Agent-based and network models

Arba Minch – Course 06

Julien Arino

December 2023

Motivation – More about incidence

Agent-based models (ABM)

Network models

Motivation – More about incidence

Agent-based models (ABM)

Network models

Incidence function versus force of infection

Two different forms for the rate of movement of S individuals from S to whatever infected compartment they end up in:

- $S' = -f(S, I, N)$ is an **incidence function**
- $S' = -\lambda(S, I, N)S$ is a **force of infection**

The two are of course essentially equivalent, the context tends to drive the form used. Advanced PDE models that consider for instance an age-of-infection structure need to integrate over $I(t, a)$ and thus often use force of infection, others are somewhat random..

Interactions – Infection

- Rate at which new cases appear per unit time is the *incidence function*

$$f(S, I, N)$$

- Depends of the number S of susceptible individuals, I of infectious individuals and, sometimes, of the total population N ▶ Incidence includes two main components
    - a denumeration of the number of contacts taking place
    - a description of the probability that such a contact, when it takes place, results in the transmission of the pathogen
- Choosing an appropriate function is hard and probably one of the flunkiest part of epidemic modelling

Two most frequently used functions

The two most frequently used incidence functions are **mass action incidence**

$$f(S,I,N)=\beta SI$$

and **standard** (or **proportional**) **incidence**

$$f(S,I,N)=\beta\frac{SI}{N}$$

In both cases, β is the *disease transmission coefficient*

Units of β

Recall that if $X(t)$ is the population in compartment X at time t, then X' has units number/time

In a differential equation, left and right hand side must have same units, so..

Mass action incidence

$$\beta SI \propto \beta \times \text{number} \times \text{number}$$

has units number/time if β has units 1/(number × time)

Standard incidence

$$\beta SI/N \propto \beta \times \text{number} \times \text{number}/\text{number} \propto \beta \times \text{number}$$

has units number/time if β has units 1/time

Mass action incidence

$$f(S, I, N) = \beta SI \tag{1}$$

- There is homogenous mixing of susceptible and infectious individuals
- Strong hypothesis: each individual potentially meets every other individual

In this case, one of the most widely accepted interpretations is that all susceptible individuals can come across all infectious individuals (hence the name, by analogy with gas dynamics in chemistry/physics)

When population is large, the hypothesis becomes unrealistic

Standard (proportional) incidence

The other form used frequently:

$$f(S, I, N) = \beta \frac{SI}{N} \tag{2}$$

Each susceptible individual meets a fraction of the infectious individuals

Or vice-versa! See, e.g., Hethcote, Qualitative analyses of communicable disease models, *Mathematical Biosciences* (1976)

Case of a larger population

Constant population $\implies$ (1) $\equiv$ (2)

When the total population is constant, a lot of incidence function are equivalent (to units)

Suppose $N(t) \equiv N_0$, then

$$\beta SI = \tilde{\beta}\frac{SI}{N} \iff \tilde{\beta} = N_0 \beta$$

and if the right hand side is satisfied, then (1) and (2) identical

Keep in mind units are different, though

General incidence

$$f(S, I, N) = \beta S^q I^p \tag{3}$$

These functions were introduced with data fitting in mind: fitting to data, find the p, q best matching the available data

Incidence with refuge

The following implements a refuge effect; it assumes that a proportion $0 < q < 1$ of the population is truly susceptible, because of, e.g., spatial heterogenities

$$f(S, I, N) = \begin{cases} \beta I \left(N - \dfrac{I}{q} \right), & \text{if } I < qN \\ 0, & \text{if } I \geq qN \end{cases} \tag{4}$$

Negative binomial incidence

$$f(S, I, N) = kS \ln \left(1 + \beta \frac{I}{k}\right) \tag{5}$$

For small values of k, this function describes a very concentrated infection process, while when $k \to \infty$, this function reduces to a mass action incidence

Asymptotic contact

$$f(S, I, N) = \frac{N}{1 - \varepsilon + \varepsilon N} \frac{F(S, I)}{N}$$

where F is one of the functions we just described
When $\varepsilon = 0$, contacts are proportionnal to N, whereas when $\varepsilon = 1$, contacts are independent from N

Asymptomatic transmission

$$f(S, I, N) = \beta \frac{SI}{c + S + I}$$

where c is a constant. E.g.,

$$\frac{C(N)}{N} F(S, I)$$

with $C(N) = N/(1 - \varepsilon + \varepsilon N)$ the function describing the contact rate and $F(S, I)$ the function describing disease spread, assumed here to be of negative binomial incidence-type

Switching incidence

Arino & McCluskey, [Effect of a sharp change of the incidence function on the dynamics of a simple disease](https://doi.org/10.1080/17513751003793017), *Journal of Biological Dynamics* (2010)
Scale population so switch occurs at $N = 1$ and suppose

$$F(S,I,N) = \begin{cases} \beta SI & \text{if } N \leq 1 \\ \beta \dfrac{SI}{N} & \text{if } N > 1 \end{cases}$$

In SIS with non-constant population

$$S' = bN - dS - F(S,I,N) + \gamma I \tag{6}$$
$$I' = F(S,I,N) - (d + \delta + \gamma)I \tag{7}$$

$b \neq d$ and δ disease-induced death rate, periodic solutions found

Why all this mess?

Because contact processes are really really complicated and that describing them mathematically is **hard**!

Motivation – More about incidence

Agent-based models (ABM)

Network models

Agent-based models (ABM)

- What are agent-based models
- When to use ABM
- When not to use ABM
- Some examples

ABM $\neq$ IBM

Early in the life of these models, they were called IBM (individual-based models)

Over the years, a “philosophical” distinction has emerged:

- IBM are mathematical models that consider individuals as the units; e.g., DTMC, CTMC, branching processes, etc.
- ABM are computational models whose study is, for the most part, only possible numerically

ABM vs Network models

Network models endow vertices with simple systems and couple them through graphs

Can be ABM, but some networks can also be studied analytically

Agent-based models (ABM)

- What are agent-based models
- When to use ABM
- When not to use ABM
- Some examples

ABM are very useful to decipher contact processes

Classic mathematical models capture contact by using approximations of what contact could be like

Classic models allow some flexibility (see section about incidence functions in Lecture X but they remain limited

ABM can model actual trajectories of individuals, so given a definition of what a contact is (how close do you need to be for a contact to take place), can count them efficaciously

ABM are very useful to understand behavioural responses

Agent-based models (ABM)
    What are agent-based models
    When to use ABM
    When not to use ABM
    Some examples

As with *all* tools, beware!

There is a law of large numbers effects happening often: if you have many units, unless some emergent behaviour arises, you get the same results using ODEs...

With this specific tool, beware!

There is a certain tendency in CS people to create *yet another* system and seek *adoption* by users

Agent-based models (ABM)

- What are agent-based models
- When to use ABM
- When not to use ABM
- Some examples

Antibiotic resistance in hospitals

D'Agata, Magal, Olivier, Ruan & Webb. Modeling antibiotic resistance in hospitals: The impact of minimizing treatment duration, Journal of Theoretical Biology (2007)

An IBM that's almost an ABM

This work is a good illustration of the "cultural proximity" between IBM and ABM

Model is stochastic and individual-based, in good enough form that approximating ODE can be derived

Allows for very specific tracking of the status of individuals through the process (almost an ABM in this sense)

The setup

Three processes:

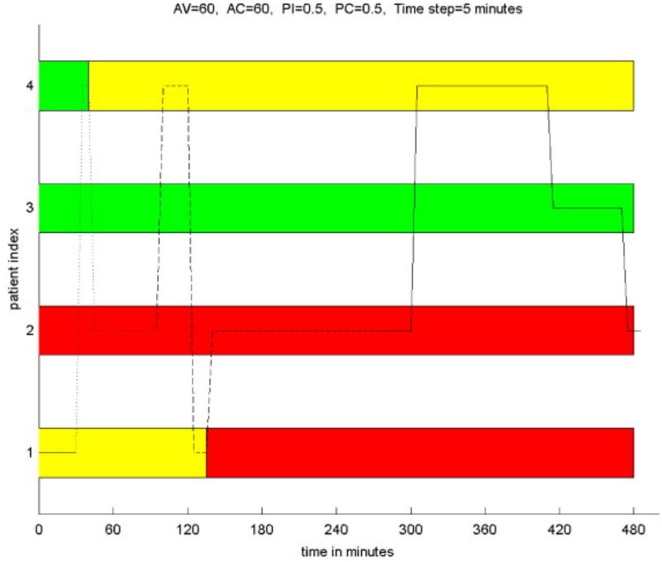
1. admission and exit of patients
2. infection of patients by HCW (health care workers)
3. contamination of HCW by patients

