Social Network Analysis

Epidemics

Course Outline

- Graph Theory and Social Networks
- Visualizing Social Networks
- Game Theory
- Information Networks and the World Wide Web
- **Network Dynamics**
- Applications of SNA in various domains

Influence in a Social Network

- Information Cascade
- Epidemics
- Application – Viral Marketing
- Application – Churn Prediction

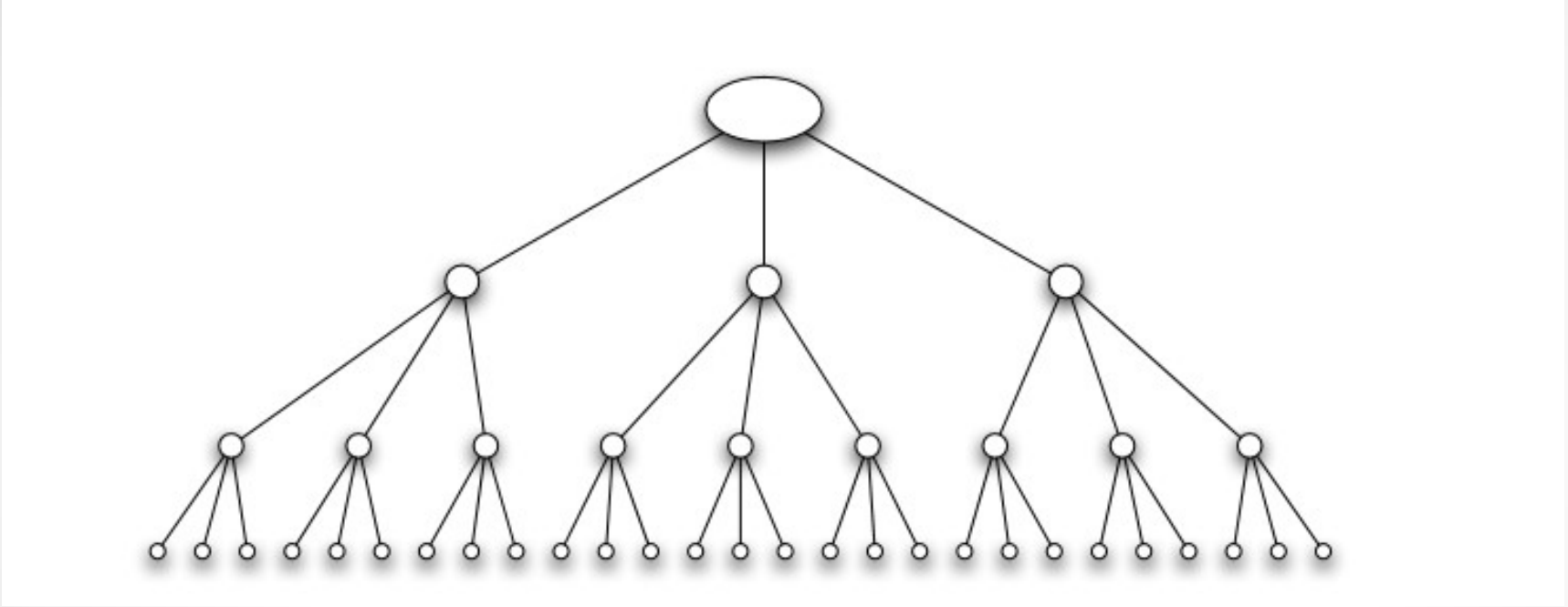
Epidemics

- Contagious diseases caused by biological pathogens
- Epidemics can be of different kinds
  - Sudden explosion
  - Persist for a long time at low levels
  - Sudden flare-ups
  - Cyclic patterns
- Can be devastating
  - Bubonic plague that killed 20% of Europe in a 7 year period in 1300s
  - **COVID 19**

Networks Influence on Epidemics

- Epidemics determined by
  - Pathogens: Contagiousness, Length of infection
  - **Social Network**
- Pathogens determine Social Network
  - Air-borne: Lot of links in the network
  - Close contact: Sparse network
- Similar for computer viruses

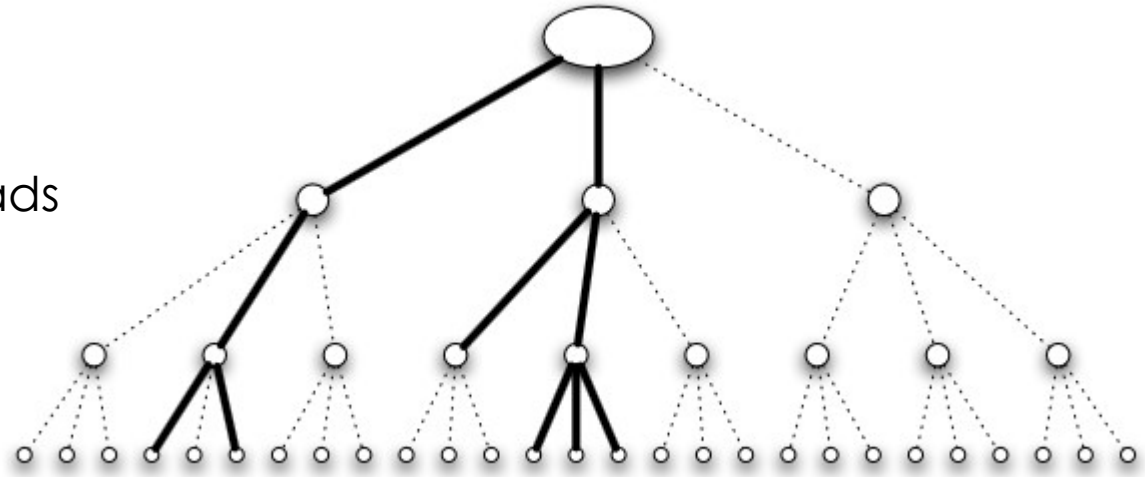
Branching Process



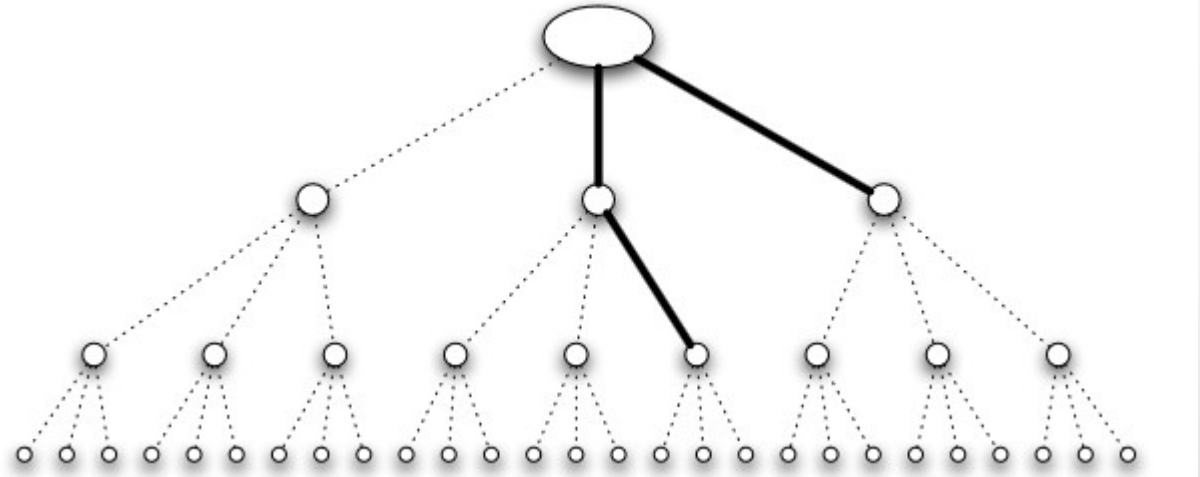
Suppose that a person carrying a new disease enters a population, and transmits it to each person he meets independently with a probability of q (or p) (**Contagion probability**). Further, suppose that he meets d (or k) people while he is contagious.

Effect of Contagion Probability

High q: Disease spreads



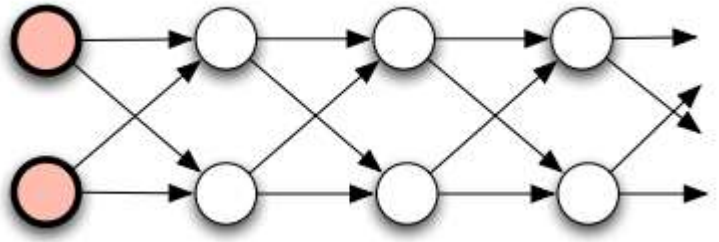
Low q: Disease dies out



Reproductive Number

- $R_0 = q.d$
  - If $R_0 < 1$ then with probability 1 the disease dies out after a finite number of waves
  - If $R_0 > 1$ then with probability, then with probability greater than 0 the disease persists by infecting at least one person in each wave
  - If R_0 is near 1 a small change in k or p can have a significant change in the effect of the disease
- Reducing R_0
  - q: Encouraging better sanitary practices reduces germs spreading
  - d: Quarantining people
- **Only R_0 matters**
  - HIV has an R_0 between 2 and 5
  - Measles has an R_0 between 12 and 18
  - Ebola has an R_0 between 1.5 and 2
  - **Covid 19?**

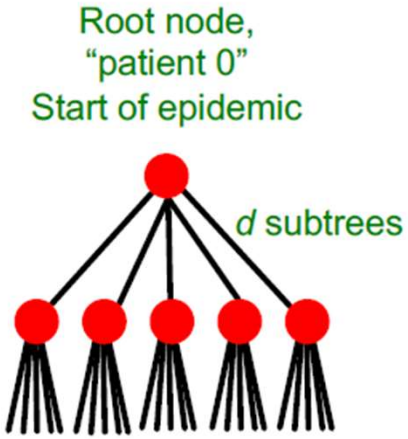
Reproductive Number – Relevant only for Tress



- $R_0 = 2/3 * 2 = 4/3$
- However the disease will die out after reaching only a finite number of steps.
- In each layer, there are four edges leading to the next layer, and each will independently fail to transmit the disease with probability 1/3. Therefore, with probability $(1/3)^4 = 1/81$, all four edges will fail to transmit the disease — and at this point, these four edges become a "roadblock" guaranteeing the disease can never reach the portion of the network beyond them.
- Thus, as the disease moves along layer-by-layer, there is a probability of at least 1/81 that each layer will be its last.
- Therefore, with probability 1, it must come to an end after a finite number of layers.

