MBA
USP
ESALQ

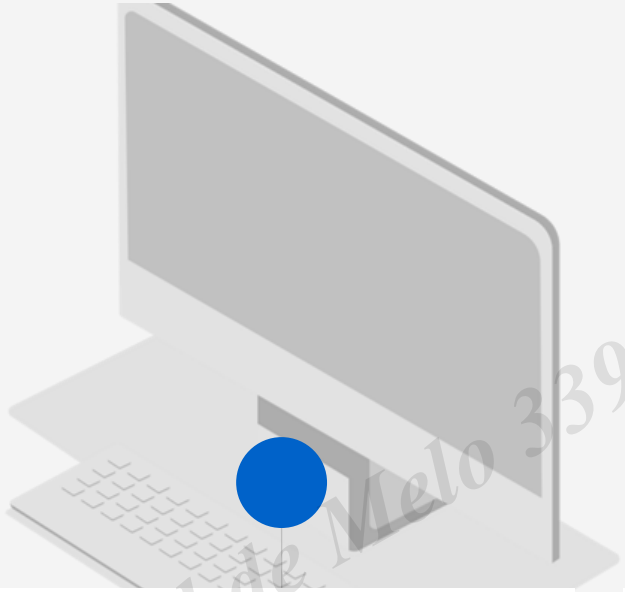
SOCIAL NETWORK ANALYSIS

Prof. Adriana Silva

*The responsibility for trustworthiness, originality, and legality of the didactic content presented is responsibility of the professor.

Program Content

Topics:



- What is SNA
  - What is for
  - How is it applied
  - Data structure

- Concepts and theory about the technique
  - Types of networks
  - Communities
  - Roles

- Calculations by hand for understanding the metrics
  - Centrality Metrics
    - Degree
    - Influence
    - Clustering Coefficient
    - Closeness
    - Betweenness

- Community
  - Resolution List
  - Application in Gephi for understanding
  - R for SNA
  - Exercises

Introduction to SNA

- *Social Network Analysis*
- *Customer Link Analysis*
- The power of knowledge of the relationships...

- It is not related to Facebook or social media

- It is an analysis based on interactions (calls, SMS, 3G, etc.)
- that are differentiated by the number and origin of the calls to others.

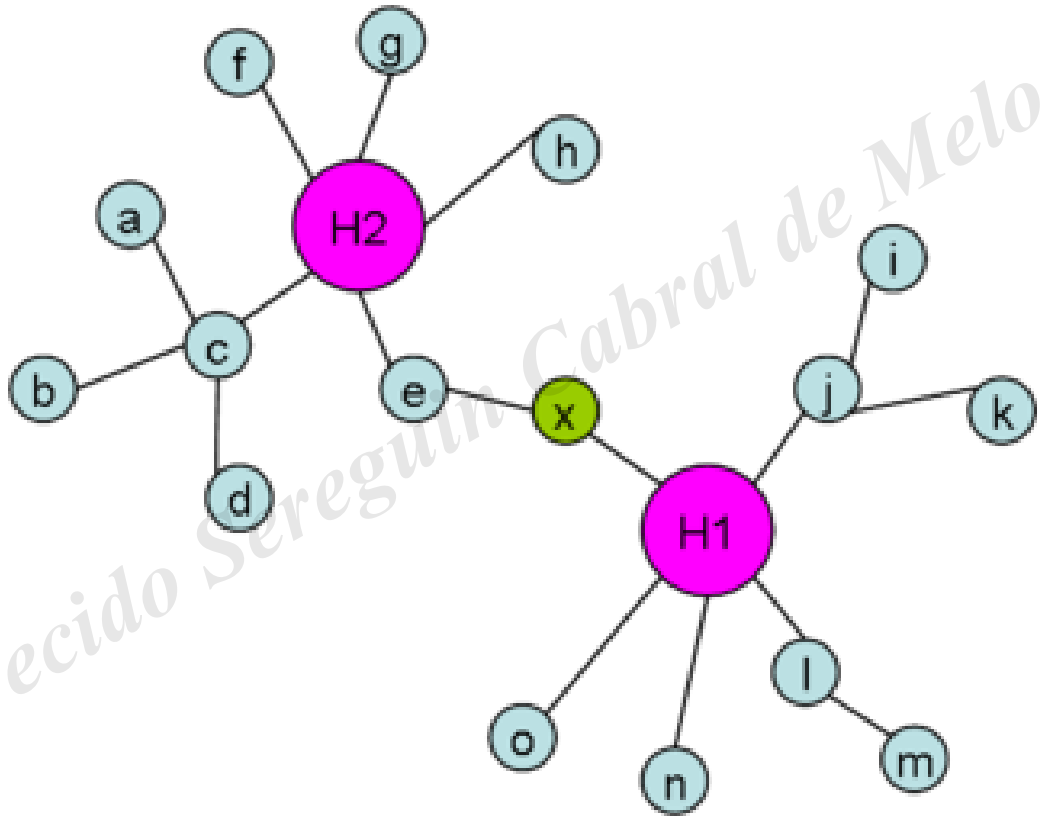
Social Network Analysis

SNA allows, for example, to the marketing analysts improve the customer vision, through the identification and incorporation of consumption relationships (and its attributes) and its strength, in profiles and segmentation.