Contamination of HCW is “transient”: they are carriers, if they wash their hands properly, they become OK

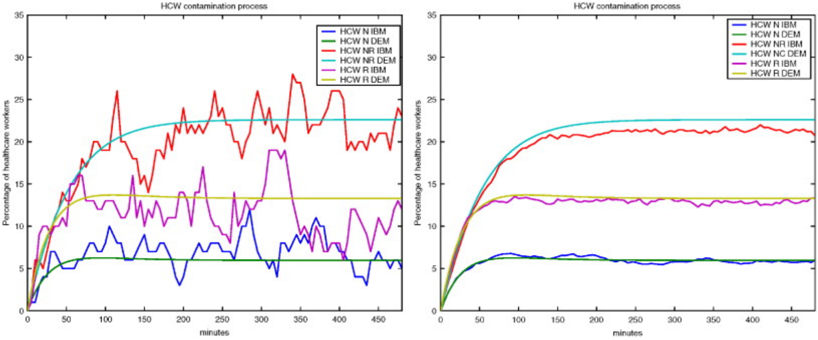
Each day has 3 shifts of 8h for HCW

Patients are put in contact by visits of HCW

Rules for contaminations per unit time



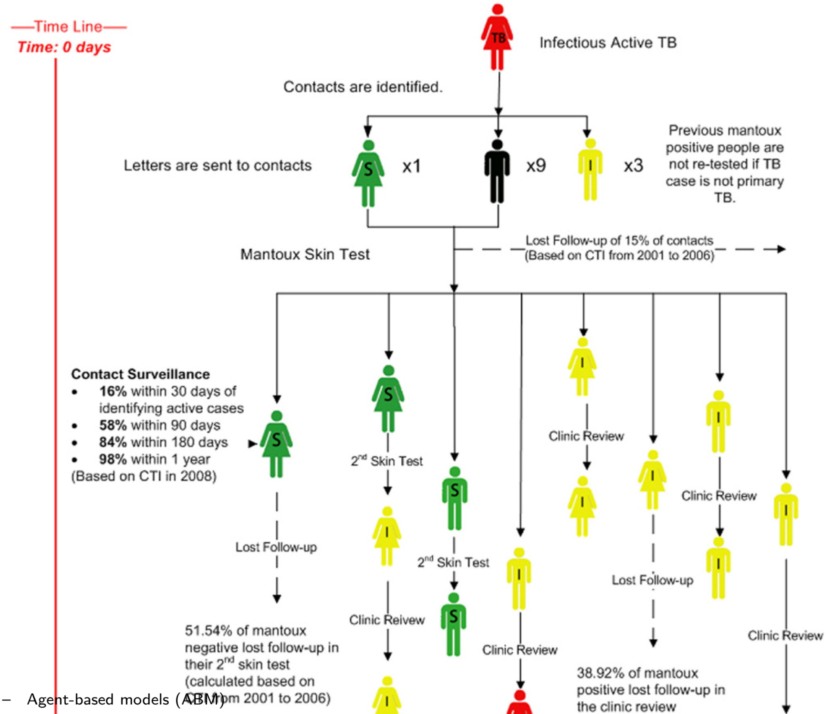
Patient–HCW contact diagram for four patients and one HCW during one shift. Patient status: uninfected (green), infected with the non-resistant strain (yellow), infected with the resistant strain (red). HCW status: uncontaminated (plain), contaminated with the non-resistant strain (dotted), contaminated with the resistant strain (dashed)



The left (respectively the right) figure corresponds to 1 trajectory (respectively the average over 80 trajectories) of the IBM during one shift, with no exit and admission of patients, and no changes in the infection status of patients

Effectiveness of contact tracing in TB

Tian, Osgood, Al-Azem & Hoeppner. Evaluating the Effectiveness of Contact Tracing on Tuberculosis Outcomes in Saskatchewan Using Individual-Based Modeling, Health Education & Behavior (2013)



—Time Line—

Time: 0 days

Infectious Active TB

Contacts are identified.

Letters are sent to contacts

Previous mantoux positive people are not re-tested if TB case is not primary TB.

Mantoux Skin Test

Lost Follow-up of 15% of contacts (Based on CTI from 2001 to 2006)

Contact Surveillance

- **16%** within 30 days of identifying active cases
- **58%** within 90 days
- **84%** within 180 days
- **98%** within 1 year

(Based on CTI in 2008)

Lost Follow-up

51.54% of mantoux negative lost follow-up in their 2nd skin test (calculated based on CTI from 2001 to 2006)