Probabilistic Spreading Models

- **Epidemic Model based on Random Trees**
  - **(a variant of branching processes)**
  - A patient meets ***d*** new people
  - With probability ***q > 0*** she infects each of them
- **Q: For which values of *d* and *q* does the epidemic run forever?**

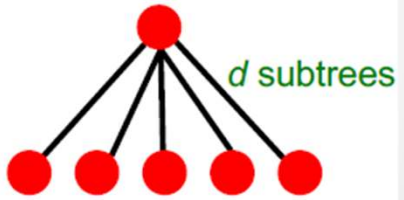


- **Run forever:** $\lim_{h\to\infty} \boldsymbol{P}\left[\begin{matrix}\boldsymbol{a\ node\ at\ depth\ h}\\ \boldsymbol{is\ infected}\end{matrix}\right] > 0$
- **Die out:** $\lim_{h\to\infty} \boldsymbol{P}\left[\begin{matrix}\boldsymbol{a\ node\ at\ depth\ h}\\ \boldsymbol{is\ infected}\end{matrix}\right] = 0$

Probabilistic Spreading Models

- p_h = prob. a node at depth h is infected
- **We need:** $\lim_{h\to\infty} p_h = ?$ (based on q and d)
  - We are reasoning about a behavior at the root of the tree. Once we get a level out, we are left with identical problem of depth $h-1$.
- **Need recurrence for p_h**

$$p_h = 1 - \underbrace{(1 - q \cdot p_{h-1})^d}_{\text{No infected node at depth } h \text{ from the root}}$$



- $\lim_{h\to\infty} p_h$ = result of iterating

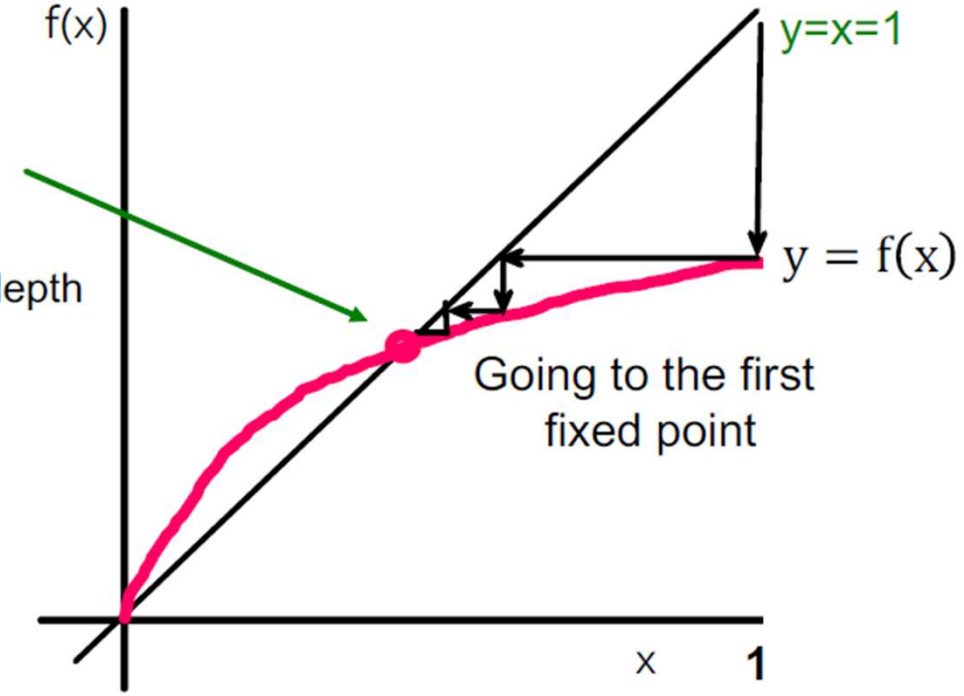
$$f(x) = 1 - (1 - q \cdot x)^d$$

  - Starting at the root: $x = 1$ (since $p_1 = 1$)

We iterate:
$x_1 = f(1)$
$x_2 = f(x_1)$
$x_3 = f(x_2)$

Fixed Point: $f(x) = 1 - (1 - qx)^d$

Fixed point:
$f(x) = x$
This means that prob. there is an infected node at depth h *is constant (>0)*



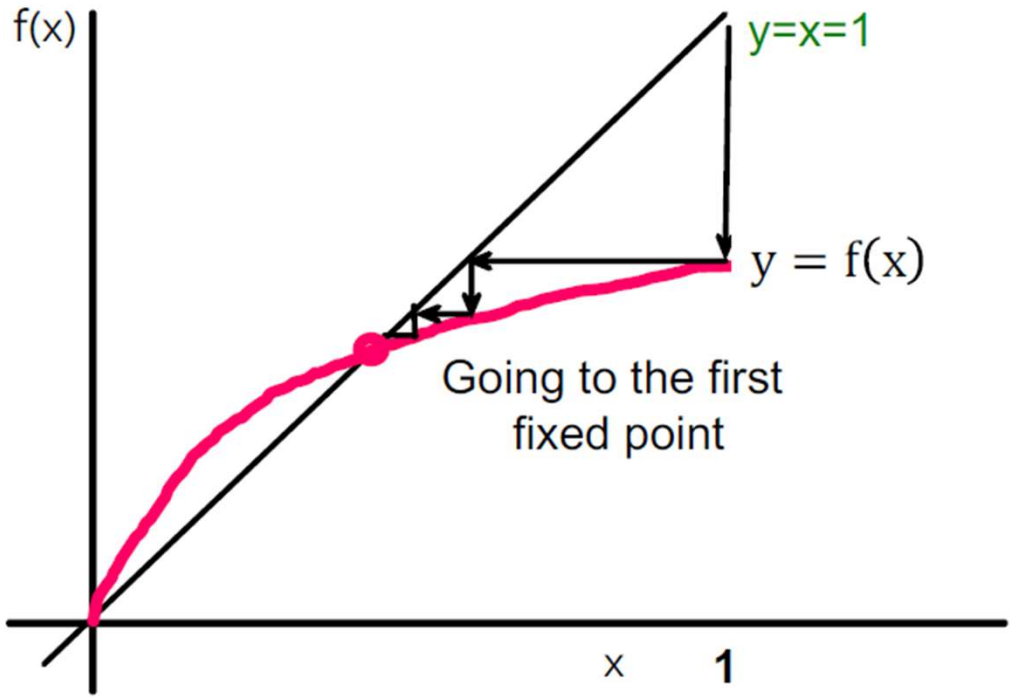
x … prob. a node at level *h-1* is infected
We start at x=1 because p_1=1.
f(x) … prob. a node at level *h* is infected
q … infection prob.
d … degree

We iterate:
$x_1 = f(1)$
$x_2 = f(x_1)$
$x_3 = f(x_2)$

If we want to epidemic to die out, then iterating $f(x)$ must go to zero. So, $f(x)$ must be **below** $y = x$.

- **What's the shape of $f(x)$?**

Fixed Point: $f(x) = 1 - (1 - qx)^d$



x … prob. a node at level *h-1* is infected
We start at x=1 because p_1=1.
f(x) … prob. a node at level *h* is infected
q … infection prob.
d … degree

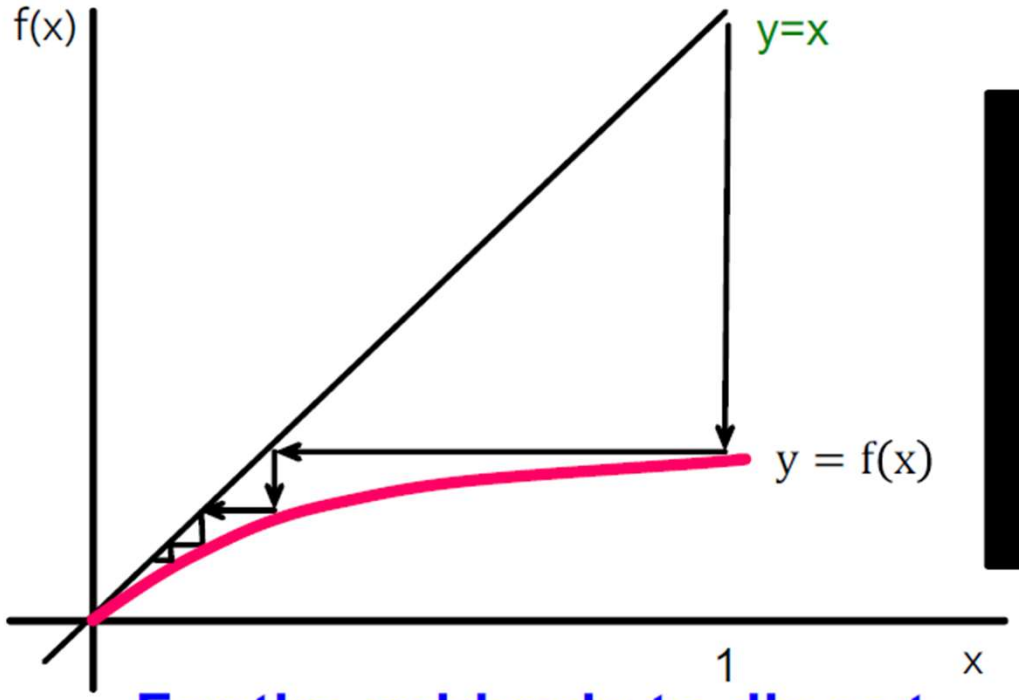
What do we know about the shape of $f(x)$?