Users can:

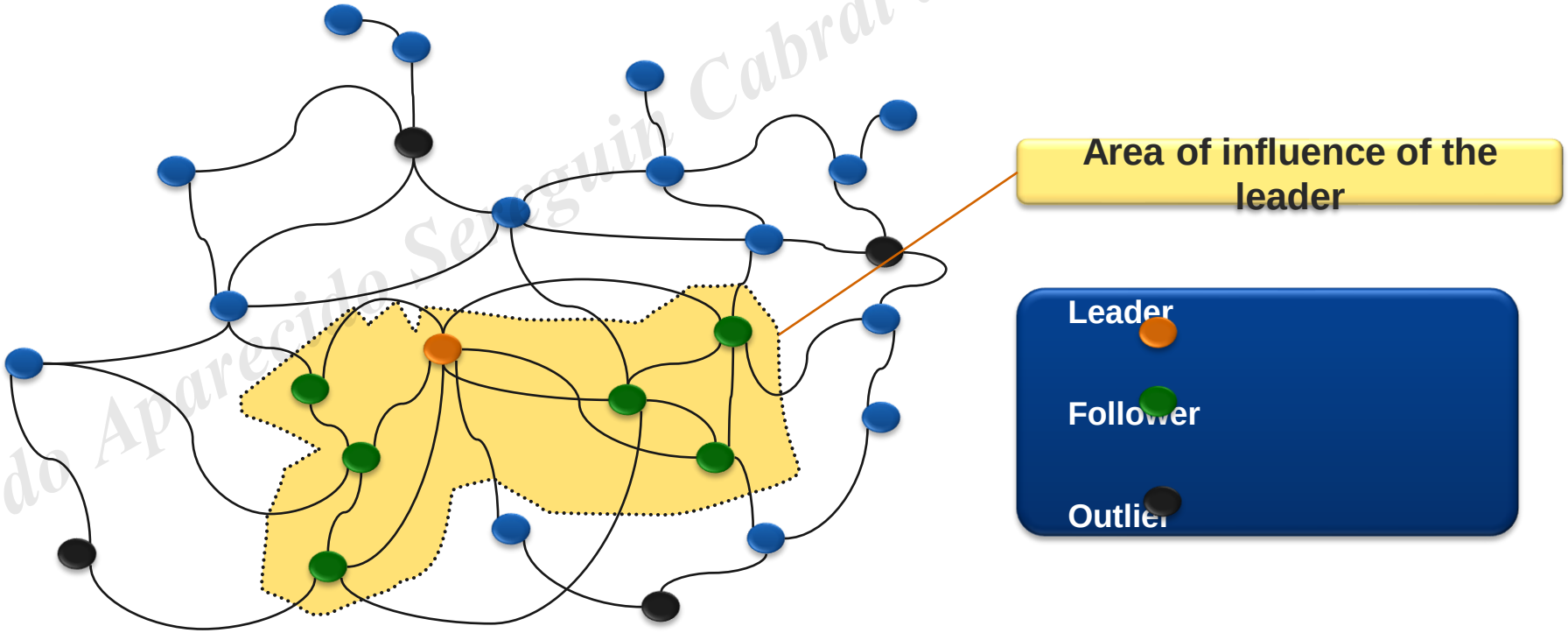
- Use in customer analysis as Network Metrics in order to improve the efficiency of the campaign
- Churn detection / prevention
- Up-sell and cross-sell
- Better understand how the products / services are adopted in the network (that is, the viral marketing)

Graph

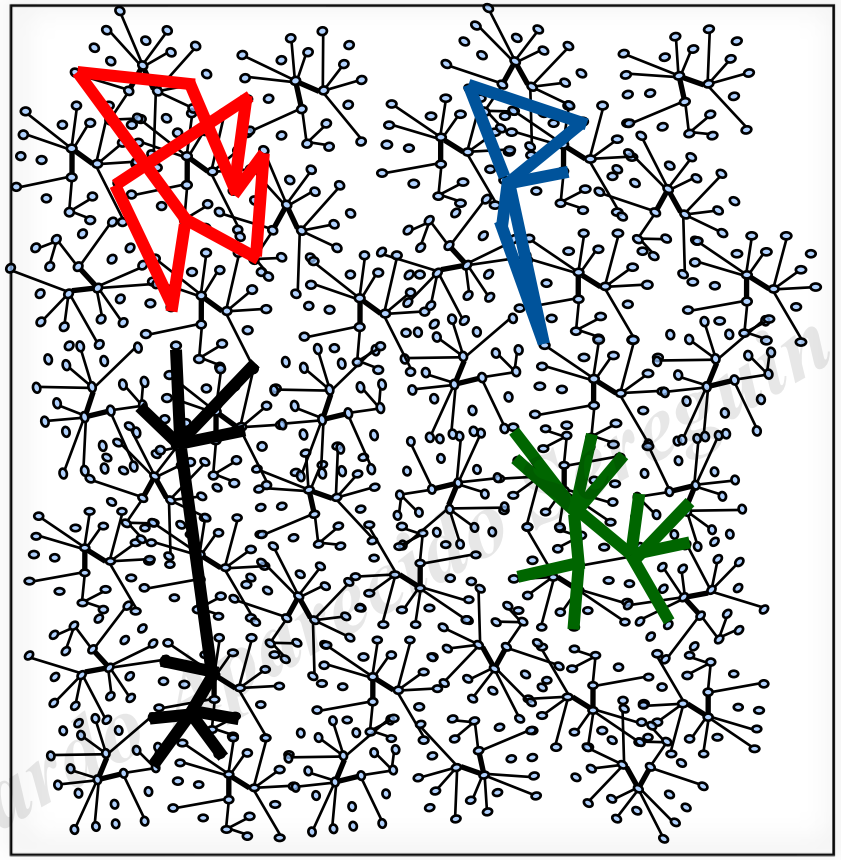


What is SNA?