2nd Skin Test

Clinic Reivew

2nd Skin Test

Clinic Review

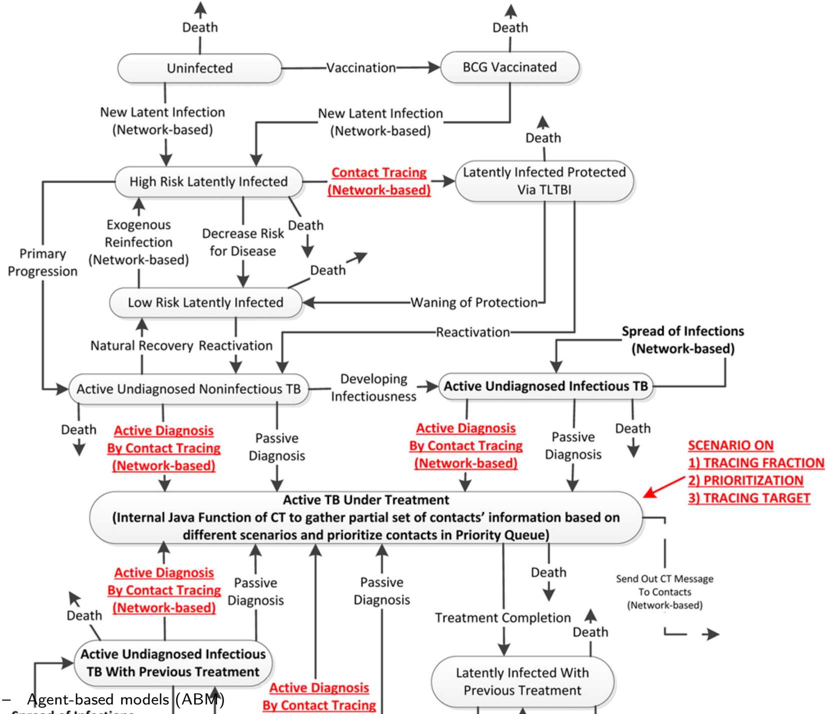
Clinic Review

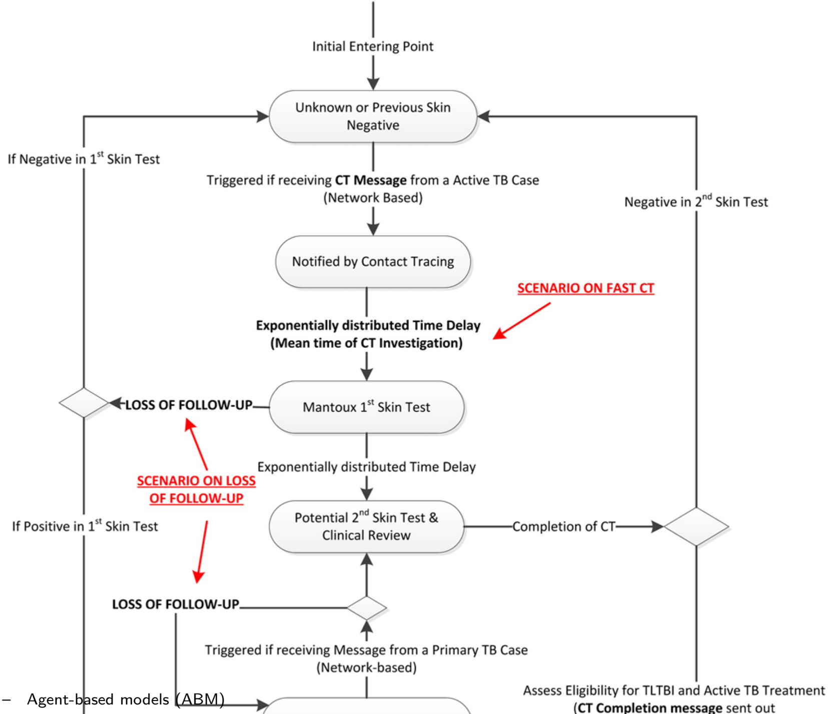
Lost Follow-up

38.92% of mantoux positive lost follow-up in the clinic review

Clinic Review

Clinic Review


Death
Death
Uninfected
Vaccination
BCG Vaccinated
New Latent Infection (Network-based)
New Latent Infection (Network-based)
Death
High Risk Latently Infected
Contact Tracing (Network-based)
Latently Infected Protected Via TLTBI
Primary Progression
Exogenous Reinfection (Network-based)
Decrease Risk for Disease
Death
Death
Low Risk Latently Infected
Waning of Protection
Natural Recovery
Reactivation
Reactivation
Spread of Infections (Network-based)
Active Undiagnosed Noninfectious TB
Developing Infectiousness
Active Undiagnosed Infectious TB
Death
Active Diagnosis By Contact Tracing (Network-based)
Passive Diagnosis
Active Diagnosis By Contact Tracing (Network-based)
Passive Diagnosis
Death
SCENARIO ON
1) TRACING FRACTION
2) PRIORITIZATION
3) TRACING TARGET
Active TB Under Treatment
(Internal Java Function of CT to gather partial set of contacts' information based on different scenarios and prioritize contacts in Priority Queue)
Active Diagnosis By Contact Tracing (Network-based)
Passive Diagnosis
Passive Diagnosis
Death
Send Out CT Message To Contacts (Network-based)
Death
Treatment Completion
Death
Active Undiagnosed Infectious TB With Previous Treatment
Active Diagnosis By Contact Tracing
Latently Infected With Previous Treatment
Spread of Infections


Initial Entering Point
Unknown or Previous Skin Negative
If Negative in 1st Skin Test
Triggered if receiving CT Message from a Active TB Case (Network Based)
Negative in 2nd Skin Test
Notified by Contact Tracing
SCENARIO ON FAST CT
Exponentially distributed Time Delay (Mean time of CT Investigation)
LOSS OF FOLLOW-UP
Mantoux 1st Skin Test
Exponentially distributed Time Delay
SCENARIO ON LOSS OF FOLLOW-UP
If Positive in 1st Skin Test
Potential 2nd Skin Test & Clinical Review
Completion of CT
LOSS OF FOLLOW-UP
Triggered if receiving Message from a Primary TB Case (Network-based)
Assess Eligibility for TLTBI and Active TB Treatment
(CT Completion message sent out

They can then formulate scenarios

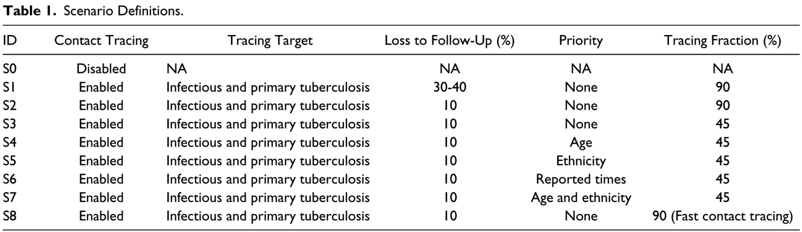
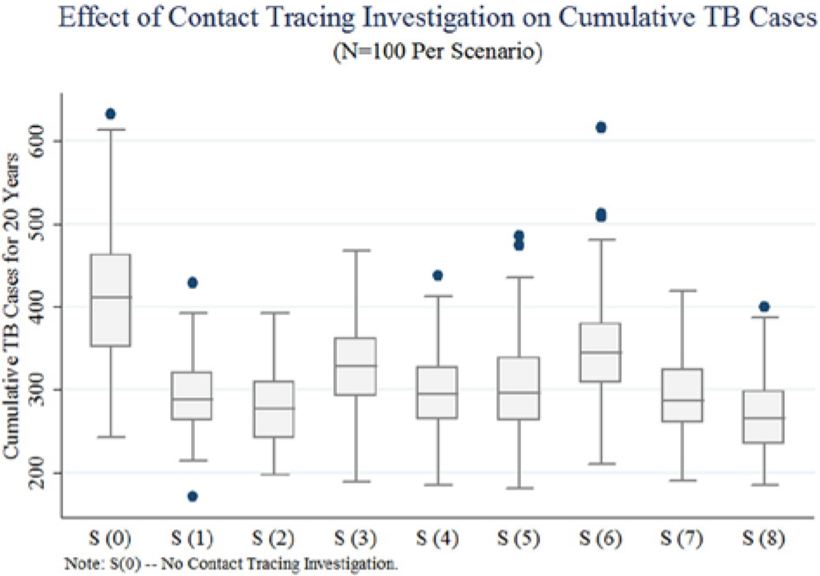


Table 1. Scenario Definitions.

ID	Contact Tracing	Tracing Target	Loss to Follow-Up (%)	Priority	Tracing Fraction (%)
S0	Disabled	NA	NA	NA	NA
S1	Enabled	Infectious and primary tuberculosis	30-40	None	90
S2	Enabled	Infectious and primary tuberculosis	10	None	90
S3	Enabled	Infectious and primary tuberculosis	10	None	45
S4	Enabled	Infectious and primary tuberculosis	10	Age	45
S5	Enabled	Infectious and primary tuberculosis	10	Ethnicity	45
S6	Enabled	Infectious and primary tuberculosis	10	Reported times	45
S7	Enabled	Infectious and primary tuberculosis	10	Age and ethnicity	45
S8	Enabled	Infectious and primary tuberculosis	10	None	90 (Fast contact tracing)

They then run these scenarios and compare results



Effect of Contact Tracing Investigation on Cumulative TB Cases

(N=100 Per Scenario)

Note: S(0) -- No Contact Tracing Investigation.

Contacts during Hajj

Tofighi, Asgary, Tofighi, Najafabadi, Arino, Amiche, Rahman, McCarthy, Bragazzi, Thommes, Coudeville, Grunnill, Bourouiba and Wu. Estimating Social Contacts in Mass Gatherings for Disease Outbreak Prevention and Management (Case of Hajj Pilgrimage), Tropical Diseases, Travel Medicine and Vaccines

Contacts during Hajj

In a mass gathering event like Hajj, lots of people come together originating from many countries

So if propagation occurs during the event, this has the capacity to spread infection far and wide when individuals (pilgrims here) return home

Contacts during part of the event are really specific in their configuration

The setup

Word of warning: I am quite fuzzy on the specifics :)

Pilgrims enter Masjid al-Haram mosque through several gates

Proceed to Mataaf (area around Kaaba), circle the Kaaba 7 times counterclockwise (process is the *Tawaf*)

Then do seven trips between Safa and Marwah (process is the *Sa'ee*)

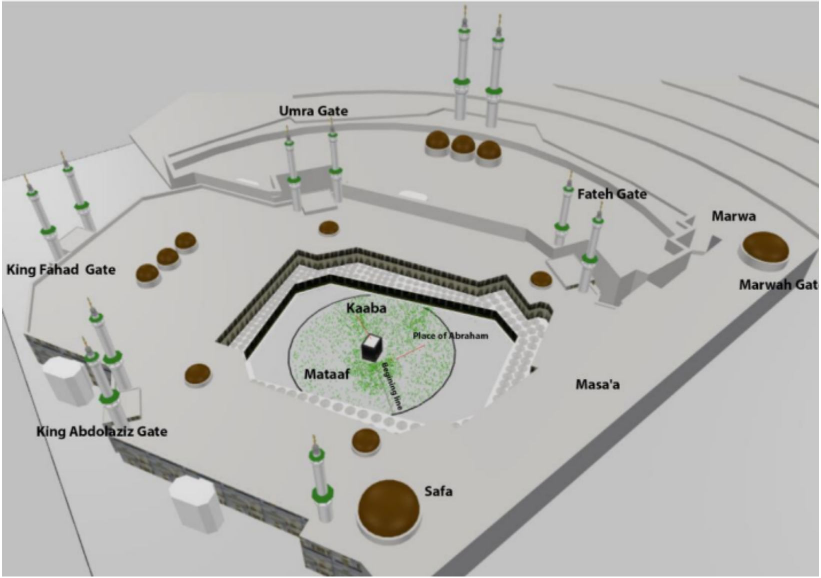
As you can gather from this:

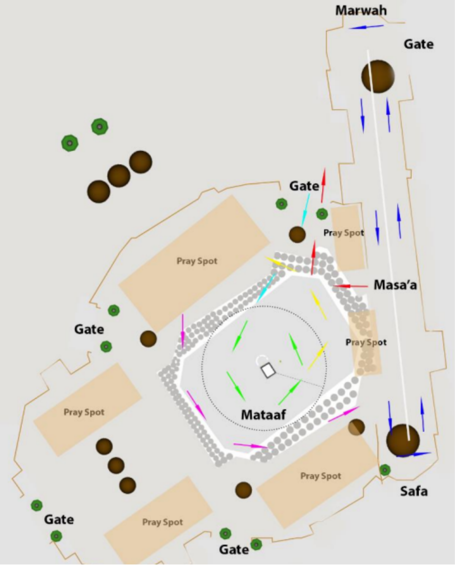
- Typically high density crowds
- Very specific mixing patterns

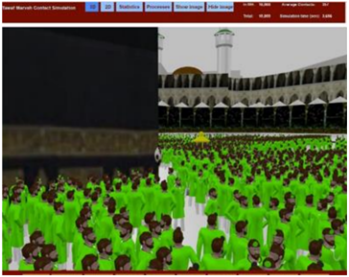
Opportunities for transmission are very high

However, control mechanisms are also available

$\Longrightarrow$ understanding contact patterns and frequency would help


Umra Gate
Fateh Gate
Marwa
King Fahad Gate
Marwah Gat
Kaaba
Place of Abraham
Mataaf
Begining line
Masa'a
King Abdolaziz Gate
Safa


Marwah
Gate
Gate
Pray Spot
Pray Spot
Masa'a
Gate
Pray Spot
Pray Spot
Mataaf
Pray Spot
Safa
Gate
Pray Spot
Gate


Tawaf Marwah Contact Simulation
3D
2D
Statistics
Processes
Show image
Hide image
Average Contacts:
Total:
Simulation time (sec):


Tawaf Marwah Contact Simulation
3D
2D
Statistics
Processes
Show image
Hide image
Average Contacts:
Total:
Simulation time (sec):

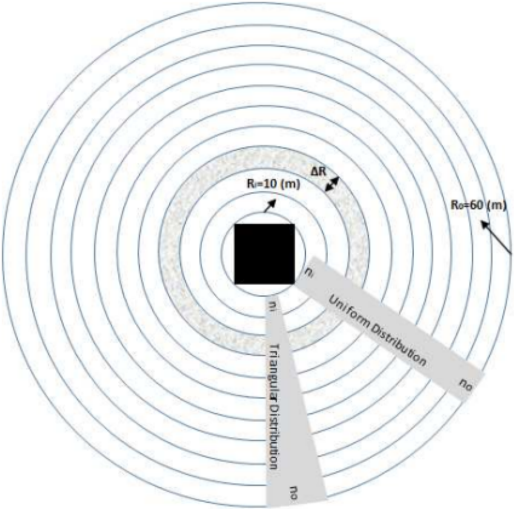


Figure 5. The general setting of Tawaf and pilgrim distribution in social distancing

Motivation – More about incidence

Agent-based models (ABM)

Network models

Network models
  - Why use network models
  - Social networks
  - Measures specific to vertices
  - Measures at the graph level
  - General context of network models

Understand contact processes

Classic models allow a certain degree of flexibility, for instance by using specific incidence functions or group models, but this remains limited and an approximation

Like ABM, network models are used to make more realistic descriptions of the transmission of pathogens

Human life is organised in networks

Family

Friends

Workplace

. . .

Social network theory has been used for years, e.g., in a professional context (e.g., how to fluidify interactions within a company)

Network models
    Why use network models
    Social networks
    Measures specific to vertices
    Measures at the graph level
    General context of network models

Before considering epidemics in networks, it is useful to learn a few notions of social network theory, as this is very useful to understand networks

Social network methods introduce measures that allow to evaluate some properties of graphs

A network is a (mathematical) graph, oriented or not, in which edges/arcs represent connections (whathever they are) between individuals, who make up the vertices of the graph

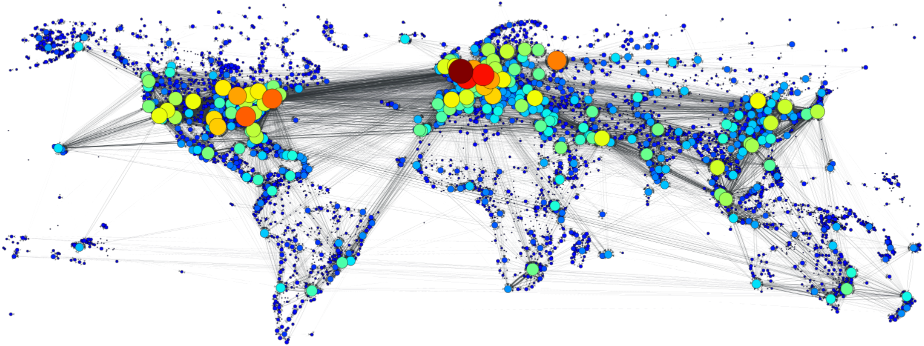
Context

- $\mathcal{G}(\mathcal{V}, \mathcal{E})$ an undirected graph
- $\mathcal{D}(\mathcal{V}, \mathcal{A})$ a digraph (directed graph)
- $\mathcal{V}$ the set of vertices (or nodes)
- $\mathcal{E}$ the set of edges (undirected case)
- $\mathcal{A}$ the set of arcs (directed graph)

Example of the global air transport network

- Je vais illustrer avec des données du réseau de transport aérien mondial - Données assez bonnes (très bonnes parfois), et un avantage flagrant: - Quand un avion part de quelque part et arrive ailleurs, c'est quelque chose d'assez .. déterministe

The global air transportation network



Example of spread of p-H1N1

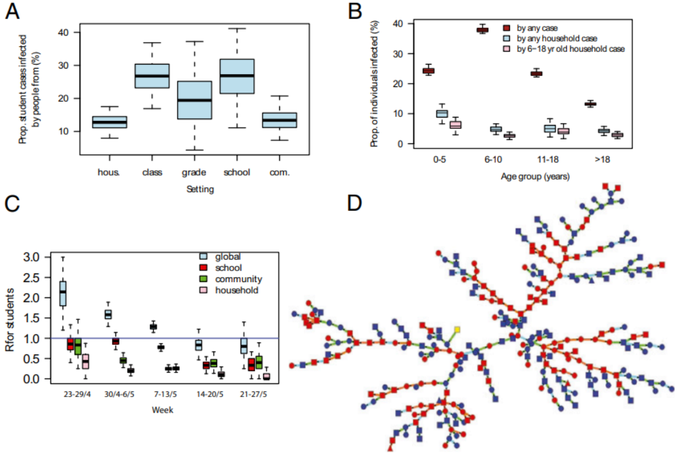
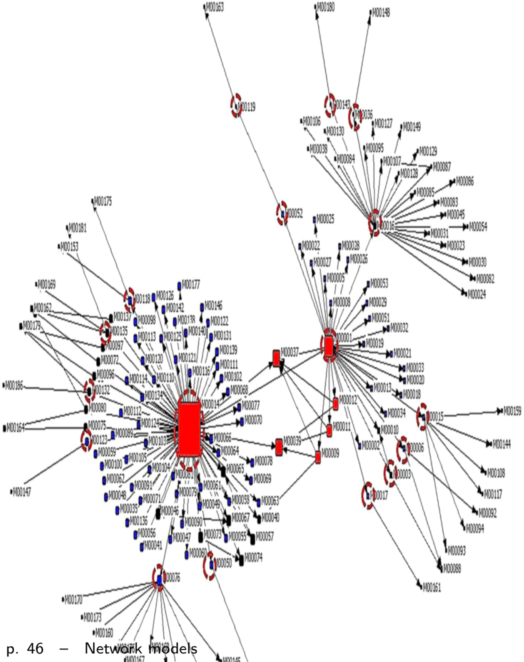


Fig. 4. Reconstruction of the transmission tree. (A) Proportion of student cases infected by people from their household, class, grade, school, or from the community. (*B*) Proportion of individuals infected by any other case (red), by any household case (blue), or by a household case aged 6–18 y (pink), as a function of the age of the individual. (*C*) Weekly estimates of the effective reproduction number in the outbreak ("global" R) and in places (school, household, and community). (*D*) Reconstructed transmission tree drawn from its predictive distribution (color of the nodes, yellow, first case; red, student of the school; blue, household member of a student; color of the lines for the type of transmission, orange, among students of the school; light blue, among household members; green, in the community; shape of the nodes, circle, female; square, male; triangle, sex is missing). Boxplots give percentiles 2.5%, 25%, 50%, 75%, and 97.5% of the predictive distribution.