- $f(0) = 0$
- $f(1) = 1 - (1 - q)^d < 1$
- $f'(x) = q \cdot d(1 - qx)^{d-1}$
- $f'(0) = q \cdot d$

f'(x) is monotone: If g'(y)>0 for all y then g(y) is monotone.
In our case, 0≤x,q≤1, d>1 so f'(x)>0, so f(x) is monotone.
f'(x) non-increasing: since term $(1-qx)^{d-1}$ in *f'(x)* is decreasing as x decreases.

$f'(x)$ is monotone non-increasing on [0,1]!

Fixed Point: When is this zero?



Reproductive number $R_0 = q \cdot d$:
There is an epidemic if $R_0 \geq 1$

For the epidemic to die out we need $f(x)$ to be below $y = x$!

So: $f'(0) = q \cdot d < 1$

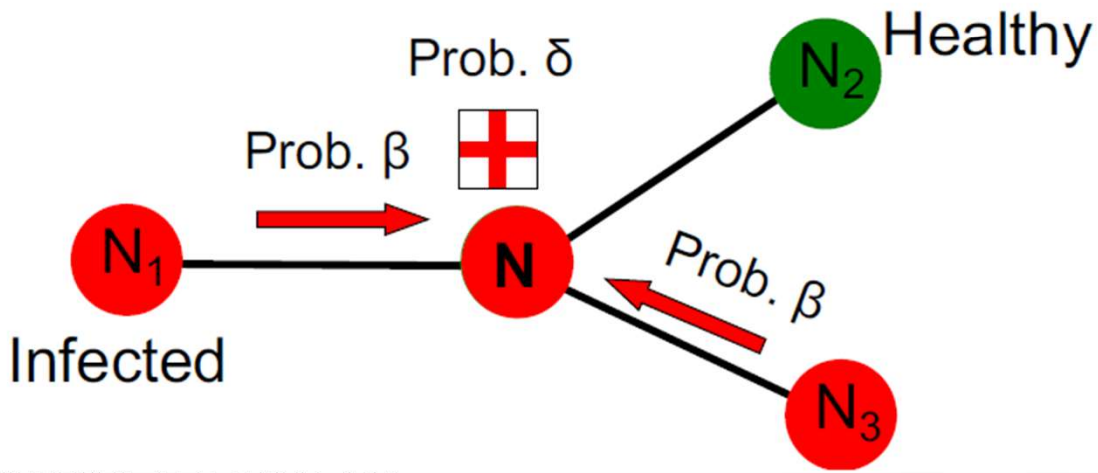
$$\lim_{h \to \infty} p_h = 0 \ \ when \ \ q \cdot d < 1$$

$q \cdot d$ = expected # of people that get infected

Spreading Models of Viruses

Virus Propagation: 2 Parameters:

- **(Virus) Birth rate β:**
  - probability that an infected neighbor attacks
- **(Virus) Death rate δ:**
  - Probability that an infected node heals



SIR Epidemic Model

- An individual node in the branching process model goes through three potential stages during the course of the epidemic:
  - Susceptible: Before the node has caught the disease, it is susceptible to infection from its neighbors.
  - Infectious: Once the node has caught the disease, it is infectious and has some probability of infecting each of its susceptible neighbors.
  - Removed: After a particular node has experienced the full infectious period, this node is removed from consideration, since it no longer poses a threat of future infection
- The progress of the epidemic is controlled by the contact network structure and by two additional quantities:
  - p (the probability of contagion)
  - t_I (the length of the infection).
- Initially, some nodes are in the I state and all others are in the S state.
- Each node v that enters the state remains infectious for a fixed number of steps t_I
- During each of these t_I steps, v has a probability p of passing the disease to each of its susceptible neighbors.
- After t_I steps, node v is no longer infectious or susceptible to further bouts of the disease; we describe it as Removed (R)

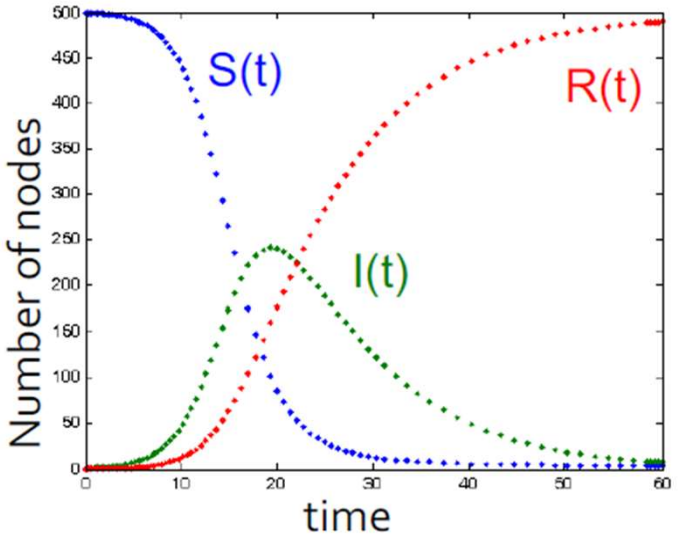
SIR Model

- **SIR model:** Node goes through phases

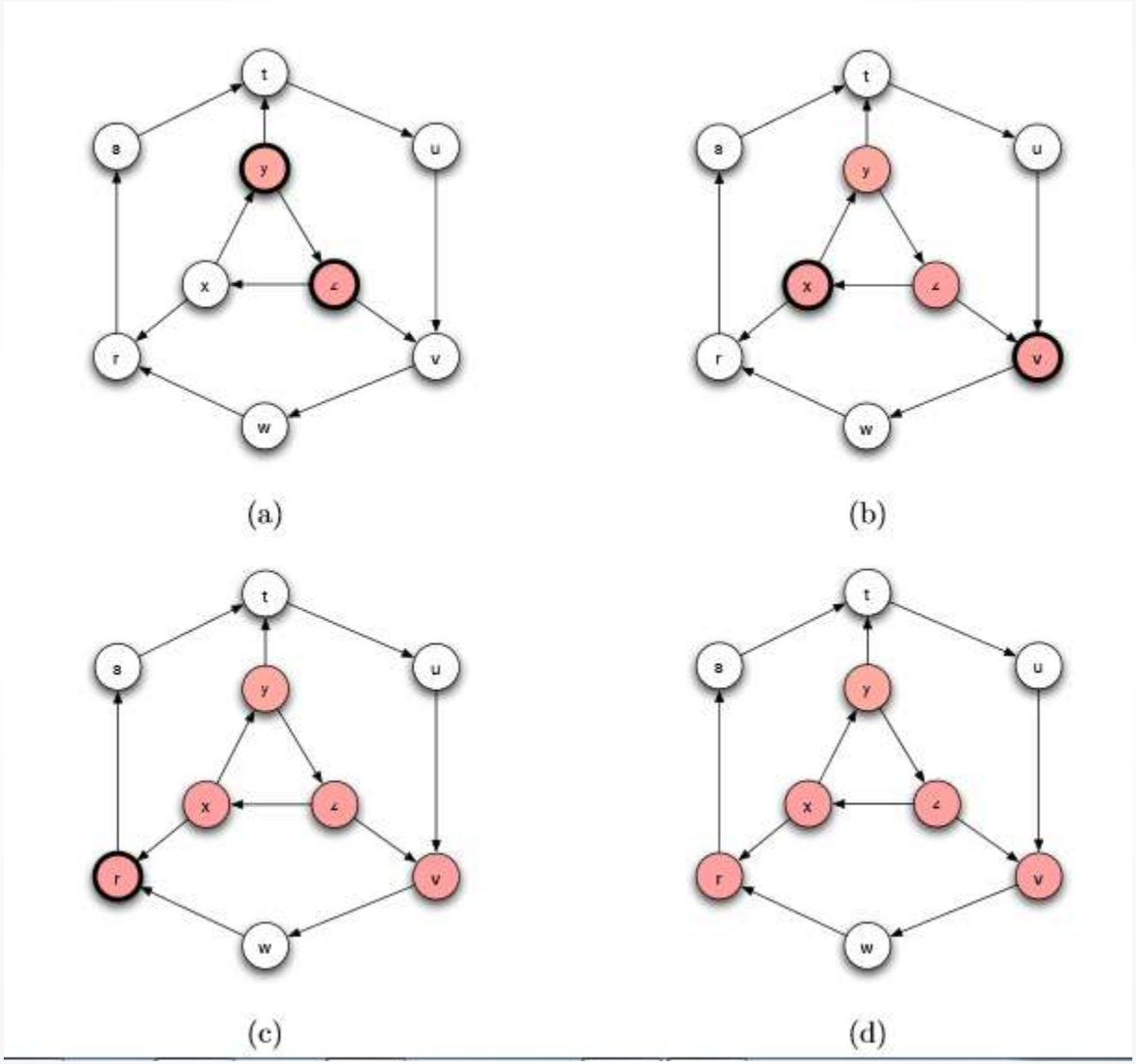
  - Models chickenpox or plague:
    - Once you heal, you can never get infected again
- **Assuming perfect mixing** (The network is a complete graph) **the model dynamics are:**