- **Social Network Analysis** is the study of social structures composed by:
  - **nodes** (usually individuals or entities) that are linked
  - (**links**) by one or more types of interdependence, such as emails, telephone contacts, financial transactions, address, etc.



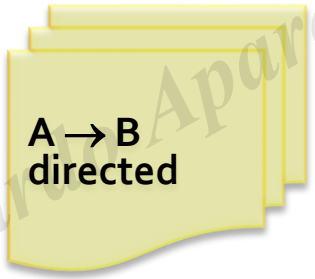
SNA: the origins



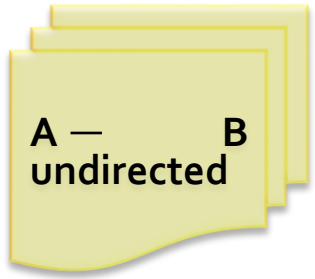
•Theory of "Six Degrees of Separation", Stanley Milgram

Social ties:

- The **nodes** of a Social Network are the elements that compose a given population.
- The **link** between two nodes defines the interaction between them, determined through "evidence".
- The study of all links between the nodes allows to measure the **structure** and **level of activity** of a population.
- Knowing the way that individuals are inter-related provides a analytical **base** for **understanding** their behaviors, opportunities, limitations, etc.



EXAMPLES OF EVIDENCES OF SOCIAL TIES:
Emails
Telephone bills
Addresses
Financial transactions



SNA applications

Application 1– Segmentation

Functionality:

•Point out customers according to their status in the network and links

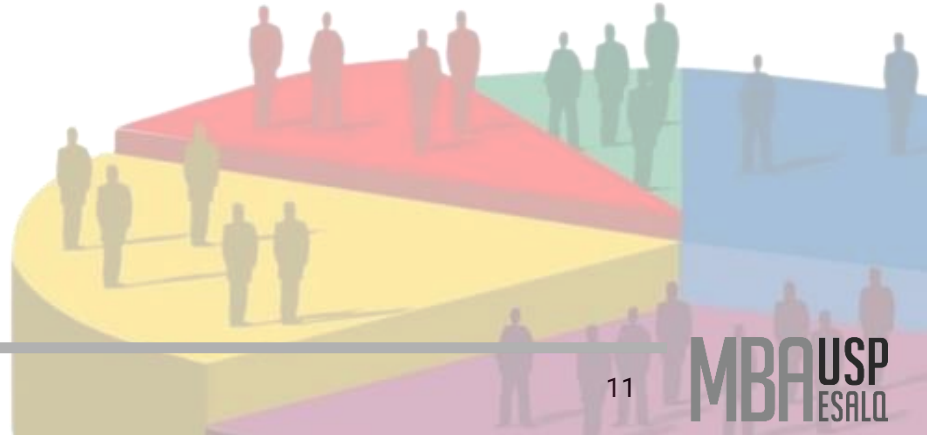
Marketing Action:

•Improve segmentation schemes beyond the profitability, life cycle, demography, etc)

Benefits:

•Strategies of "treatment" more appropriate for customers based on social networks

•More efficient marketing



SNA applications

Application 2a – "Leaders" Retention

Functionality:

•Identify "Leaders"

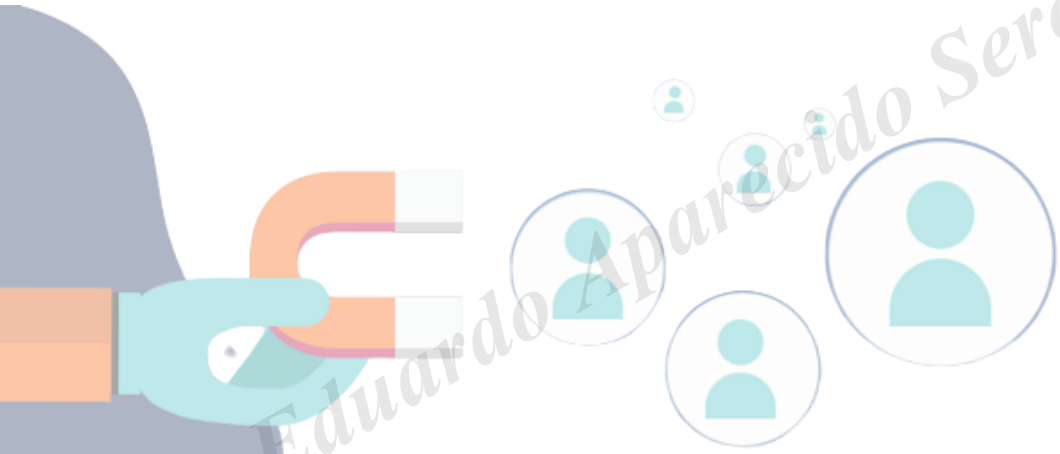
Marketing Action:

•Retention strategies for key leaders

Benefits:

•More efficiency in Marketing expenses

•Churn reduction / greater retention



SNA applications

Application 2b – "Followers" Retention

Functionality:

•Detect when a leader can "abandon"

Marketing Action:

•Retention strategies for followers at risk

Benefits:

•Efficiency in Marketing expenses
•Churn reduction / greater retention

SNA applications

Application 3 – Cross-sell in Leaders – Viral Effect

Functionality:

•Identify leaders and better understand the adoption of news

Marketing Actions:

•Cross strategies / up-sell firstly for leaders, promoting the viral adoption

Benefits:

•It rationalizes expenses, mainly in subsidy

SNA applications

Application 4 – Best selection in Acquisition

Functionality:

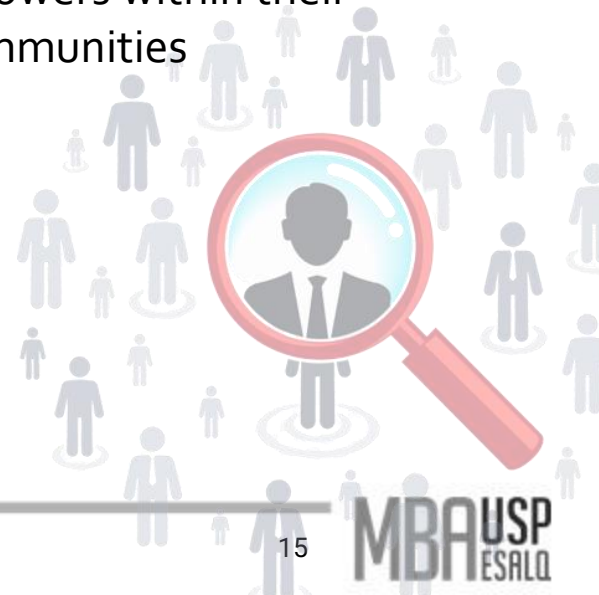
•Determine and understand the profile of target leaders

Marketing Actions:

•Acquisition strategies promoting the influencing power of leaders

Benefits:

•Attract leaders and followers within their communities



SNA concepts

Types of networks

Communities

Roles of individuals within the network

SNA concepts

Types of networks

Networks can be directed or undirected

- Undirected links means that there is no distinction between two vertices associated with each link, that is, there is no direction in the relation of the link that connects two links.
- Directed link means that there is a direction but relationships between nodes. If A calls 100 times for B

SNA concepts

Communities

- One of the most important objectives in networks analysis is the detection of structures known as Communities.
- They are defined intuitively as groups of nodes that are more strongly linked with each other than with the rest of the network.

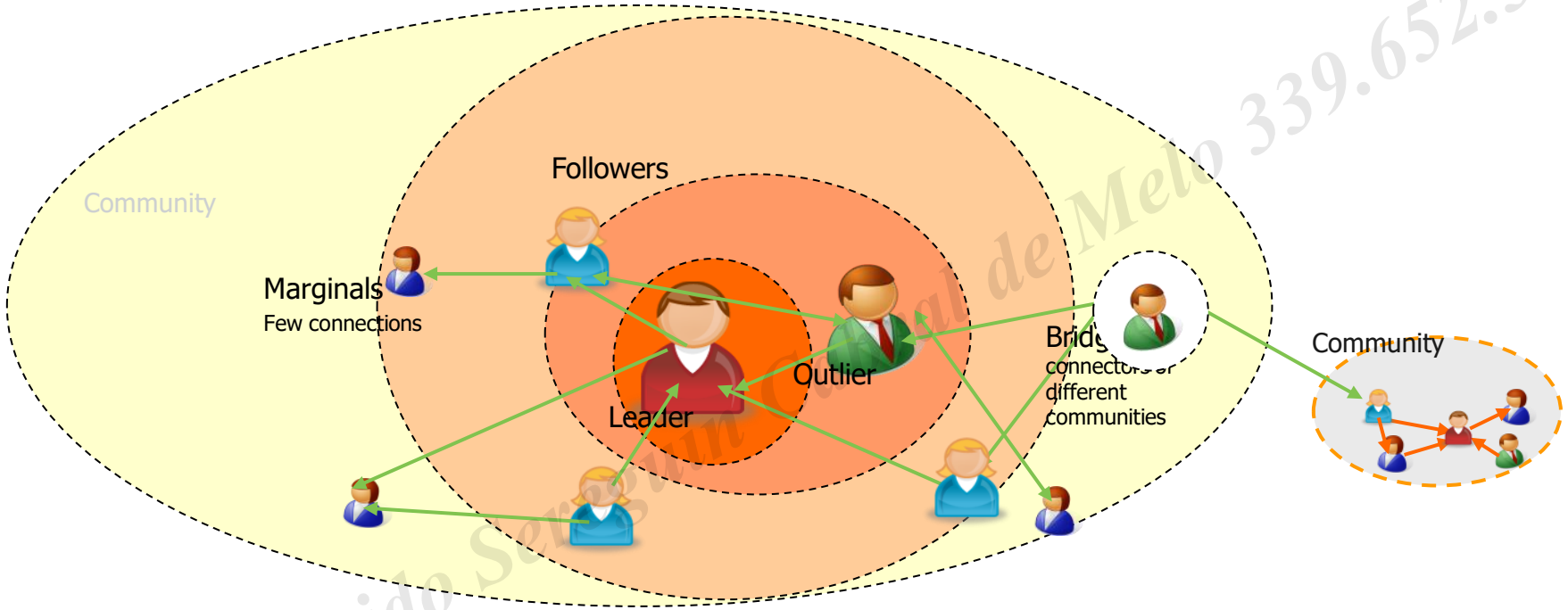
Application:

Internal Communities – when the predominance is of my brand.

External Community – when the predominance is of some competition.