2828 | www.pnas.org/cgi/doi/10.1073/pnas.1008895108 Cauchemez et al.

Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza, Cauchemez *et al*, PNAS **108**(7):2825-2830 (2011)

Example of spread of MERS



Topological dynamics of the 2015 South Korea MERS-CoV spread-on-contact networks, Yang & Jung, Scientific Reports **10**:4327 (2020)

Some “measures” concern the vertices, others the graph as a whole
In all that follows, unless otherwise indicated, $G = (X, A)$ is a digraph. If undirected, we write $G = (X, E)$.

Network models

- Why use network models
- Social networks
- Measures specific to vertices
- Measures at the graph level
- General context of network models

Network models

- Measures specific to vertices
  - Centre of a graph
  - Centrality – Betweenness and closeness
  - Periphery of a graph
  - Degree distribution
  - Circumference & Girth
  - Graph density
  - Graph connectivity
  - Cliques
  - k-cores
  - General context of network models

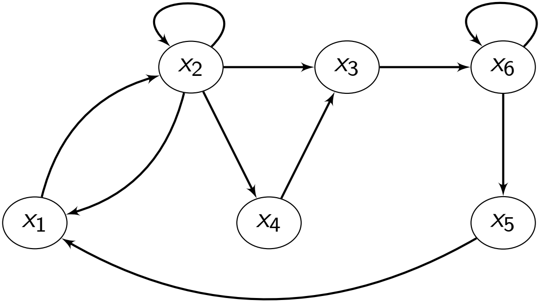
Geodesic distance

Definition 1 (Geodesic distance)

For $x, y \in X$, the **geodesic distance** $d(x, y)$ is the length of the shortest path from x to y, with $d(x, y) = \infty$ if no such path exists

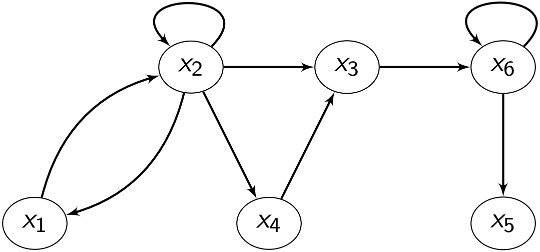
- $d(x_1, x_2) = 1$
- $d(x_1, x_3) = 2$
- $\ldots$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$



- $d(x_5, x_1) = \infty$
- $d(x_3, x_1) = \infty$
- ...

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ \infty & \infty & 0 & \infty & 2 & 1 \\ \infty & \infty & 1 & 0 & 3 & 2 \\ \infty & \infty & \infty & \infty & 0 & \infty \\ \infty & \infty & \infty & \infty & 1 & 0 \end{pmatrix}$$



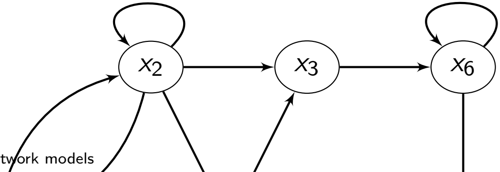
Eccentricity

Definition 2 (Vertex eccentricity)

The **eccentricity** $e(x)$ of vertex $x \in X$ is

$$e(x) = \max_{\substack{y \in X \\ y \neq x}} d(x, y)$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & \mathbf{4} & 3 \\ 1 & 0 & 1 & 1 & \mathbf{3} & 2 \\ 3 & 4 & 0 & \mathbf{5} & 2 & 1 \\ 4 & \mathbf{5} & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & \mathbf{4} \\ 2 & 3 & \mathbf{4} & \mathbf{4} & 1 & 0 \end{pmatrix}$$



Central points, radius and centre

Definition 3 (Central point)

A **central point** of G is a vertex x_0 with smallest eccentricity

Definition 4 (Radius)

The **radius** of G is $\rho(G) = e(x_0)$, where x_0 is a centre of G In other words,

$$\rho(G) = \min_{x \in X} e(x)$$

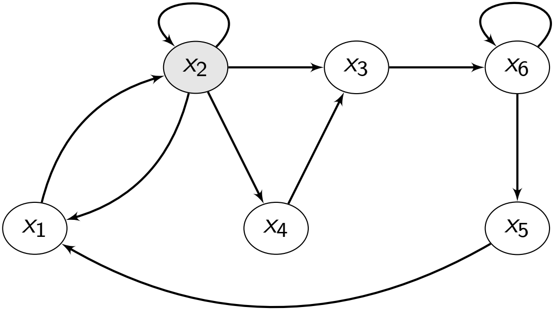
Definition 5 (Centre)

The **centre** of G is the set of vertices that are central points of G, i.e.,

$$\{x \in X : e(x) = \rho(G)\}$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & \mathbf{4} & 3 \\ 1 & 0 & 1 & 1 & \mathbf{3} & 2 \\ 3 & 4 & 0 & \mathbf{5} & 2 & 1 \\ 4 & \mathbf{5} & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & \mathbf{4} \\ 2 & 3 & \mathbf{4} & \mathbf{4} & 1 & 0 \end{pmatrix}$$

Radius is 3, x_2 is a central point (the only one) and the centre is $\{x_2\}$



Network models
  Measures specific to vertices
    Centre of a graph
    Centrality – Betweenness and closeness
    Periphery of a graph
    Degree distribution
    Circumference & Girth
    Graph density
    Graph connectivity
    Cliques
    k-cores
    General context of network models

How *central* is a vertex?

Centrality tries to answer the question: what are the most influent vertices?
We have seen central vertices and vertices on the periphery, let us consider two other measures of centrality

- Betweenness centrality
- Closeness centrality

Many other forms (we will come back to this, e.g., degree centrality)

Betweenness

Definition 6 (Betweenness)

$G = (X, U)$ a graph, $x \in X$. The **betweenness** of v is

$$b_{\mathcal{D}}(v) = \sum_{s \neq t \neq v \in X} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where

- σ_{st} number of shortest geodesic paths from s to t
- $\sigma_{st}(v)$ number of shortest geodesic paths from s to t through v

In other words

- For each pair of vertices (s, t), compute the shortest paths between them
- For each pair of vertices (s, t), determine the fraction of shortest paths that pass through vertex v
- Sum this fraction over all pairs of vertices (s, t)

Normalising betweenness

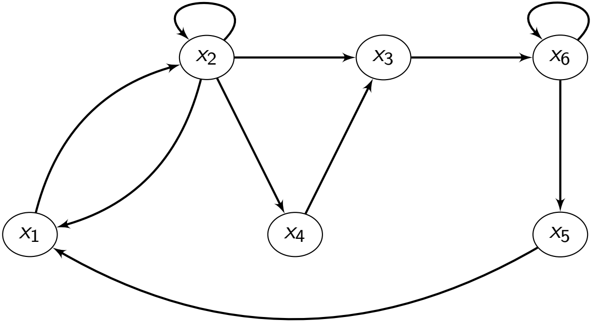
Betweenness may be normalized by dividing through the number of pairs of vertices not including v:

- for directed graphs, $(n-1)(n-2)$
- for undirected graphs, $(n-1)(n-2)/2$

Example of betweenness

```
distances(G,
mode="out")
```

$$\begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$



Number of shortest paths

Recall we found `distances(G, mode="out")`

$$\mathcal{D} = \begin{pmatrix} 0 & 1 & 2 & 2 & 4 & 3 \\ 1 & 0 & 1 & 1 & 3 & 2 \\ 3 & 4 & 0 & 5 & 2 & 1 \\ 4 & 5 & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & 4 \\ 2 & 3 & 4 & 4 & 1 & 0 \end{pmatrix}$$

To find the number of shortest paths between pairs of vertices, we can use powers of the adjacency matrix

Write $\mathcal{D} = [d_{ij}]$, for a given (i, j) $(i \neq j)$, if $d_{ij} = k$, then pick the (i, j) in A^k

We find

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Recall that betweenness of v is

$$b_{\mathcal{D}}(v) = \sum_{s \neq t \neq v \in X} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

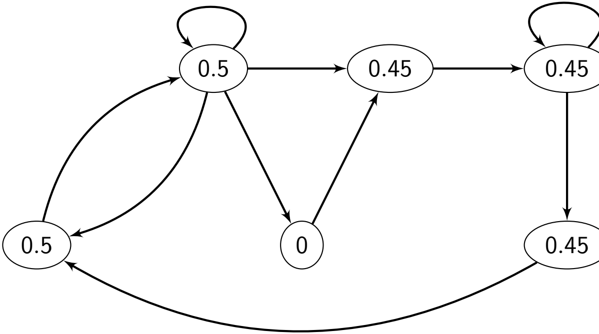
σ_{st} ($\#$ shortest paths from s to t) is found in the matrix above

What about $\sigma_{st}(v)$, $\#$ of those shortest paths that go through v?