$$\frac{dS}{dt} = -\beta SI \qquad \frac{dR}{dt} = \delta I$$

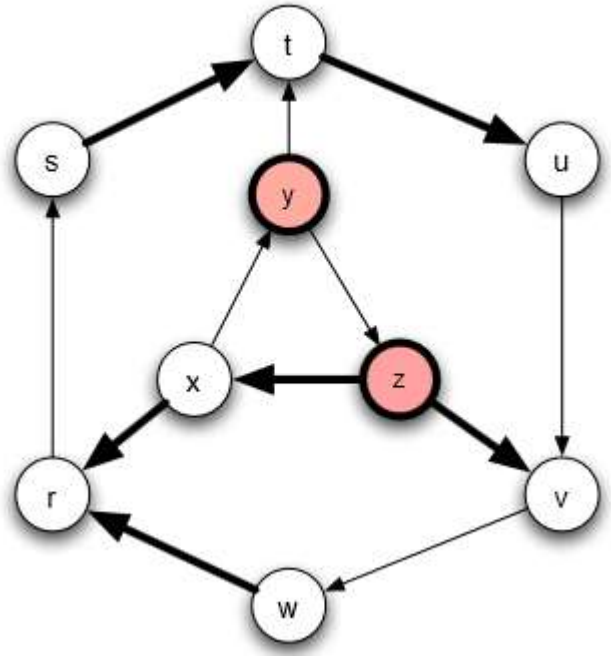
$$\frac{dI}{dt} = \beta SI - \delta I$$



SIR Epidemic Model



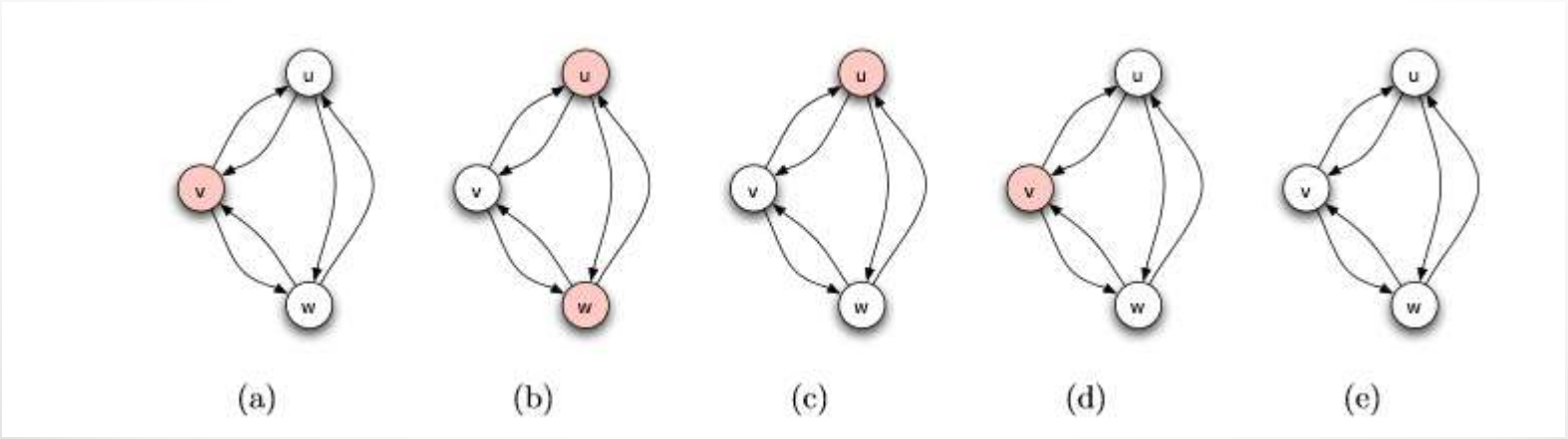
Percolation



- The static view of the model
- We decide in advance which edges will transmit infection
- A node v will become infected during the epidemic if and only if there is a path to v from one of the initially infected nodes that consists entirely of open edges

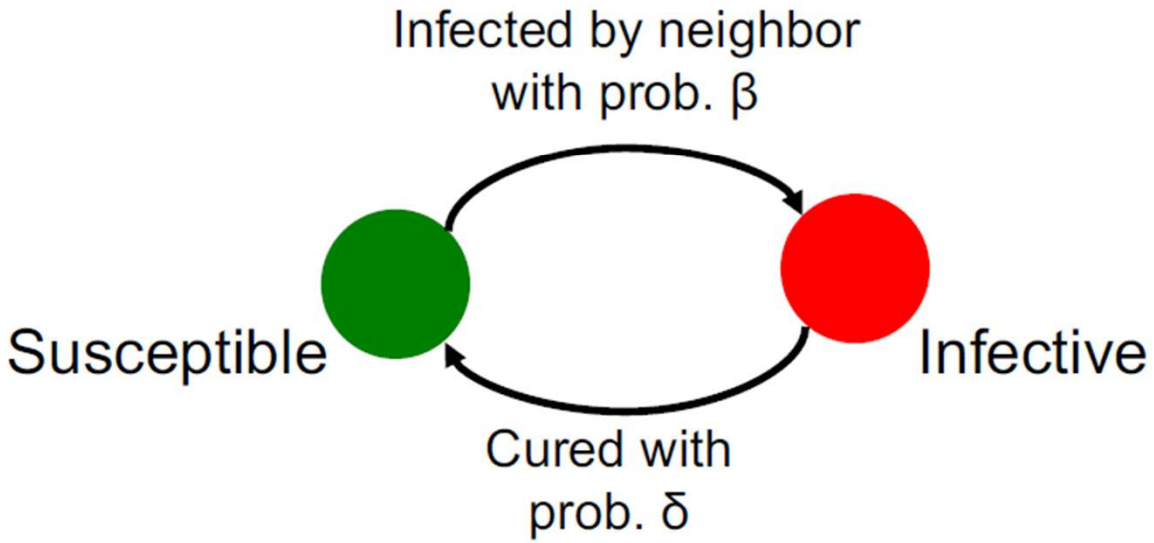
SIS Epidemic Model

- There is no Removed state
  - After a node is done with the Infectious state, it cycles back to the Susceptible state and is ready to catch the disease again
- Initially, some nodes are in the I state and all others are in the S state.
- Each node v that enters the state remains infectious for a fixed number of steps t_I
- During each of these t_I steps, v has a probability p of passing the disease to each of its susceptible neighbors.
- After t_I steps, node v is no longer infectious and it returns to the S state

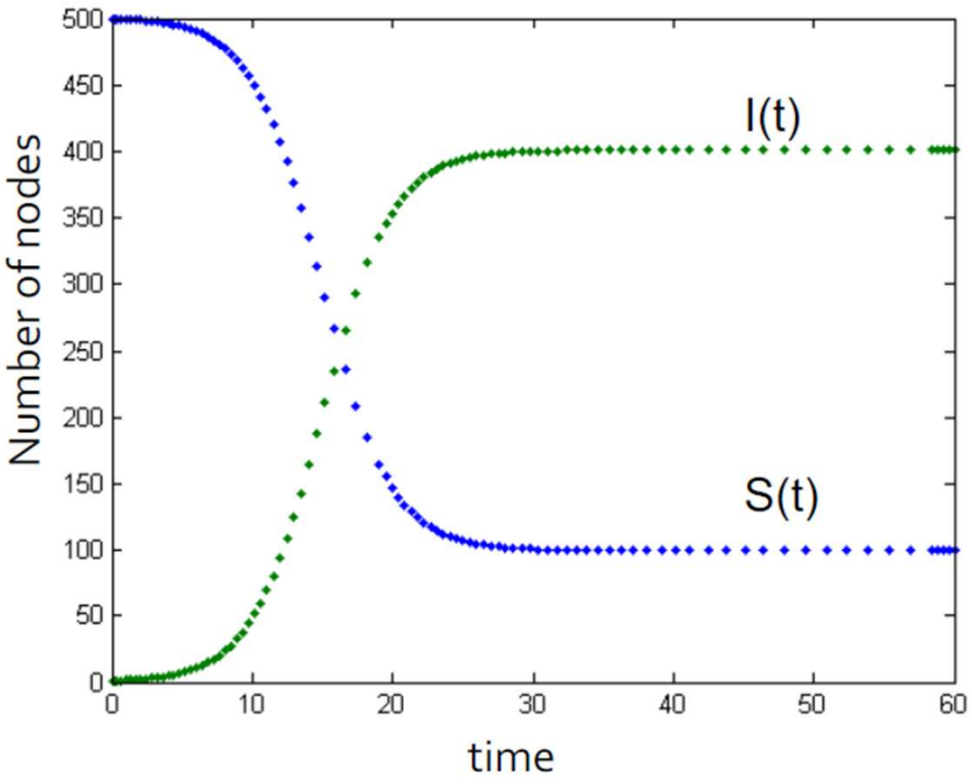


SIS Model

- **Susceptible-Infective-Susceptible (SIS) model**
- Cured nodes immediately become susceptible
- **Virus "strength": $s = \beta / \delta$**
- **Node state transition diagram:**



SIS Model



- **Models flu:**
  - Susceptible node becomes infected
  - The node then heals and become susceptible again
- **Assuming perfect mixing (a complete graph):**

