  receives a new link to the new-born node at time  $t$  is  $m \frac{d_i(t)}{\sum_{j=1}^{N_0+t} d_j(t)}$ . The number of edges at time 0 is  $L_0 = \frac{N_0(N_0 - 1)}{2}$  so  $L_t = L_0 + mt$

When  $t$  is very large, both  $L_t \approx mt$  and  $N_t \approx t$  so  $m \frac{d_i(t)}{\sum_{j=1}^{N_0+t} d_j(t)} = m \frac{d_i(t)}{2L_t} = \frac{d_i(t)}{2t}$

# GENERATING ALGORITHM

We can write the evolution of degrees in continuous time as

$$\frac{d}{dt}(d_i(t)) = \frac{d_i(t)}{2t}$$

with initial condition  $d_i(i) = m$ . Using separation of variables, we get that  $d_i(t) = m \left( \frac{t}{i} \right)^{1/2}$

The CDF is given by  $P(d_i(t) < d) = P\left(m \left( \frac{t}{i} \right)^{1/2} < d\right) = P\left(i > \frac{m^2 t}{d^2}\right) = 1 - P\left(i < \frac{m^2 t}{d^2}\right)$

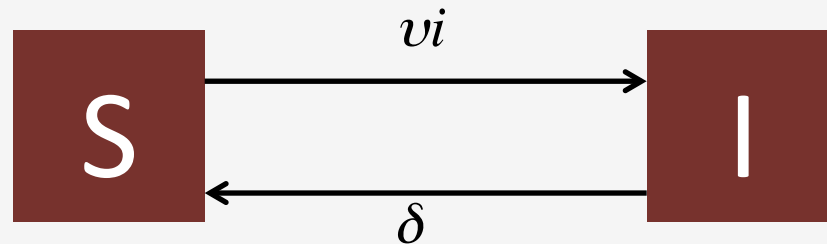
Entry time  $i$  is a random variable that is uniform in time. At time  $t$ , there are roughly  $t$  nodes. Thus  $i$  has a uniform distribution  $U(0, t)$

$$\text{So } P(d_i(t) < d) = 1 - P\left(i < \frac{m^2 t}{d^2}\right) = 1 - \frac{m^2 t}{d^2} \frac{1}{t} = 1 - \frac{m^2}{d^2}$$

The density function of the nodes' degree is hence given by  $P(d) = 2m^2 d^{-3}$

# SIS MODEL

**Compartmental Model** : 2 possible discrete states (susceptible and infectious)



The **transmission** parameter  $\nu$ : probability of a contact between a susceptible and an infected person resulting in transmission of the infectious disease

The **recovery** rate  $\delta$ : percent of the population that recovers from a disease each period.

$$\frac{d}{dt}s = -\nu kis + \delta i$$

$$\frac{d}{dt}i = \nu kis - \delta i$$

# SIS MODEL

The **Reproductive number**: the average number of secondary infections produced by an infected individual in a wholly susceptible population.

When the network is **homogeneous** :  $\rho_0 = \frac{\nu \langle k \rangle}{\delta}$

When the network is **heterogeneous**:  $R_0 = \rho_0 (1 + CV^2)$  where  $CV$  is the coefficient of variation of the nodes' degree distribution

Infection spreads when  $R_0 > 1$  and dies out when  $R_0 < 1$

**Consequence: NO THRESHOLD FOR A SCALE FREE NETWORK**

# SIS MODEL

The **Pastor-Satorras** and **Vespignani** analytic approach (2000):

We define the effective spreading rate as  $\lambda = \frac{\nu}{\delta}$  and WLOG we assume that  $\delta = 1$

If  $\rho_k(t)$  is the relative density of infected nodes with degree  $k$ , then:

$$\frac{d}{dt} \rho_k(t) = -\rho_k(t) + \lambda k [1 - \rho_k(t)] \Theta(\lambda)$$

where  $\Theta(\lambda)$  is the probability that any given link points to an infected node

At the steady state (i.e. :  $\frac{d}{dt} \rho_k(t) = 0$ ) we get that  $\rho_k(t) = \frac{k\lambda\Theta(\lambda)}{1 + k\lambda\Theta(\lambda)}$

The probability that a meeting of node  $l$  is with a degree  $d$  node is  $\frac{P(d)d}{\langle d \rangle}$

$$\text{so } \Theta(\lambda) = \frac{\sum k P(k) \rho_k}{\sum s P(s)}$$



# SIS MODEL

With few calculus steps, the average density of infected nodes at the steady state is given by:

$$\rho \cong 2e^{-1/m\lambda}$$

This proves again the absence of threshold or critical point in the model, i.e.  $\lambda_c = 0$

# SIR MODEL

**Compartmental Model** : 3 possible discrete states (susceptible, infectious and recovered)



The **transmission** parameter  $\nu$  : probability of a contact between a susceptible and an infected person resulting in transmission of the infectious disease

The **recovery** rate  $\delta$  : percent of the population that recovers from a disease each period. If the duration of the infection is denoted  $D$ , then  $\nu = 1/D$ , since an individual experiences one recovery in  $D$  units of time.

$$\frac{d}{dt}s = -\nu kis$$

$$\frac{d}{dt}i = \nu kis - \delta i$$

$$\frac{d}{dt}r = \delta i$$

# CONCLUSION

Most computer networks have a very specific structure: we can model using a scale-free network.

With the SIS model, we found that there is no threshold: the disease will spread for any value of the spreading rate

With the introduction of antivirus software, we change our model and use the SIR where the epidemic disease ultimately dies out

Our analysis for computer networks were directly derived from analysis on infected human populations