We can use `all_shortest_paths(G, from = s, to = t, mode = "out")`

Example of betweenness

`betweenness(G, directed = FALSE, normalized = TRUE)`
Values shown in the vertices.



Closeness

Definition 7

$G = (X, U)$, $v \in X$. The **closeness** of v is

$$c_{\mathcal{D}}(v) = \frac{1}{n-1} \sum_{t \in X \setminus \{v\}} d_{\mathcal{D}}(v, t)$$

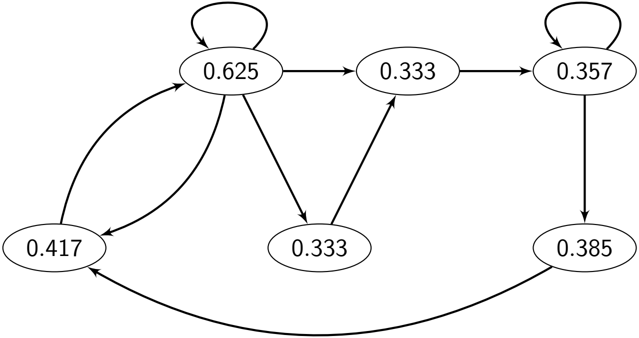
i.e., mean geodesic distance between a vertex v and all other vertices it has access to

Another definition is

$$c_{\mathcal{D}}(v) = \frac{1}{\sum_{t \in X \setminus \{v\}} d_{\mathcal{D}}(v, t)}$$

Example of (out) closeness

```
closeness(G, normalized = TRUE, mode=``out'')
```



- Network models
  - Measures specific to vertices
    - Centre of a graph
    - Centrality – Betweenness and closeness
    - Periphery of a graph
    - Degree distribution
    - Circumference & Girth
    - Graph density
    - Graph connectivity
    - Cliques
    - *k*-cores
    - General context of network models

Diametre and periphery of a graph

Definition 8 (Diametre of a graph)

The **diametre** of G is

$$\delta(G) = \max_{\substack{x,y\in X \\ x\neq y}} d(x,y)$$

or, in other words,

$$\delta(G) = \max_{x\in X} e(x)$$

$\delta(G) < \infty \iff G$ strongly connected

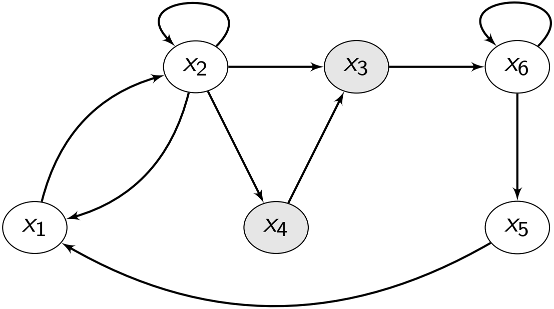
Definition 9 (Periphery)

The **periphery** of a graph is the set of vertices whose eccentricity achieves the diametre, i.e.,

$$\{x \in X : e(x) = \delta(G)\}$$

$$\begin{pmatrix} 0 & 1 & 2 & 2 & \mathbf{4} & 3 \\ 1 & 0 & 1 & 1 & \mathbf{3} & 2 \\ 3 & 4 & 0 & \mathbf{5} & 2 & 1 \\ 4 & \mathbf{5} & 1 & 0 & 3 & 2 \\ 1 & 2 & 3 & 3 & 0 & \mathbf{4} \\ 2 & 3 & \mathbf{4} & \mathbf{4} & 1 & 0 \end{pmatrix}$$

Diametre is 5 and the periphery is $\{x_3, x_4\}$



Definition 10 (Antipodal vertices)

Network models
  Measures specific to vertices
    Centre of a graph
    Centrality – Betweenness and closeness
    Periphery of a graph
    Degree distribution
    Circumference & Girth
    Graph density
    Graph connectivity
    Cliques
    k-cores
    General context of network models

Degree distribution

Definition 11 (Arc incident to a vertex)

If a vertex x is the initial endpoint of an arc u, which is not a loop, the arc u is **incident out of vertex** x

The number of arcs incident out of x plus the number of loops attached to x is denoted $d_G^+(x)$ and is the **outer demi-degree** of x

An arc **incident into vertex** x and the **inner demi-degree** $d_G^-(x)$ are defined similarly

Definition 12 (Degree)

The **degree** of vertex x is the number of arcs with x as an endpoint, each loop being counted twice. The degree of x is denoted $d_G(x) = d_G^+(x) + d_G^-(x)$

If each vertex has the same degree, the graph is **regular**

Definition 13 (Isolated vertex)

A vertex of degree 0 is **isolated**.

Definition 14 (Average degree of G)

$d(G) = \frac{1}{|V|} \sum_{v \in V} deg_G(v)$.

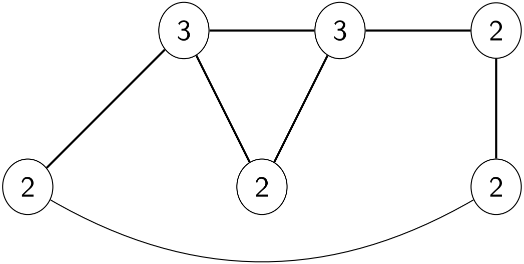
Definition 15 (Minimum degree of G)

$\delta(G) = \min\{deg_G(v) | v \in V\}$.

Definition 16 (Maximum degree of G)

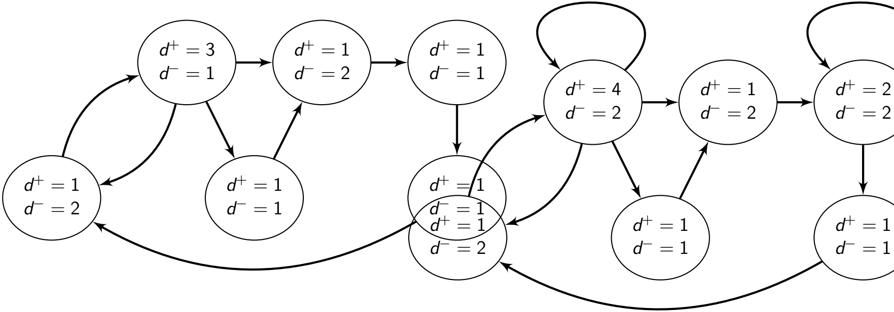
$\Delta(G) = \max\{deg_G(v) | v \in V\}$.

Degrees in an undirected graph



Here, vertices are labelled using the degree

Degrees in a directed graph



What to consider about degrees?

Degrees are often considered as a measure of popularity

Often write $k(i)$ (or k_i) for "degree of vertex i", $k^-(i)$ and $k^+(i)$ for in- and out-degree

- Minimum and maximum degree
- Minimum and maximum in/out-degree. E.g., if you consider the global air transportation network and the in/out-degree of airports, in-degree is a measure of a location's "popularity" as a travel destination
- Range of degrees in a graph: are there large discrepancies in connectivity between vertices in the graph?
- Average degree (often denoted $\langle k \rangle$ because of physicists)
- Average in/out-degree
- Variance of the degrees or in/out-degrees

- Average (nearest) neighbour degree, to encode for *preferential attachment* (one prefers to hang out with popular people)

$$k_i^{nn} = \frac{1}{k(i)} \sum_{j \in \mathcal{N}(i)} k(j)$$

or, in terms of the adjacency matrix $A = [a_{ij}]$,

$$k_i^{nn} = \frac{1}{k(i)} \sum_{j} a_{ij} k(j)$$

- *Excess degree*: take nearest neighbour degree but do not consider the edge/arc followed to get to the neighbour
- Degree, nearest neighbour and excess degree distributions

Degrees in `igraph`

- `degree` gives the degrees of the vertices
- `degree_distribution` gives numeric vector of the same length as the maximum degree plus one. The first element is the relative frequency zero degree vertices, the second vertices with degree one, etc.
- `knn` calculate the average nearest neighbor degree of the given vertices and the same quantity in the function of vertex degree
- `strength` sums up the edge weights of the adjacent edges for each vertex