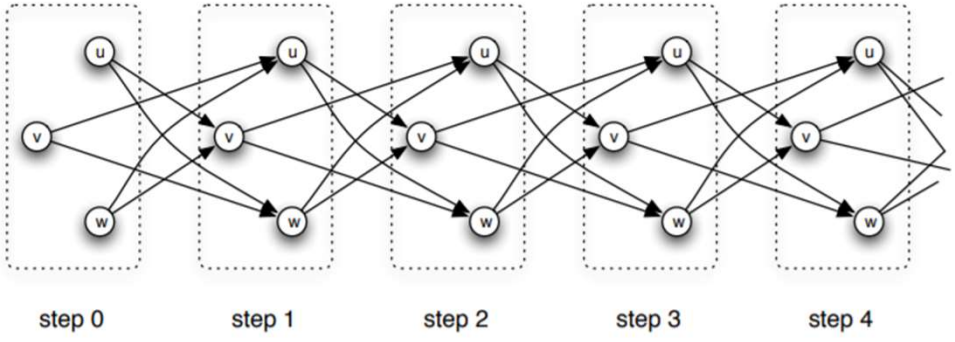
$$\frac{dS}{dt} = -\beta SI + \delta I$$

$$\frac{dI}{dt} = \beta SI - \delta I$$

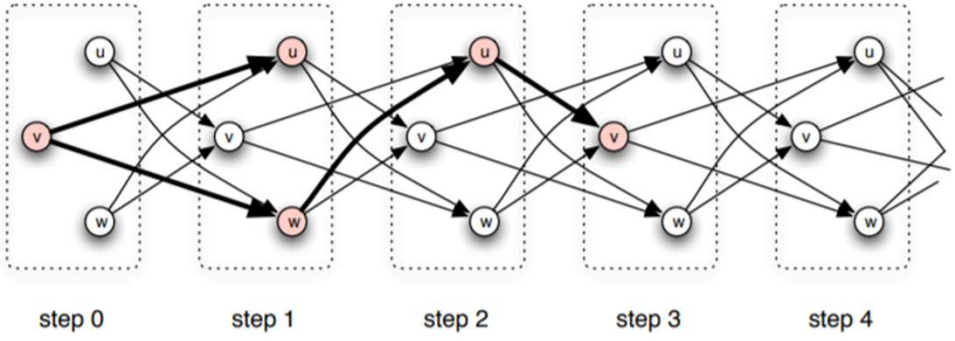
SIR vs SIS

- In a SIR epidemic graph the nodes decrease over time – so the epidemic will ultimately die down after a relatively small number of steps
- A SIS epidemic can run for an extremely long time
- At at a particular critical value of the contagion probability p, SIS epidemic on the network will undergo a rapid shift from one that dies out quickly to one that persists for a very long time
- SIS model can be converted to SIR model by a time-expanded contact network
  - A new copy of each infected node is created

Time-expanded contact network



(a) *To represent the SIS epidemic using the SIR model, we use a "time-expanded" contact network*



(b) *The SIS epidemic can then be represented as an SIR epidemic on this time-expanded network.*

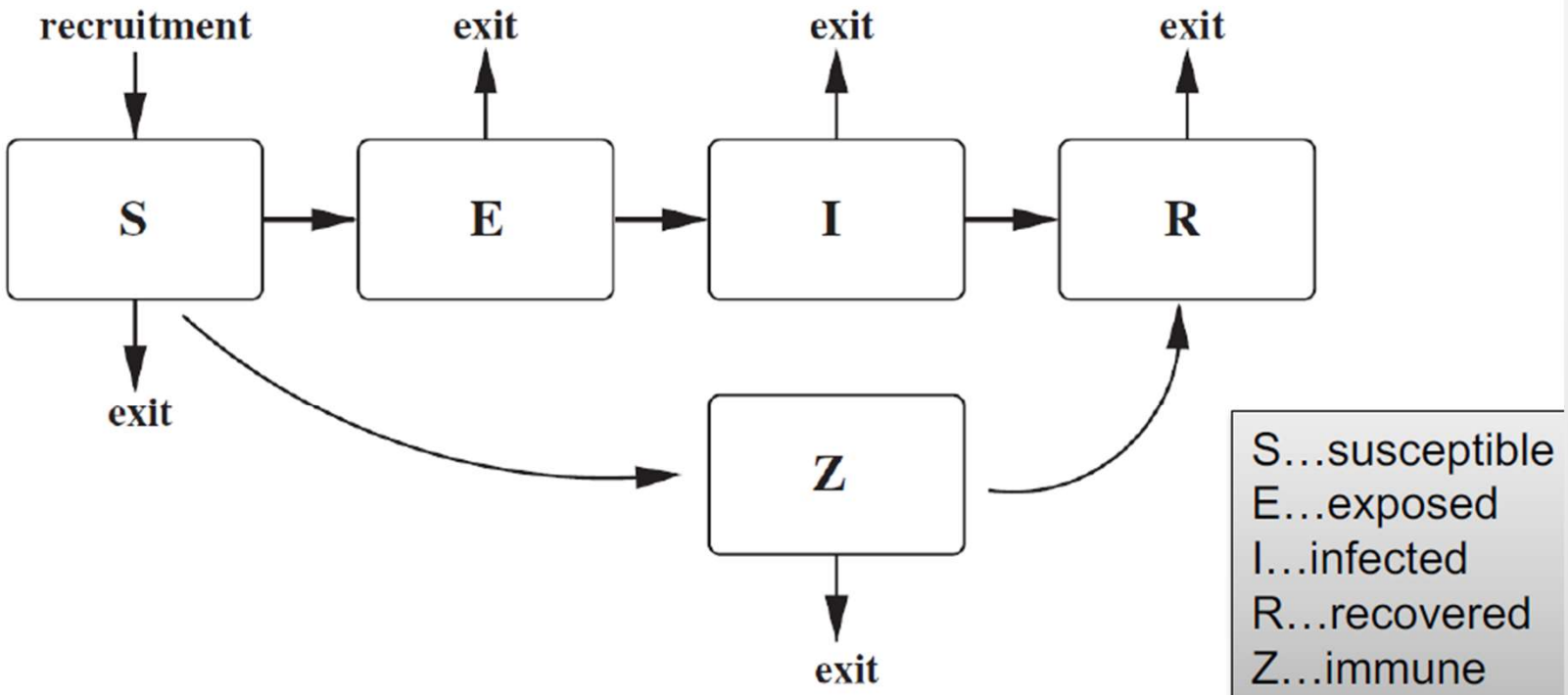
- We think about a node v as in fact being a "different individual" at each time step, then we can represent things so that nodes are never reinfected.
- We create a separate copy of each node for each time step t = 0, 1, 2, 3 and onward.
- Now, for each edge in the original contact network, linking a node v to a node w, we create edges in the time-expanded contact network from the copy of v at time t to the copy of w at time t + 1

SIRS Epidemic Model

- After an infected node recovers, it passes briefly through the R state on its way back to the S state
- Initially, some nodes are in the I state and all others are in the S state.
- Each node v that enters the state remains infectious for a fixed number of steps t_I
- During each of these t_I steps, v has a probability p of passing the disease to each of its susceptible neighbors.
- After t_I steps node v is no longer infectious. It then enters the R state for a fixed number of steps t_R. During this time, it cannot be infected with the disease, nor does it transmit the disease to other nodes.
- After t_R steps in the R state, node v returns to the S state

More Generally: S+E+I+R Models

- **General scheme for epidemic models:**
  - **Each node can go through phases:**
    - Transition probs. are governed by the model parameters



SEIR Model

- Let $S(t)$, $E(t)$, $I(t)$, and $R(t)$ represent fraction of population in each of four groups at time t.

$$S(t) + E(t) + I(t) + R(t) = 1$$

- Dynamics:

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dE}{dt} = \beta SI - \alpha E$$