SNA concepts

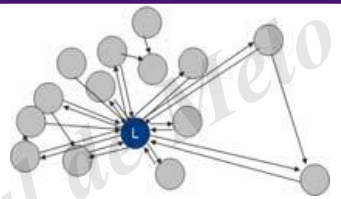
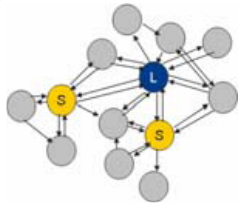
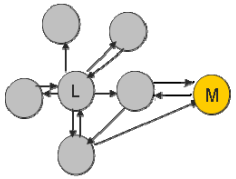
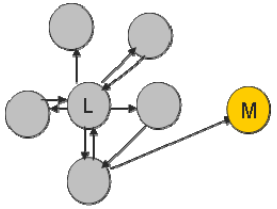
Roles and individual metrics



- Each node performs a role within its community
- Each node has several metrics of relationship, such as centrality, number of connections, etc.

SNA concepts

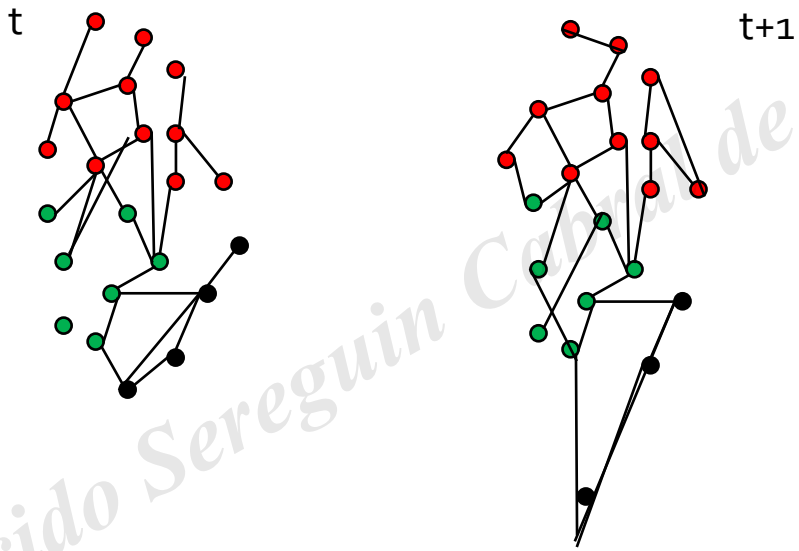
Roles and individual metrics

Rule	Description	Visualization
Leader	It is the telephone number within a community that has a reciprocal connectivity with other telephone numbers within the same community. It is the node in which the shortest path within the community (central point) is covered.	
Follower	It is the telephone number that has similar behavior to the leader, but with lower magnitude. It must have a direct connection with the leader.	
Marginal 1	Individual with a similar profile of the follower, but it is not close to the leader.	
Marginal 2	It remains on the fringe of the community. It does not have any characteristic of the other roles. It has a low proximity and scarce reciprocal relationships.	