Degree from adjacency matrix

Suppose adjacency matrix take the form $A = [a_{ij}]$ with $a_{ij} = 1$ if there is an arc from the vertex indexed i to the vertex indexed j and 0 otherwise. (Could be the other way round, using A^T, just make sure)

Let $\boldsymbol{e} = (1, \ldots, 1)^T$ be the vector of all ones

$A\boldsymbol{e} = (d_G^+(1), \ldots, d_G^+(1))^T$ (out-degree)

$\boldsymbol{e}^T A = (d_G^-(1), \ldots, d_G^-(1))$ (in-degree)

Network models
  - Why use network models
  - Social networks
  - Measures specific to vertices
  - Measures at the graph level
  - General context of network models

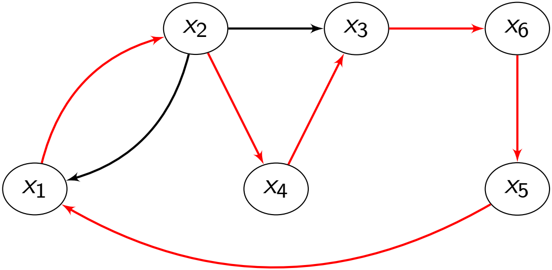
Network models
        - Centre of a graph
        - Centrality – Betweenness and closeness
        - Periphery of a graph
        - Degree distribution
    - Measures at the graph level
        - Circumference & Girth
        - Graph density
        - Graph connectivity
        - Cliques
        - k-cores
        - General context of network models

Circumference

Definition 17 (Circumference)

In an undirected (resp. directed) graph, the total number of edges (resp. arcs) in the longest cycle of graph G is the **circumference** of G

Circumference is 6.

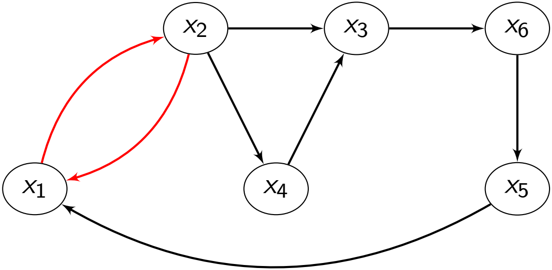


Girth

Definition 18 (Girth)

The total number of edges in the shortest cycle of graph G is the **girth** $g(G)$

Girth is 2.



Network models

- Centre of a graph
- Centrality – Betweenness and closeness
- Periphery of a graph
- Degree distribution

Measures at the graph level

- Circumference & Girth
- Graph density
- Graph connectivity
- Cliques
- k-cores
- General context of network models

Completeness

Definition 19 (Complete undirected graph)

An undirected graph is complete if every two of its vertices are adjacent.

Definition 20 (Complete digraph)

A digraph $D(V, A)$ is complete if $\forall u, v \in V,\ uv \in A$.

In case of simple graphs, completeness effectively means that "information" can be transmitted from every vertex to every other vertex quickly (1 step)

It can be useful to know how far away we are from being complete

Number of edges/arcs in a complete graph

$G = (X, E)$ undirected and simple of order n has at most

$$\frac{n(n-1)}{2}$$

edges, while $G = (X, A)$ directed and simple of order n has at most

$$n(n-1)$$

arcs

Density of a graph

Definition 21 (Density)

The fraction of maximum number of edges or arcs present in the graph is the **density** of the graph.

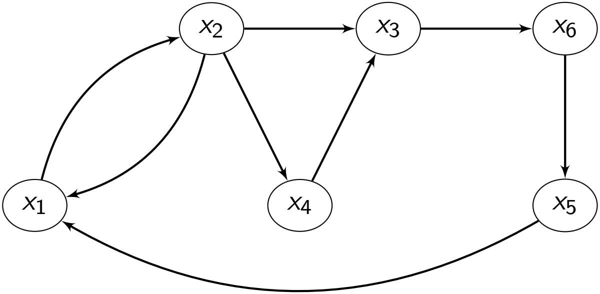
If the graph has p edges or arcs, then its density is, respectively,

$$\frac{2p}{n(n-1)}$$

or

$$\frac{p}{n(n-1)}$$

Example of density



Graph has order 6 and thus a max of 30 arcs. Here, 8 arcs $\Longrightarrow$ density 0.267 (26.7% of arcs are present)

Network models
- Centre of a graph
- Centrality – Betweenness and closeness
- Periphery of a graph
- Degree distribution

Measures at the graph level
- Circumference & Girth
- Graph density
- Graph connectivity
- Cliques
- *k*-cores
- General context of network models

Connectedness

We have already seen connectedness (quasi- or strong in the oriented case)

Connectedness is important in terms of characteristing graph properties, as it shows the capacity of the graph to convey information to all the members of the graph (the vertices)

Definition 22 (Connected graph)

A **connected graph** is a graph that contains a chain $\mu[x, y]$ for each pair x, y of distinct vertices

Denote $x \equiv y$ the relation "$x = y$, or $x \neq y$ and there exists a chain in G connecting x and y". $\equiv$ is an equivalence relation since

1. $x \equiv y$ [reflexivity]
2. $x \equiv y \implies y \equiv x$ [symmetry]
3. $x \equiv y, y \equiv z \implies x \equiv z$ [transitivity]

Definition 23 (Connected component of a graph)

The classes of the equivalence relation $\equiv$ partition X into connected sub-graphs of G called **connected components**

Articulation set

Definition 24 (Articulation set)

For a connected graph, a set A of vertices is called an **articulation set** (or a **cutset**) if the subgraph of G generated by $X - A$ is not connected

`articulation_points(G)` in `igraph` (assumes the graph is undirected, makes it so if not)

Strongly connected graphs

$G = (X, U)$ connected. A **path of length 0** is any sequence $\{x\}$ consisting of a single vertex $x \in X$

For $x, y \in X$, let $x \equiv y$ be the relation "there is a path $\mu_1[x, y]$ from x to y as well as a path $\mu_2[y, x]$ from y to x". This is an equivalence relation (it is reflexive, symmetric and transitive)

Definition 25 (Strong components)

Sets of the form

$$A(x_0) = \{x : x \in X, x \equiv x_0\}$$

are equivalence classes; they partition X and are the **strongly connected components** of G

Definition 26 (Strongly connected graph)

G **strongly connected** if it has a single strong component

Definition 27 (Minimally connected graph)

G is **minimally connected** if it is strongly connected and removal of any arc destroys strong-connectedness

Definition 28 (Contraction)

$G = (X, U)$. The **contraction** of the set $A \subset X$ of vertices consists in replacing A by a single vertex a and replacing each arc into (resp. out of) A by an arc with same index into (resp. out of) a

Quasi-strong connectedness

Definition 29 (Quasi-strong connectedness)

G **quasi-strongly connected** if $\forall x, y \in X$, exists $z \in X$ (denoted $z(x, y)$ to emphasize dependence on x, y) from which there is a path to x and a path to y

Strongly connected $\implies$ quasi-strongly connected (take $z(x, y) = x$); converse not true

Quasi-strongly connected $\implies$ connected

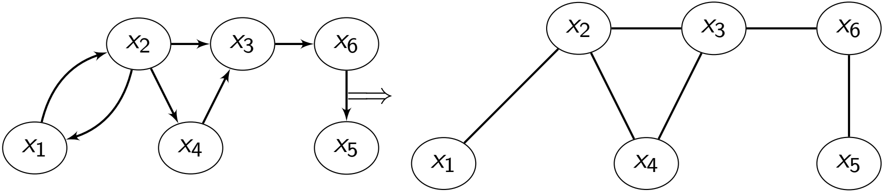
Lemma 30

$G = (X, U)$ has a root $\iff$ G quasi-strongly connected

Weak-connectedness

Definition 31 (Weakly connected graph)

$G = (X, U)$ **weakly connected** if $G = (X, E)$ connected, where E is obtained from U by ignoring the direction of arcs



Weak components

Define for $x, y \in X$ the relation $x \equiv y$ as "$x = y$ or $x \neq y$ and there is a chain in G connecting x and y" [like for components in an undirected graph, except the graph is directed here]

This defines an equivalence relation

Definition 32 (Weak components)

Sets of the form

$$A(x_0) = \{x : x \in X, x \equiv x_0\}$$

are equivalence classes partitioning X into the **weakly connected components** of G