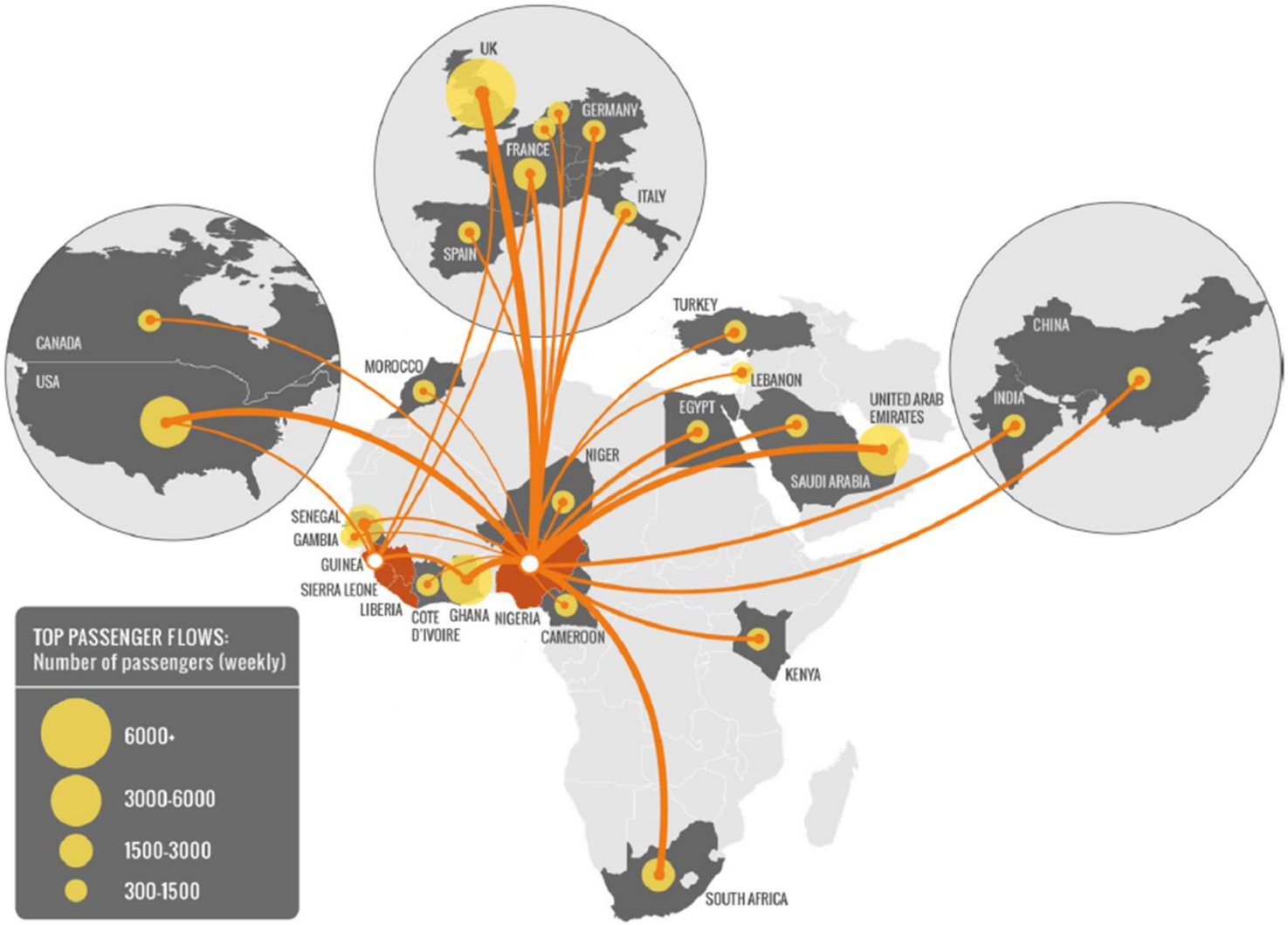
$$\frac{dI}{dt} = \alpha E - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

Adding E and I reduces to SIR model.

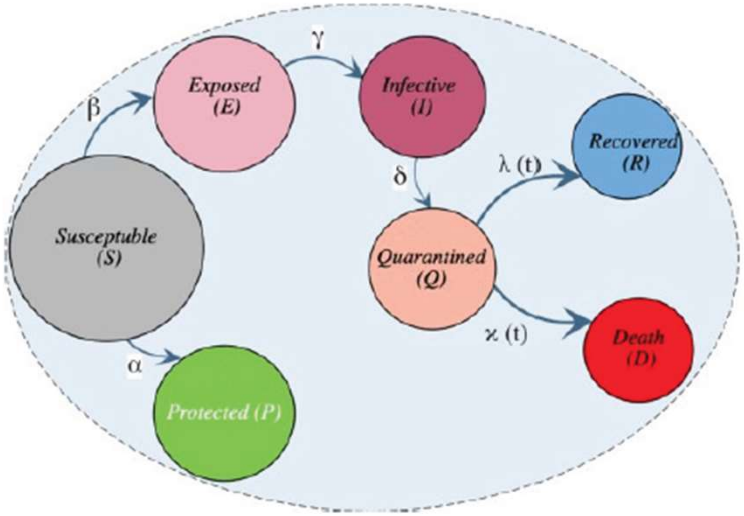
[Gomes et al., 2014]

Modeling Ebola with SEIR



[Gomes et al., Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak, *PLOS Current Outbreaks*, '14]

SEIQRDP model



(1) S_t corresponding to the percentage of population susceptible to Corona
(2) E_t corresponds to the proportion of the population exposed to Corona
(3) I_t represents the proportion of the population that is infected with Corona
(4) Q_t represents the proportion of the population that is quarantined. The current SEIQRDP model's usage of Quarantine is defined as the separation of the healthy and asymptomatic persons to prevent the virus from spreading rapidly.
(5) R_t corresponds to the proportion of the population who have recovered from Corona
(6) D_t is a subset of the population that have died as a result of Corona
(7) P_t represents the proportion of the population that is immune to the virus [3, 16].

$$\frac{dS(t)}{dt} = -\alpha S(t) - \beta \frac{S(t)I(t)}{N}$$

$$\frac{dE(t)}{dt} = -\gamma E(t) + \beta \frac{S(t)I(t)}{N}$$

$$\frac{dI(t)}{dt} = \gamma E(t) - \delta I(t)$$

$$\frac{dQ(t)}{dt} = \delta I(t) - \lambda(t)Q(t) - \kappa(t)Q(t) \quad . \qquad (1)$$

$$\frac{dR(t)}{dt} = \lambda(t)Q(t)$$

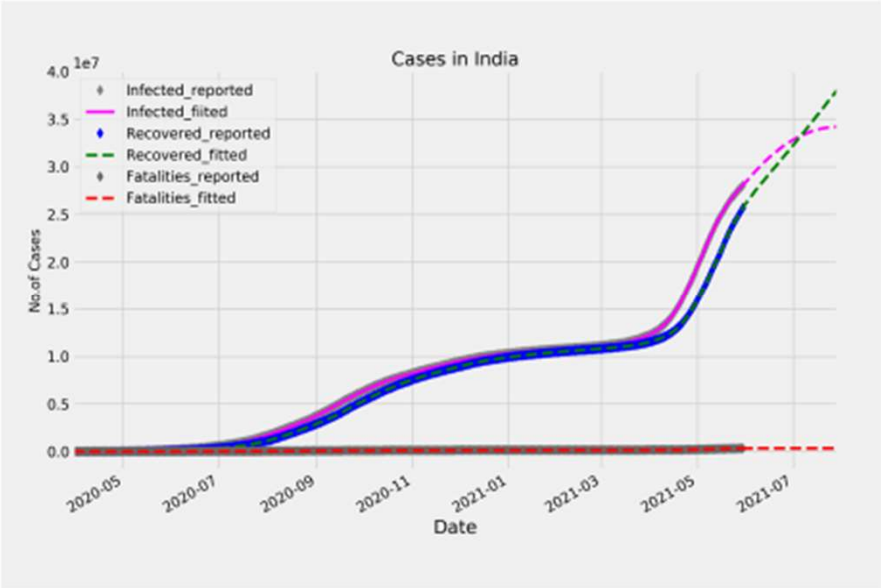
$$\frac{dD(t)}{dt} = \kappa(t)Q(t)$$

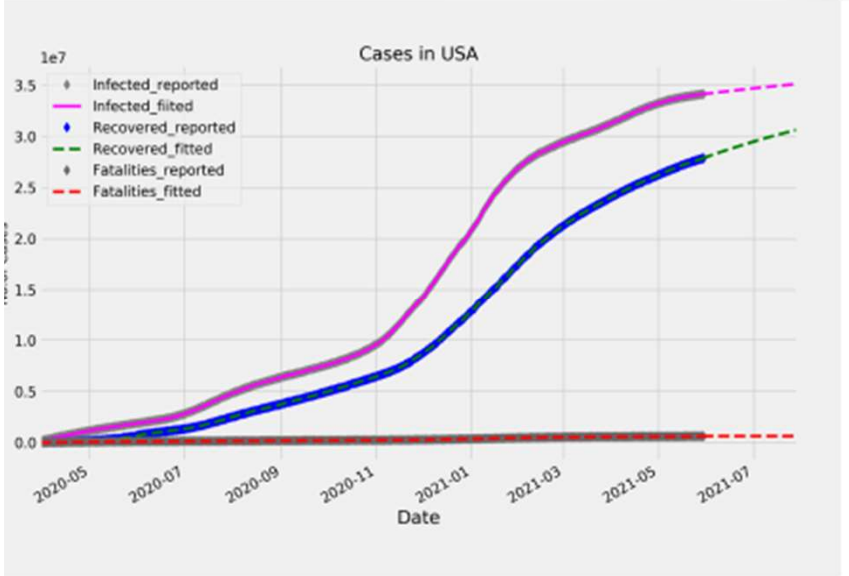
$$\frac{dP(t)}{dt} = \alpha S(t)$$

where β is the infection rate, γ is the inverse of the average latent time, δ is the rate of infected person sent to quarantine, $\lambda(t)$ is cure rate at time t and $\kappa(t)$ is mortality rate at time t.