SNA concepts

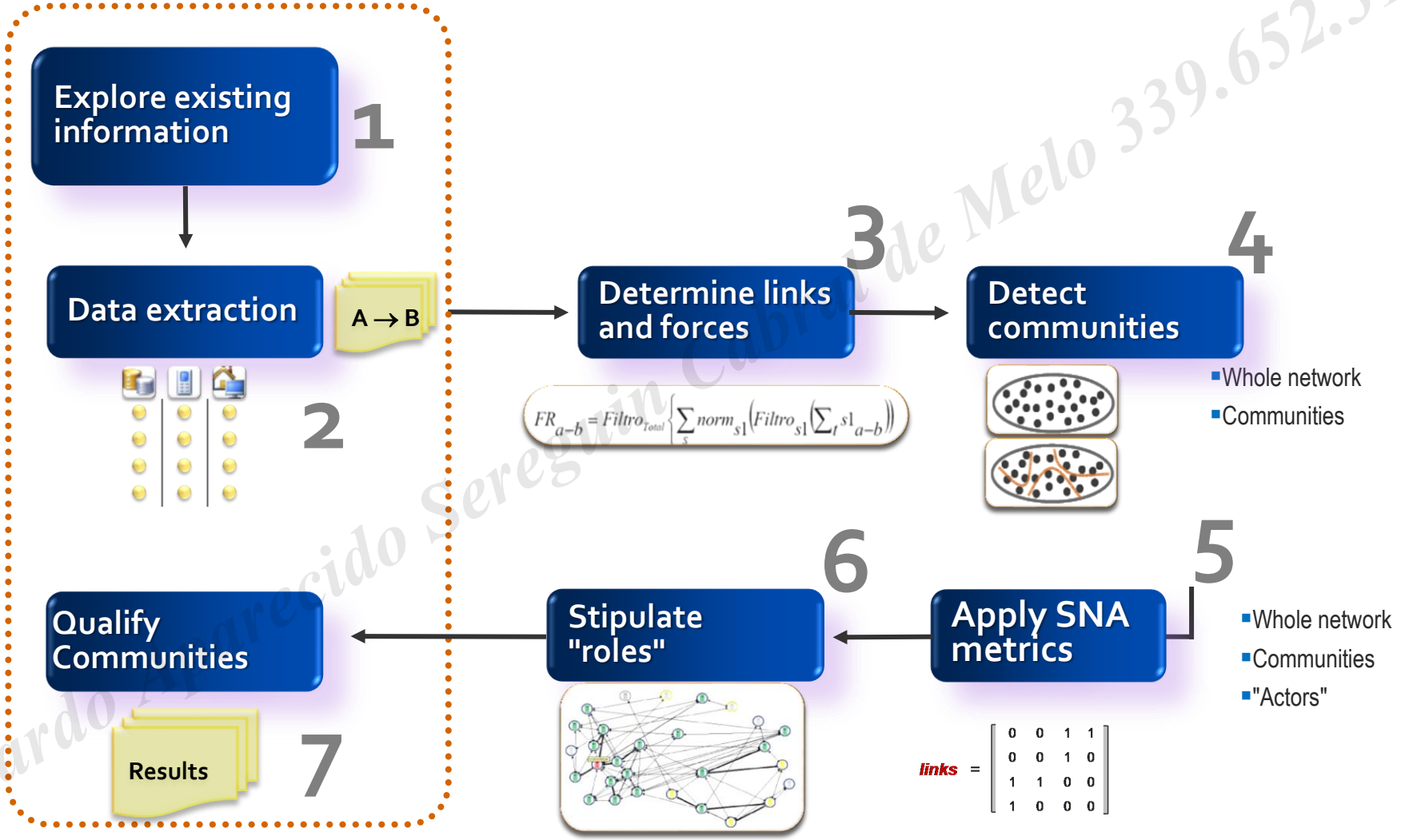
Static View versus Dynamic View

Communities and their members can change over time



- SNA helps directing these changes:
  - Actions in communities change communities
  - Actions in roles change individually roles or events of influence

SNA process (suggestion): "step by step"



Investments

Centrality Metrics

- Centrality
  - Centrality measures provides metrics of importance of a node in a network.
  - Among them, there are:
    - Degree
    - Weighted Degree
    - Clustering Coefficient
    - Closeness (closeness centrality)
    - Betweenness (betweenness centrality)

Centrality Metrics

Degree - Centrality

- *Degree*
  - Measure regarding the amount of *links* that a certain node has (for an *undirected* network)
  - For a *Directed* network, *out-degree* refers to the number of *links* that proceed from a certain node, *in-degree* is the number of links that the same node receives. And *degree* is the sum of *in-degree* with *out-degree* of the node.

Centrality Metrics

Weighted Level - Centrality

- *Weighted Degree*
  - It is a generalization of the degree that takes into account the weight of links (when existing).

Centrality Metrics

Clustering Coefficient - Centrality

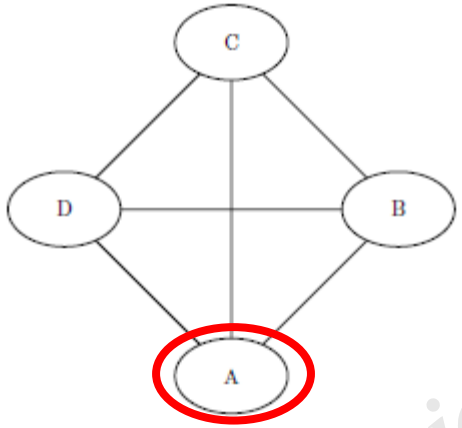
- *Clustering Coefficient*
  - The *clustering coefficient* for a node is the number of links between their neighbor nodes, divided by the number of links that could exist between them.
  - Measure that presents how the neighbors are connected and not the node itself.

$$C(i) = \frac{|\{(u,v) \in A : u,v \ \in \delta_i|}{|\delta_i|(|\delta_i| - 1)}$$

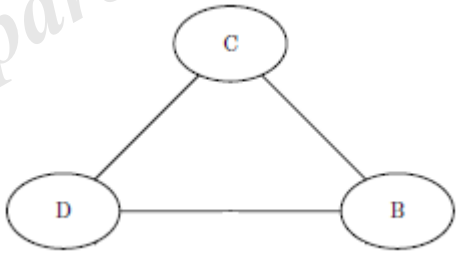
Centrality Metrics

Clustering Coefficient - Centrality

- *Example*



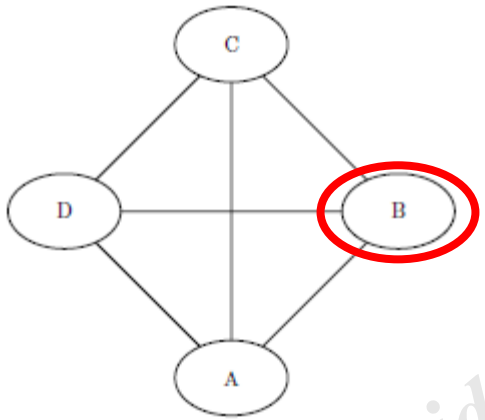
$$C(A) = \frac{\textit{how many links between them}}{\textit{how many links possible}} = \frac{3}{3} = 1$$