$G = (X, U)$ is weakly connected if there is a single weak component

Components in `igraph`

- `is_connected` decides whether the graph is weakly or strongly connected
- `components` finds the maximal (weakly or strongly) connected components of a graph
- `count_components` does almost the same as components but returns only the number of clusters found instead of returning the actual clusters
- `component_distribution` creates a histogram for the maximal connected component sizes
- `decompose` creates a separate graph for each component of a graph
- `subcomponent` finds all vertices reachable from a given vertex, or the opposite: all vertices from which a given vertex is reachable via a directed path

Network models

- Centre of a graph
- Centrality – Betweenness and closeness
- Periphery of a graph
- Degree distribution

Measures at the graph level

- Circumference & Girth
- Graph density
- Graph connectivity
- Cliques
- *k*-cores
- General context of network models

Cliques

Definition 33 (Clique in undirected graphs)

$G = (X, E)$ a simple undirected graph. A **clique** is a subgraph G' of G such that all vertices in G' are adjacent

Definition 34 (n-clique)

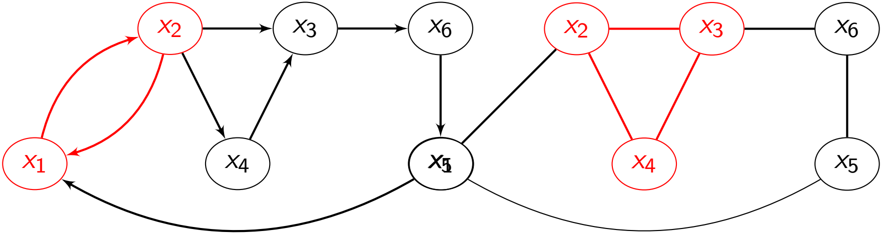
A simple, complete graph on n vertices is called an ***n*-clique** and is often denoted K_n

Definition 35 (Clique in directed graphs)

$G = (X, U)$ a simple directed graph. A **clique** is a subgraph G' of G such that all vertices in G' are mutually adjacent

Definition 36 (Maximal clique)

A **maximal clique** is a clique that cannot be extended by adding another adjacent vertex


x_1
x_2
x_3
x_4
x_5
x_6
x_2
x_3
x_4
x_5
x_6

Cliques in `igraph`

- `cliques` find all complete subgraphs in the input graph, obeying the size limitations given in the min and max arguments
- `largest_cliques` finds all largest cliques in the input graph
- `max_cliques` finds all maximal cliques in the input graph (The largest cliques are always maximal, but a maximal clique is not necessarily the largest)
- `count_max_cliques` counts the maximal cliques
- `clique_num` calculates the size of the largest clique(s)

Network models

- Centre of a graph
- Centrality – Betweenness and closeness
- Periphery of a graph
- Degree distribution

Measures at the graph level

- Circumference & Girth
- Graph density
- Graph connectivity
- Cliques
- *k*-cores
- General context of network models

k-core

Definition 37 (*k*-core of a graph)

$G = (X, U)$ a graph. The **k-core** of G is a maximal subgraph in which each vertex has degree at least k

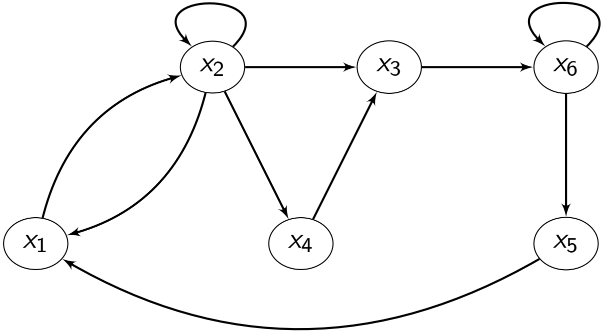
Definition 38 (Coreness of a vertex)

$G = (X, U)$ a graph, $x \in X$. The **coreness** of x is k if x belongs to the k-core of G but not to the $k + 1$ core of G

For directed graphs, in-cores or out-cores depending on whether in-degree or out-degree is used

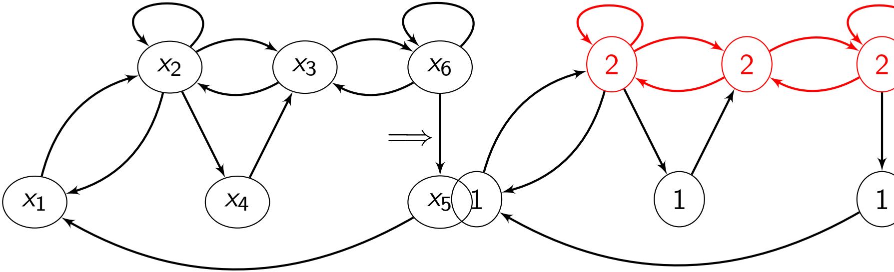
In `igraph`: `coreness`

Coreness in the directed case

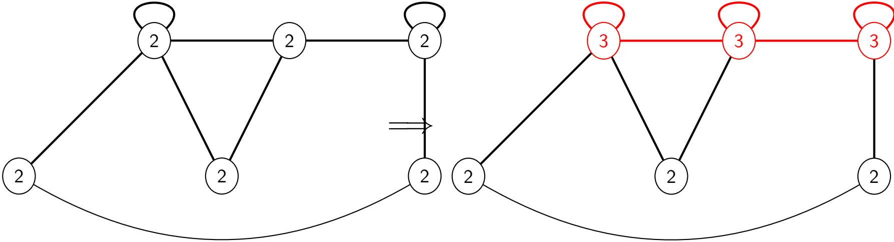


G has only a 1-in-core and 1-out-core: there is no (maximal) subgraph in which the in- or out-degree is larger than 1

In-coreness in the directed case



Coreness in the undirected case



Network models
  Why use network models
  Social networks
  Measures specific to vertices
  Measures at the graph level
  General context of network models

Network models
- Centre of a graph
- Centrality – Betweenness and closeness
- Periphery of a graph
- Degree distribution
- Circumference & Girth
- Graph density
- Graph connectivity
- Cliques
- *k*-cores

General context of network models
- General context of network models

See for instance...

- Newman. Spread of epidemic disease on networks, 2002
- Keeling & Eames. Networks and epidemic models, 2005
- Meyers, Pourbohloul, Newman, Skowronski & Brunham. Network theory and SARS: predicting outbreak diversity, 2005
- Meyers, Newman & Pourohloul. Predicting epidemics on directed contact networks, 2006
- Bansal, Read, Pourbohloul & Meyers. The dynamic nature of contact networks in infectious disease epidemiology, 2010

Typically, consider a graph (or digraph, depending on the cases) in which

- Each vertex is an individual
- Existence of an arc from i to j indicates that i is in contact with j and can transmit the pathogen to them
- dans le cas non orienté, l'existence d'un arc de i vers j implique celle d'un arc (le même) de j vers i et établit que les deux individus sont connectés
- La connexion n'est pas permanente, mais décrit plutôt la possibilité d'une connexion: i et j entrent en contact de façon régulière

Matrice d'adjacence
On utilisera souvent la **matrice d'adjacence** $A = [a_{ij}]$, dans laquelle $a_{ij} = 1$ si le nœud i a un lien vers le nœud j et $a_{ij} = 0$ sinon
On écrit parfois $A(\mathcal{D})$ pour indiquer que A est la matrice d'adjacence du digraphe $\mathcal{D}$, et dans l'autre sens, $\mathcal{D}(A)$ pour indiquer que le graphe est construit en utilisant la matrice d'adjacence
Si le graphe est non orienté, alors A est symmétrique

Nature du réseau
- Parfois on dispose de données précises sur les liens entre individus (sondages, etc.) - Souvent on idéalise des réseaux, on choisit des réseaux avec des propriétés données

Distribution of vertex degrees

The **transmissibility** of a disease in a graph/digraph is the average probability that an infectious individual transmits the disease to a susceptible individual they are in contact with

In an uncorrelated network,

$$T_c = \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

where $\langle k \rangle$ and $\langle k^2 \rangle$ are the mean and mean square degree

Il est nécessaire que $T > T_c$ pour qu'un *outbreak* devienne une épidémie majeure

The EpiModel library

`R` library providing tools to simulate and analyse network epidemiological models

Provides two types of approaches

- Simulation of ODE compartmental models (not so interesting)
- Simulation of network models

Their website has several useful tutorials

Part of the statnet meta-library