Applied economic networks - I

Ugo Bolletta

RITM, Université Paris-Saclay

September 27, 2021

About the exam

- 1 in-class presentation (1/3 of the final grade)
- 1 referee report (2/3 of the final grade)

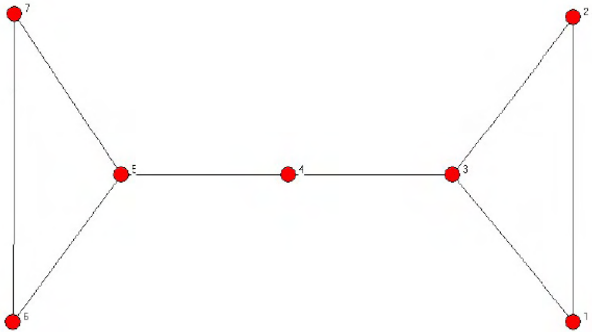
We will give you at the end of this class a list for the presentations. You can pick them up on a basis of first come first served. Each class there will be one presentation, starting next week

Centrality measures

1. **Degree centrality**: a basic measure of popularity. Take Twitter: the number of followers is the number of people who can be reached by a message; the potential to influence.
2. **Eigenvector centralities**: the popularity of your direct connections also matter.
3. **Diffusion centrality**: The number of steps needed to reach the maximum number of nodes.
4. **Betweenness centrality**: Is a node a key bridge?

Degree centrality

Often measured just by the **degree**, sometimes **normalized** by weighting for the number of nodes.



Closeness centrality

Distance based measure:
Call $\rho_g(i, j)$ the list of distances between node i and a node j.
Then:

$$\sum_j \rho_g(i, j)$$

However this measure is awkward.

Closeness centrality

Distance based measure:

Call $\rho_g(i,j)$ the list of distances between node i and a node j.
Then:

$$\sum_j \rho_g(i,j)$$

However this measure is awkward.

Closeness centrality

$$c_i^{cls}(g) = \frac{n-1}{\sum_{j \neq i} \rho_g(i,j)}$$

Decay centrality

Another measure based on distance: the decay in traveling along the shortest path.

Decay centrality

Defining with $\delta \in (0, 1)$ the decay factor:

$$c_i^\delta = \sum_{\ell \leq n-1} \delta^\ell n_i^\ell(g)$$

where $n_i^\ell(g)$ is the number of nodes at distance ℓ from node i, $n_i^\ell(g) = |\{j : \rho_g(i, j) = \ell\}|$

Betweenness centrality

The importance to connect other nodes.
It considers all paths between two nodes j, k different from i that pass through i.

Godfather centrality

$$GF_i(G) = (1 - clust_i(g))d_i(g)(d_i(g) - 1)/2 \tag{1}$$

Betweenness centrality

The importance to connect other nodes.
It considers all paths between two nodes j, k different from i that pass through i.

Godfather centrality

$$GF_i(G) = (1 - clust_i(g))d_i(g)(d_i(g) - 1)/2 \tag{1}$$

This is just one variation of *betweenness centrality*, but has a nice relation with clustering because it counts only connected pairs.

Eigenvector centralities

A step back:

The Neumann Series

If a matrix $\mathbf{A}$ has a spectral radius $r(A) < 1$ then

$$\sum_{\ell=0}^{\infty} \mathbf{A}^{\ell} = (\mathbf{I} - \mathbf{A})^{-1}$$

Eigenvector centralities

A step back:

The Neumann Series

If a matrix $\mathbf{A}$ has a spectral radius $r(A) < 1$ then

$$\sum_{\ell=0}^{\infty} \mathbf{A}^{\ell} = (\mathbf{I} - \mathbf{A})^{-1}$$

$\sum_{\ell=0}^{\infty} \mathbf{A}^{\ell}$: does it remind you something?

Bonacich centrality (Bonacich, 1984)

Now we take our matrix $\mathbf{A}$, a real number $\alpha \in (0, r(A)^{-1})$, and a vector of ones $\mathbf{b}$

Bonacich centrality

$\beta(\mathbf{A}; \alpha, \mathbf{b}) = (I - \alpha\mathbf{A})^{-1}\mathbf{b}$

Remark

The Bonacich centrality (eigenvector centralities in general) is proportional to the sum of neighbors' centralities.

Eigenvector centrality

Similar to Bonacich, but without imposing constraints on α.

The centrality is proportional to the centrality of neighbors.

$$\lambda c_i = \sum_j g_{ij} c_j$$

$$\lambda \mathbf{c} = g \mathbf{c}$$

We see that the centrality here corresponds to the eigenvector of the adjacency matrix corresponding to the greatest eigenvalue λ.

Eigenvector centrality

Similar to Bonacich, but without imposing constraints on α.