Centrality Metrics

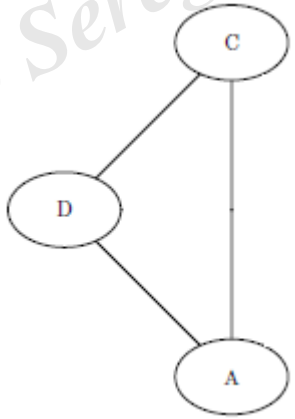
Clustering Coefficient - Centrality

- *Example*



$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{3}{3} = 1$$

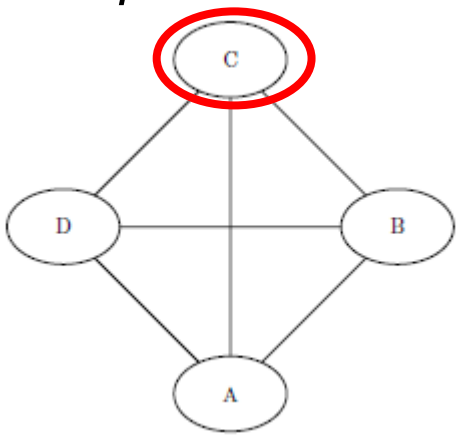
$$C(B) = \frac{3}{3} = 1$$



Centrality Metrics

Clustering Coefficient - Centrality

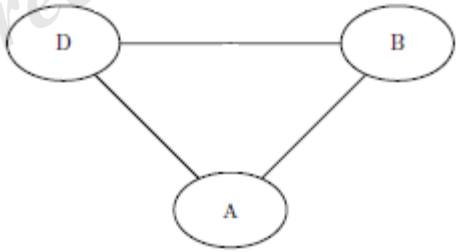
- *Example*



$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{3}{3} = 1$$

$$C(B) = \frac{3}{3} = 1$$

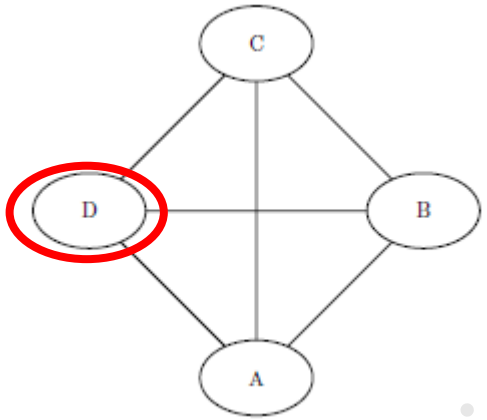
$$C(C) = \frac{3}{3} = 1$$

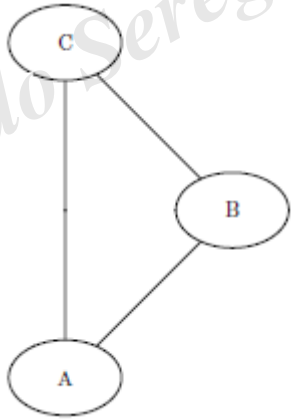


Centrality Metrics

Clustering Coefficient - Centrality

- *Example*





$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{3}{3} = 1$$

$$C(B) = \frac{3}{3} = 1$$

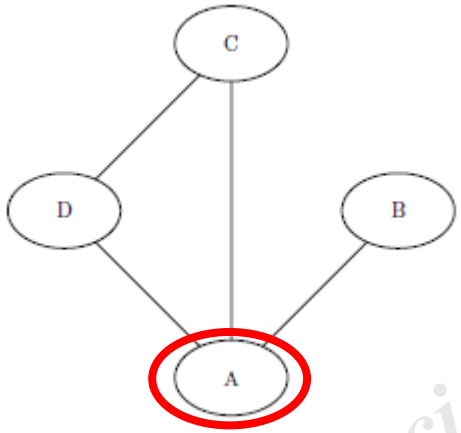
$$C(C) = \frac{3}{3} = 1$$

$$C(D) = \frac{3}{3} = 1$$

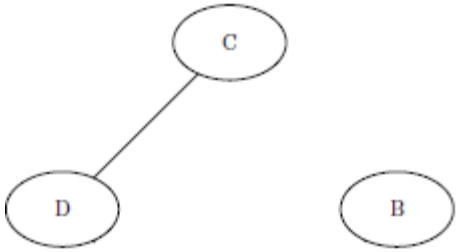
Centrality Metrics

Clustering Coefficient - Centrality

- *Example 2*



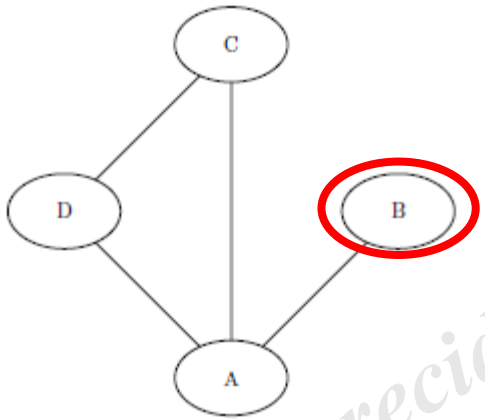
$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{1}{3} = 0.333$$