- S. Mukherjea & A.A. Nanavati; IIT-D Spring 2016

Covid 19 Prediction using SEIQRDP model



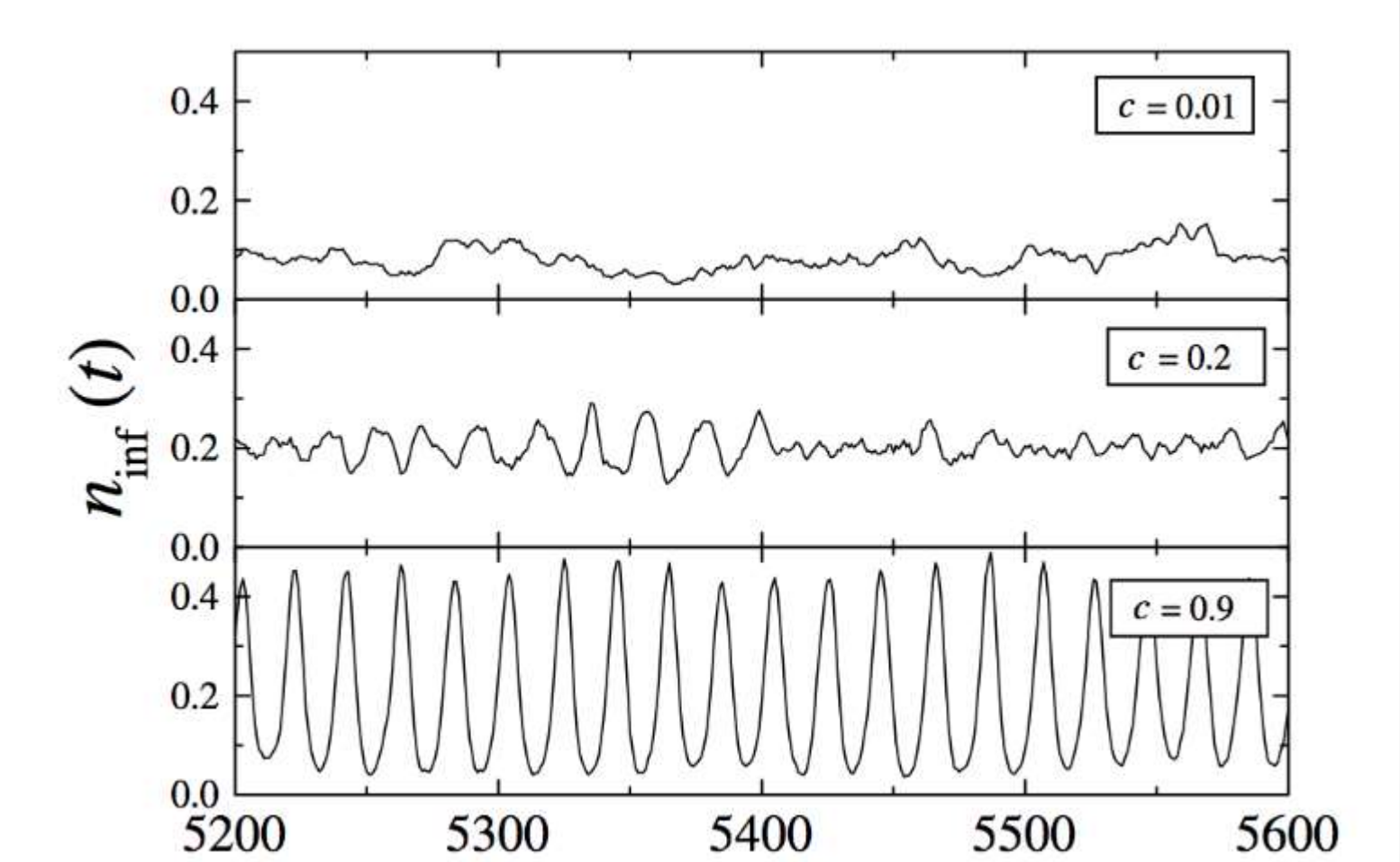


Siva Prasad Reddy – Masters Project, IIT Delhi 2021

Oscillations

- The tendency of epidemics for certain diseases to synchronize across a population, sometimes producing strong oscillations in the number of affected individuals over time.
  - Such effects are well-known for diseases including measles and syphilis.
- Main requirements:
  - Long-range link: Produce coordination in the timing of flare-ups across dispersed parts of the network
  - Temporary immunity: The temporary immunity produces a network-wide deficit in the number and connectivity of susceptible individuals, yielding a large "trough" in the size of the outbreak that directly follows the "peak" from the earlier flare-ups

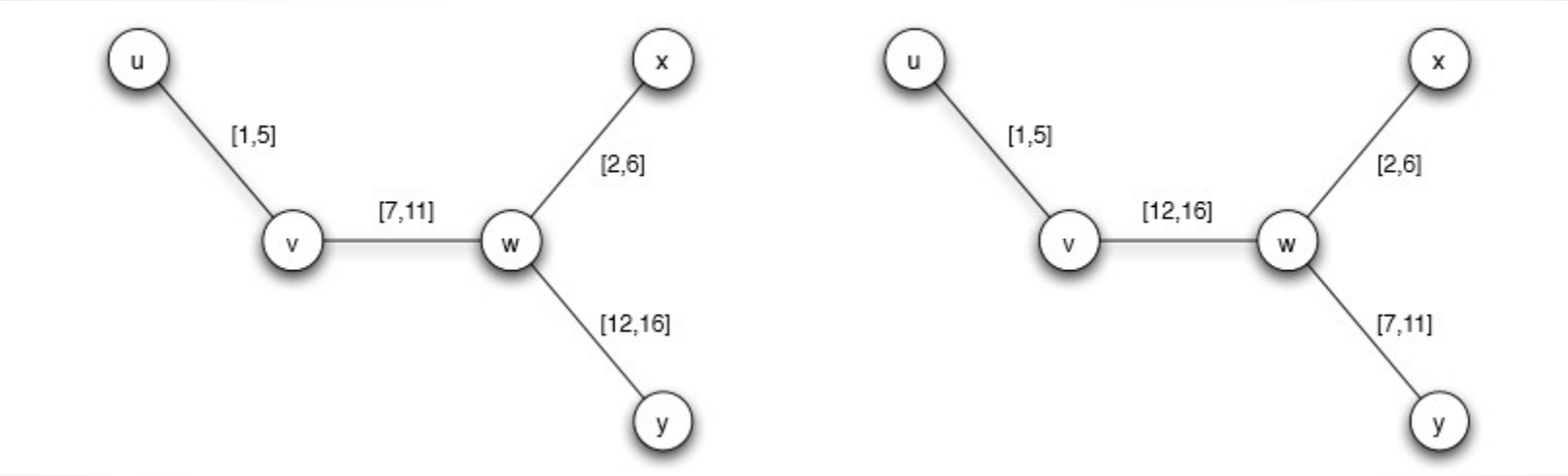
Synchronization



c: Fraction on long range links

- Absence of oscillations for small c (c = 0.01)
- Wide oscillations for large c (c = 0.9)
- A transitional region (c = 0.2) where oscillations intermittently appear and then disappear

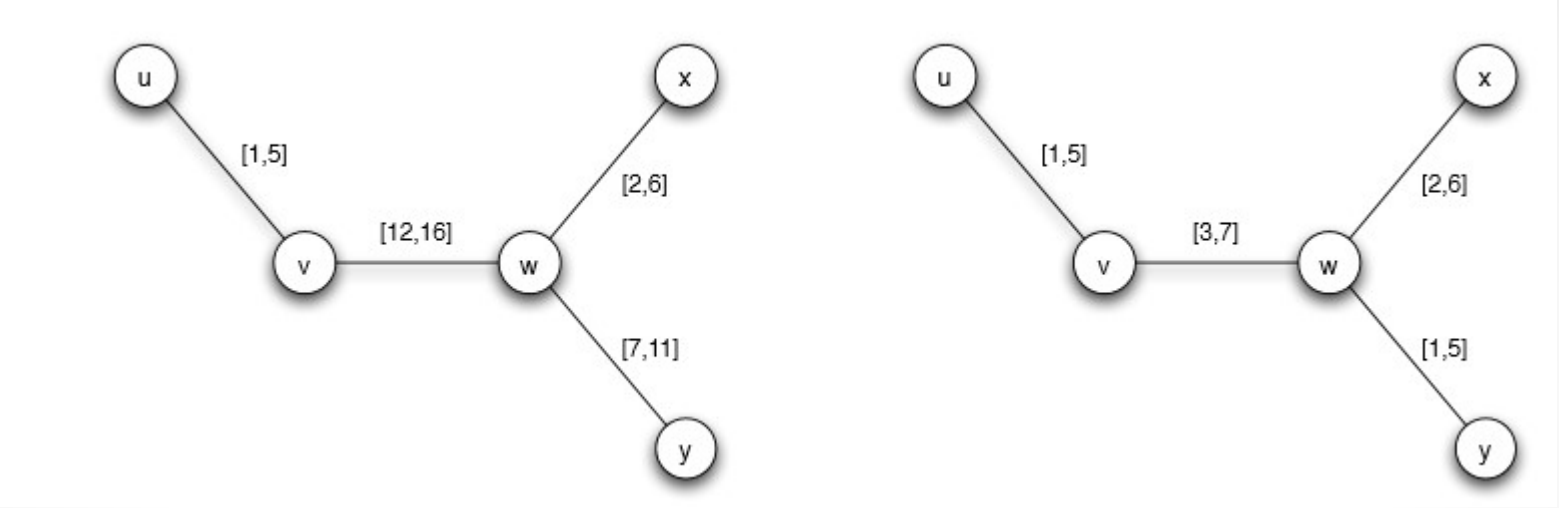
Transient and Concurrent Contacts



The disease can potentially pass all the way from u to y

The disease can't potentially pass from u to y

Transient and Concurrent Contacts

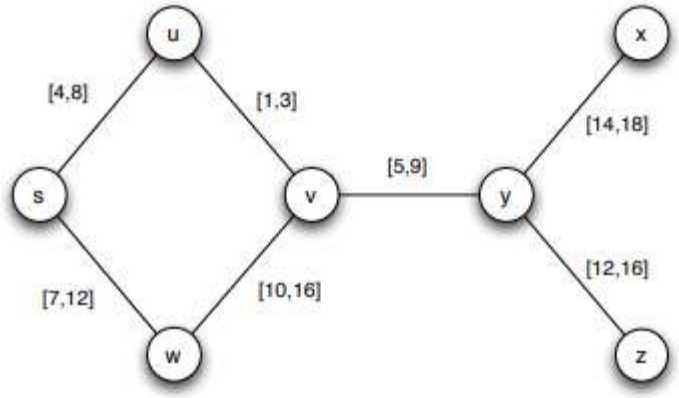


Relations are not concurrent

Relations are concurrent

In larger networks, the effects of concurrency on disease spreading can become particularly pronounced