The centrality is proportional to the centrality of neighbors.

$$\lambda c_i = \sum_j g_{ij} c_j$$

$$\lambda \mathbf{c} = g\mathbf{c}$$

We see that the centrality here corresponds to the eigenvector of the adjacency matrix corresponding to the greatest eigenvalue λ.

Google page-rank

Google uses the eigenvector centrality to recommend you the links for your Google searches.

Diffusion centrality

- Based on a diffusion process starting at node i.
- In period 1 i passes a piece of information with probability δ t each neighbor.
- In any period ℓ, nodes that have received the information in period $\ell - 1$ pass the information with probability δ

Diffusion centrality

- Based on a diffusion process starting at node i.
- In period 1 i passes a piece of information with probability δ t each neighbor.
- In any period ℓ, nodes that have received the information in period $\ell - 1$ pass the information with probability δ

Diffusion centrality

$$c_i^{dif}(g, \delta, L) = \sum_{\ell=1}^{L} \sum_j \delta^\ell g_{ij}^\ell$$

- If $L = 1 \rightarrow$ degree centrality
- If $L \rightarrow \infty \rightarrow$ Bonacich centrality

Taxonomy of central measures

		Weighting			
				Infinite	
		Immediate	Extended	low δ	high δ
Nodal Statistic	Neighborhood/ Paths	**Degree**	**Decay, Closeness**	Not Applicable	
	Walks	**Degree**	**Diffusion**	**Bonacich**	**Eigenvector, Page Rank**
	Intermediary/ Geodesics	**Godfather**	**Betweenness**	Not Applicable	

Table 1: A Taxonomy of Centrality Measures

Figure: Source: Block, F. Jackson, M. and Tebaldi, P. (2021)

The strength of weak ties

Famous paper by Granovetter (1973)

- How strong is a tie? $\rightarrow$ number of times two individuals have met in a week
- Granovetter recorded the interactions, their intensity and whether the individuals found a job via the network
- 16.7% found a job through a strong tie, 55.7% through a medium tie and 27.6% through a weak tie
- Main intuition: weak ties are more likely to pass unheard information, they are bridges (the idea of betweenness)
- Weak ties are important "connectors" and allow for a proper flow of information, social influence and peer effects **even in fairly segregated** networks

Models of diffusion

Diffusion is interesting from several points of view:

- Pandemics
- Fake news
- Information..

Models of diffusion

Diffusion is interesting from several points of view:

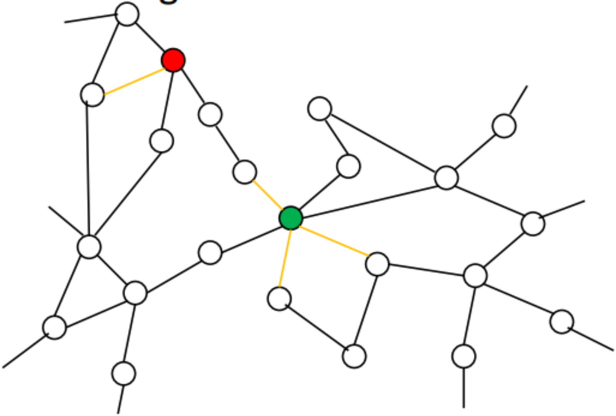
- Pandemics
- Fake news
- Information..

In these types of model we are usually interested in:

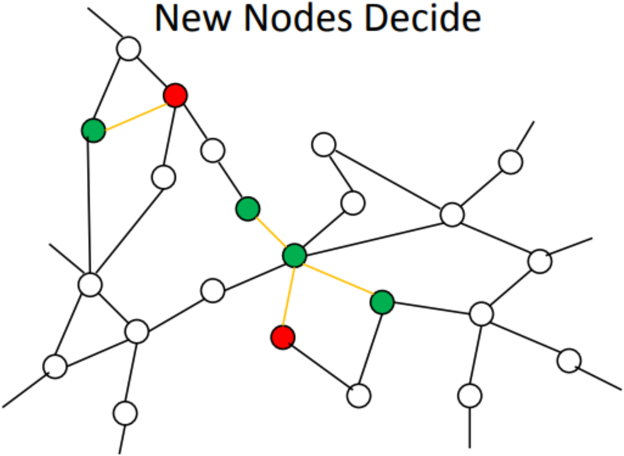
- how fast a unit of interest (virus, information..) reaches the greatest number of people,
- how many people we can reach by *seeding* one specific individual.