Centrality Metrics

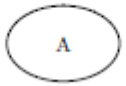
Clustering Coefficient - Centrality

- *Example 2*



$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{1}{3} = 0.333$$

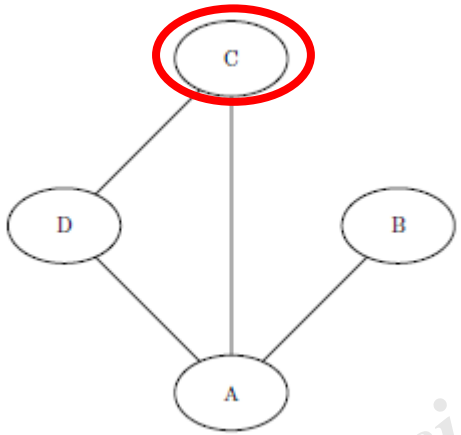
$$C(B) = 0$$



Centrality Metrics

Clustering Coefficient - Centrality

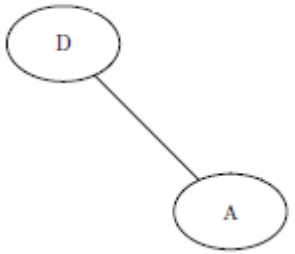
- *Example 2*



$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{1}{3} = 0{,}333$$

$$C(B) = 0$$

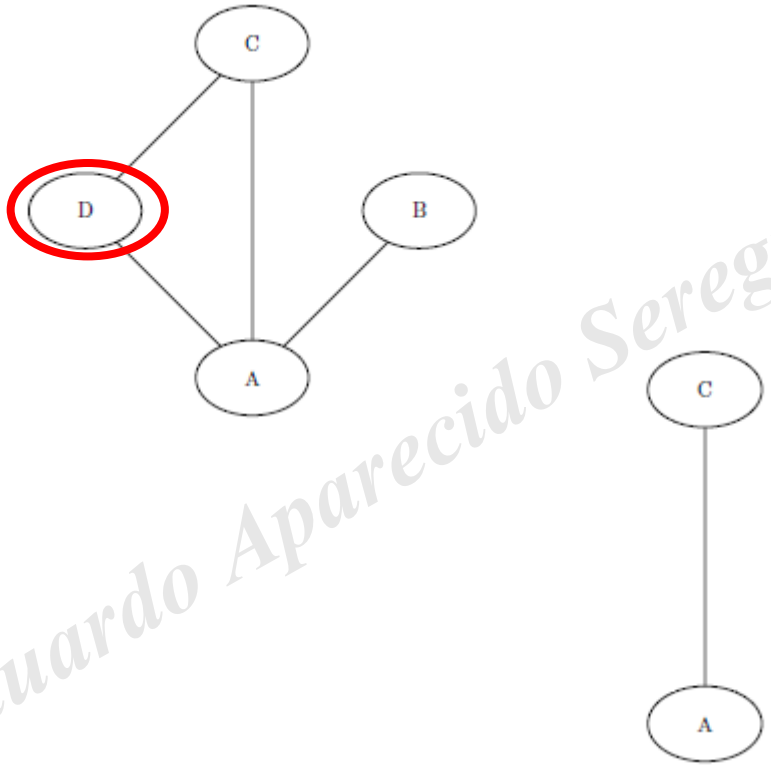
$$C(C) = 1$$



Centrality Metrics

Clustering Coefficient - Centrality

- *Example 2*



$$C(A) = \frac{quantas\ ligações\ entre\ eles}{quantas\ possíveis\ ligações} = \frac{1}{3} = 0{,}333$$

$$C(B) = 0$$

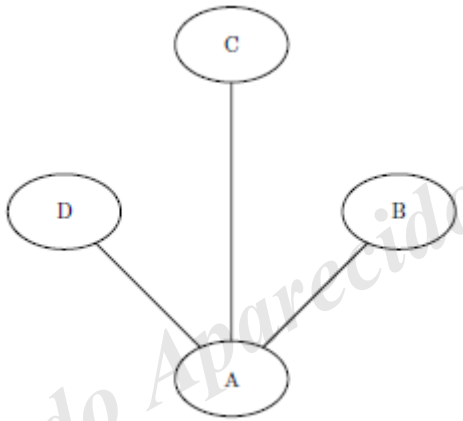
$$C(C) = 1$$

$$C(D) = 1$$

Centrality Metrics

Clustering Coefficient

- *Example 3*



$$C(A) = 0$$
$$C(B) = 0$$
$$C(C) = 0$$
$$C(D) = 0$$

Centrality Metrics

Closeness - Centrality

- *Closeness Centrality*
  - The closeness metric is a proximity measure. It is the average of shortest paths to reach another individual. Proximity can be considered a measure of time that would be necessary to spread information from a node to the other nodes in the network.
  - Mean distance from the referred node to all other nodes.

$$C_C(u) = \frac{|N| - 1}{\left(\sum_{v \in N \setminus u} d_{uv}^N + d_{vu}^N\right) / 2}$$

being $d_{uv}^N = \begin{cases} d_{uv}, se\ d_{uv} < \infty \\ |N|, otherwise \end{cases}$

in which d_{uv} is the short path between the node u and v.

Centrality Metrics

Betweenness - Centrality

- *Betweenness (betweenness centrality)*
  - The Betweenness metric counts the number of times that a private node occurs in the shortest paths among other nodes. Nodes with large betweenness are as the "porters" of information" Since they are in the shortest paths, they are the "conductors" (or "handlers") of information.
  - Betweenness measures the frequency that the node appears in the shortest paths between nodes.

$$C_b(u) = \sum_{\substack{s \neq u \neq t \in N \\ s \neq t}} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

in which σ_{st} is the number of paths

and $\sigma_{st}(u)$ is the number of short paths between s and t passing u through

Network Metrics

General Statistics

- Several statistics can be calculated for networks and for nodes of the networks:
  - **Nodes**: number of nodes in the chart (|N|)
  - **Edges**: number of links in the chart (|A|)
  - **Graph density**: the number of links in a chart (|A|), divided by the total number of links in the complete chart (|N| (|N| - 1))
  - **Average length of the paths**: sum of the shortest paths divided by the number of shortest paths
  - **Eccentricity**: distance from the node until the most distant node of the network (it's the longest way among the shortest paths)
  - **Radius**: lower eccentricity
  - **Network Diameter**: greater eccentricity (greater weight among the shortest paths)

Bibliographic references

- PINHEIRO, C. A. R. **Social Network Analysis in Telecommunications**. USA: John Wiley, 2011, 284 p.

- NEWMAN, M. E. J. **Networks**: An Introduction. New York: Oxford, 2010, 772 p.

MBA
USP
ESALQ

THANK YOU!

Prof. Adriana Silva

linkedin.com/in/adrianamms