Exercise

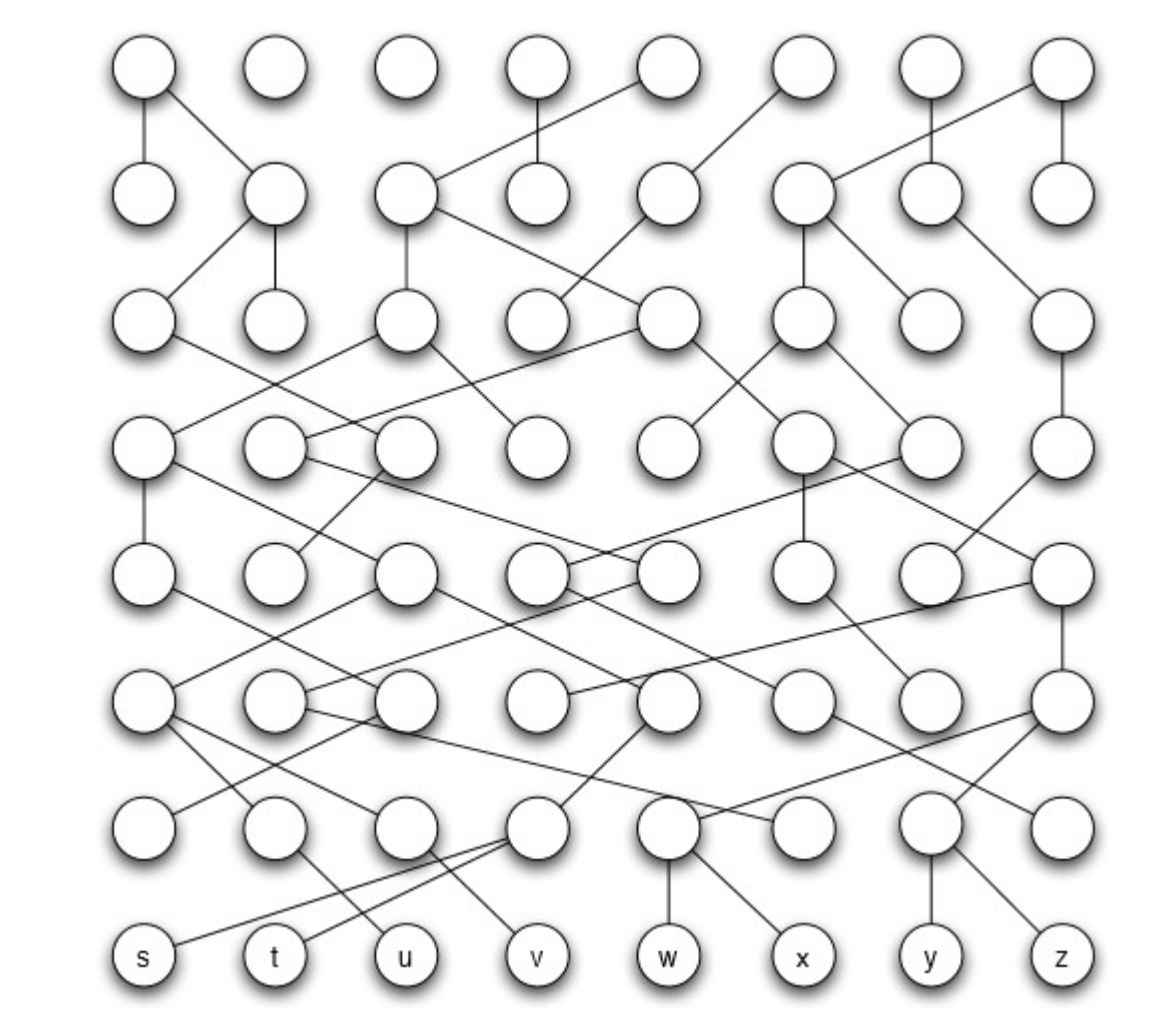


1. Suppose you are studying the spread of a rare disease among the set of people pictured in Figure above. The contacts among these people are as depicted in the network in the figure, with a time interval on each edge showing when the period of contact occurred. We assume that the period of observation runs from time 0 to time 20.
   a. Suppose that s is the only individual who had the disease at time 0. Which nodes could potentially have acquired the disease by the end of the observation period, at time 20?
   b. Suppose that you find, in fact, that all nodes have the disease at time 20. You're fairly certain that the disease couldn't have been introduced into this group from other sources, and so you suspect instead that a value you're using as the start or end of one of the time intervals is incorrect. Can you find a single number, designating the start or end of one of the time intervals, that you could change so that in the resulting network, it's possible for the disease to have flowed from s to every other node?

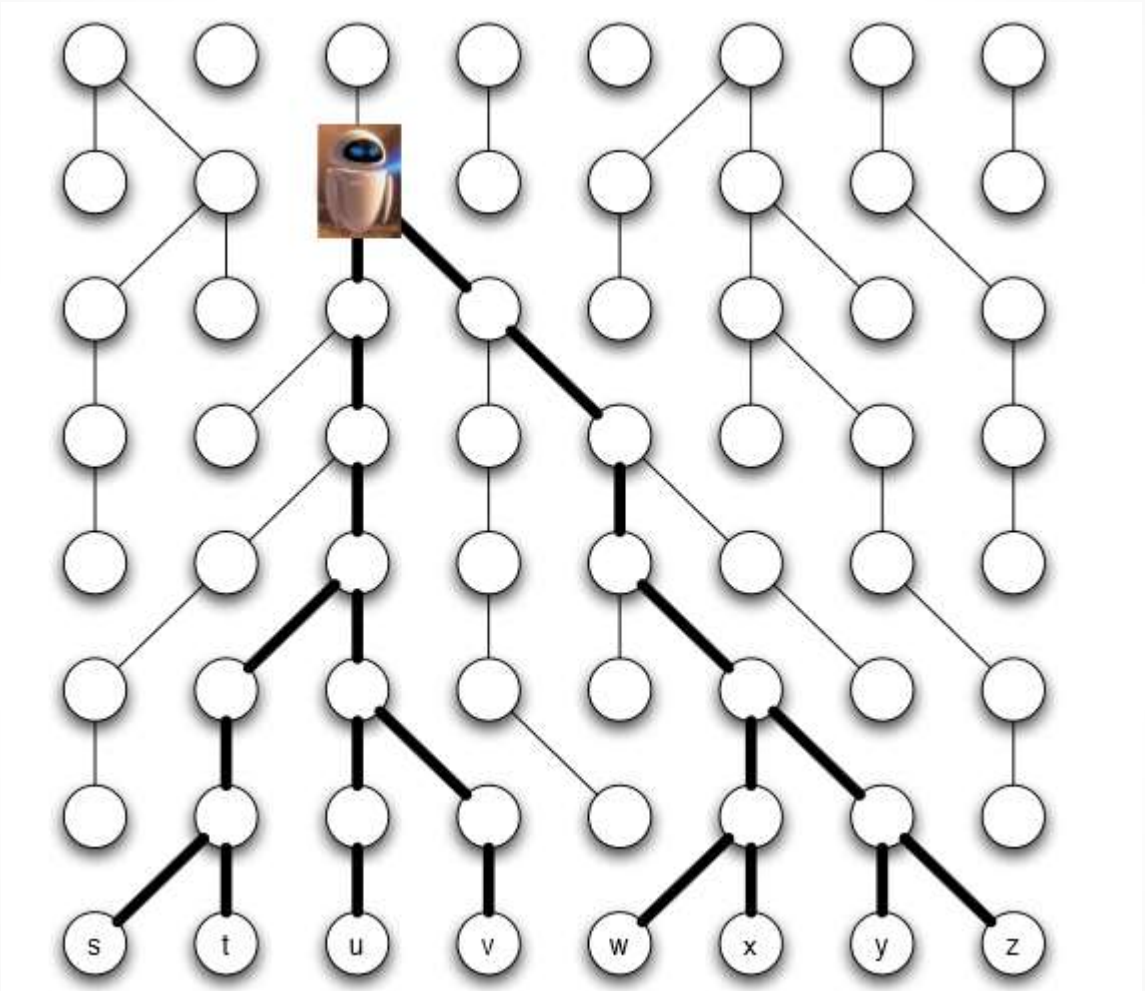
Mitochondral Eve

- Epidemics model can be used to model other things that spread through a network
- Genetic Inheritance
- Rebecca Cann, Mark Stoneking, and Allan Wilson Nature (1987)
  - Following our maternal ancestry
  - Maternal lineage
  - Meet at a single woman who lived between 100,000 and 200,000 years ago, probably in Africa
  - Study based on mitochondrial DNA which is (to a first approximation) passed to children entirely from their mothers

Study of Single Parent Ancestry



Discovering Common Ancestor



References

1. Networks, Crowds, and Markets: Reasoning about a Highly Connected World **Chapter 21.1-21.7** https://www.cs.cornell.edu/home/kleinber/networks-book/

- Mohamed A. Bahloul, Abderrazak Chahid, and Taous-Meriem Laleg-Kirati. 2020. Fractional-Order SEIQRDP Model for Simulating the Dynamics of COVID-19 Epidemic. IEEE Open Journal of Engineering in Medicine and Biology 1 (2020), 249–256. https://doi.org/10.1109/OJEMB.2020.3019758