Diffusion of microfinance

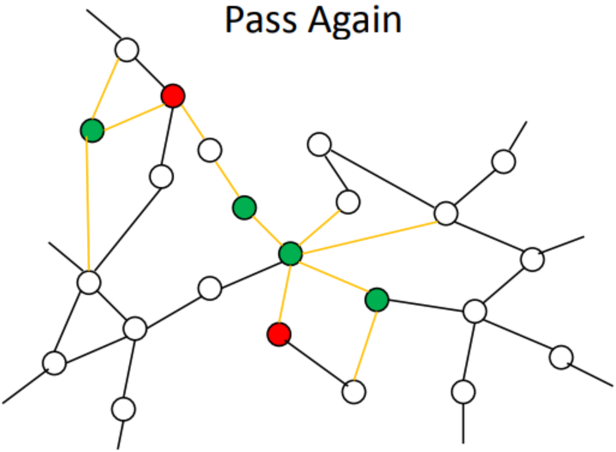
Passing: Different Probabilities



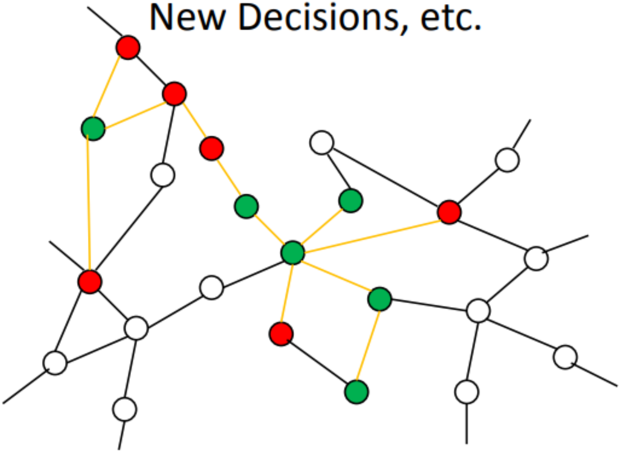
Diffusion of microfinance



Diffusion of microfinance



Diffusion of microfinance



SIR model

- $N = S(t)$

SIR model

- $N = S(t) + I(t)$

SIR model

- $N = S(t) + I(t) + R(t)$
- We set some initial conditions:
  - $S(0) = S_0, I(0) = I_0 \quad and \quad R(0) = 0$

Transition from a state to another

where β is the average number of contacts

- $\frac{dS}{dt} = -\frac{\beta IS}{N}$
- $\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I$
- $\frac{dR}{dt} = \gamma I$

SIR model (cont.)

Reproduction rate

The expected number of new infections from a new infection in a population of susceptible.

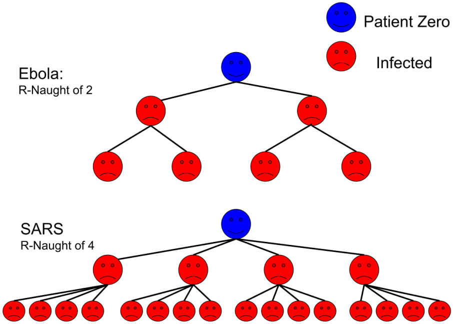
$$R_0 = \frac{\beta}{\gamma}$$

SIR model (cont.)

Reproduction rate

The expected number of new infections from a new infection in a population of susceptible.

$$R_0 = \frac{\beta}{\gamma}$$



The importance of R_0

Let's rewrite dI/dt:

$$\frac{dI}{dt} = (R_0\frac{S}{N} - 1)\gamma I$$

from which we obtain that infections increase if:

$$\frac{dI}{dt} > 0 \quad \text{iff} \quad R_0 S > N$$

and

$$\frac{dI}{dt} < 0 \quad \text{iff} \quad R_0 S < N$$

Solution of the model

On average, we need to see the number of contacts by an infectious individual before he/she recovers:

$$S(t) = S(0)e^{-\frac{\beta}{\gamma}(R(t)-R(0))/N}$$

We are interested in the dynamics for $t \to \infty$

Solution of the model

On average, we need to see the number of contacts by an infectious individual before he/she recovers:

$$S(t) = S(0)e^{-\frac{\beta}{\gamma}(R(t)-R(0))/N}$$

We are interested in the dynamics for $t \to \infty$

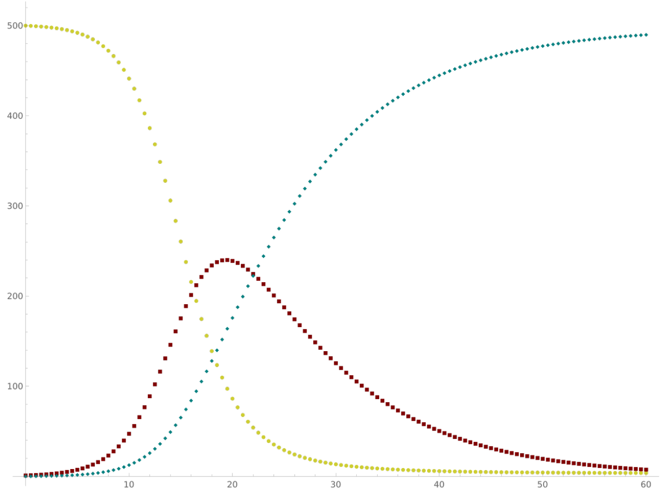
- Denote with $s_\infty = S(\infty)/N$ and with $r\infty = R(\infty)/N$. Then:

$$s_\infty = 1 - r_\infty = s_0 e^{\frac{\beta}{\gamma}(r_\infty - r_0)}$$

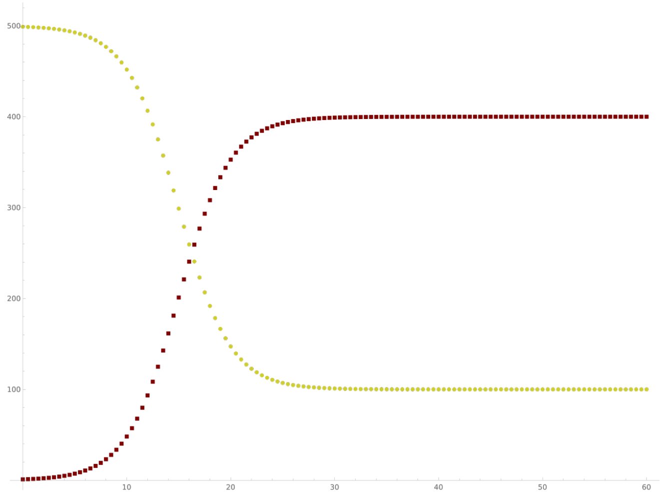
Remark

Unless $s_0 = 0$ (no-one is susceptible in the first place), the only way to fight the pandemics is by reducing infections.

Graphical representation of dynamics



What if we remain susceptible after being infected? SIS model



Trees and forests

The simplest form of the SIR model assumes that the network is very specific.

Trees

A tree is a connected network with no cycles.

- A connected network is a tree iff it has $n-1$ links.
- A tree has at least two leaves.
- In a tree there is a unique path between any two nodes.

Recap diffusion

There are a lot of models of diffusion:

- More or less sophisticated
- Used for pandemics:
  - We can include death probabilities, vaccines, and more complicated network structures.
- We can study information, rumors and overall spread into a network.
- So far one strong assumption: networks are fixed!

Diffusion and microfinance

Banerjee, Chandrasekar, Duflo, Jackson (2013)

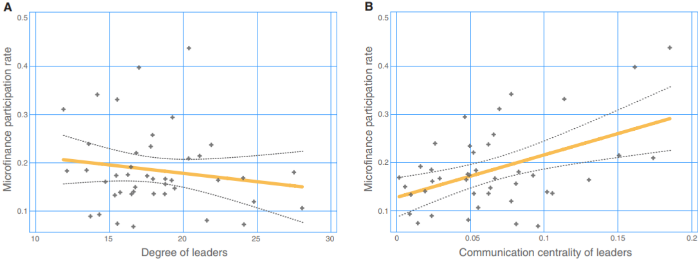
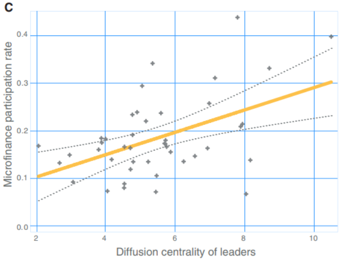


Fig. 2. Microfinance participation versus measures of leader centrality. All panels depict the correlation of village-level microfinance participation rate (*y* axis) against a measure of leader centrality (*x* axis); 95% confidence intervals are displayed. **(A)** Participation village-by-village as a function of the average degree of the leaders in the village. **(B)** Participation village-by-village as a function of the average communication centrality of the leaders the village. **(C)** Participation village-by-village as a function of the average diffusion centrality of the leaders in the village.



List of papers for presentations

Down below you can find the link to the spreadsheet to book your slot:

https://docs.google.com/spreadsheets/d/1PonIAbB4KMLvL2-CCHzDV2uPaQGOvWNZJup8Nzh0mAo/edit?usp=sharing

- Presentations are **15 minutes each**.
- The main goal of the presentation is that you convey the main message of the paper: everyone should exit the room with something learned through you.
- Time is limited so focus on central aspects, detect and report the details necessary for the overall understanding of the paper.
- Write me an email if you cannot find a slot (I counted 10 students, but I might have miscounted!)

Feel free to write me an email if you need any assistance during the presentation!

Thank you for your attention