Social Network Analysis (SNA)

Jelena Jovanovic

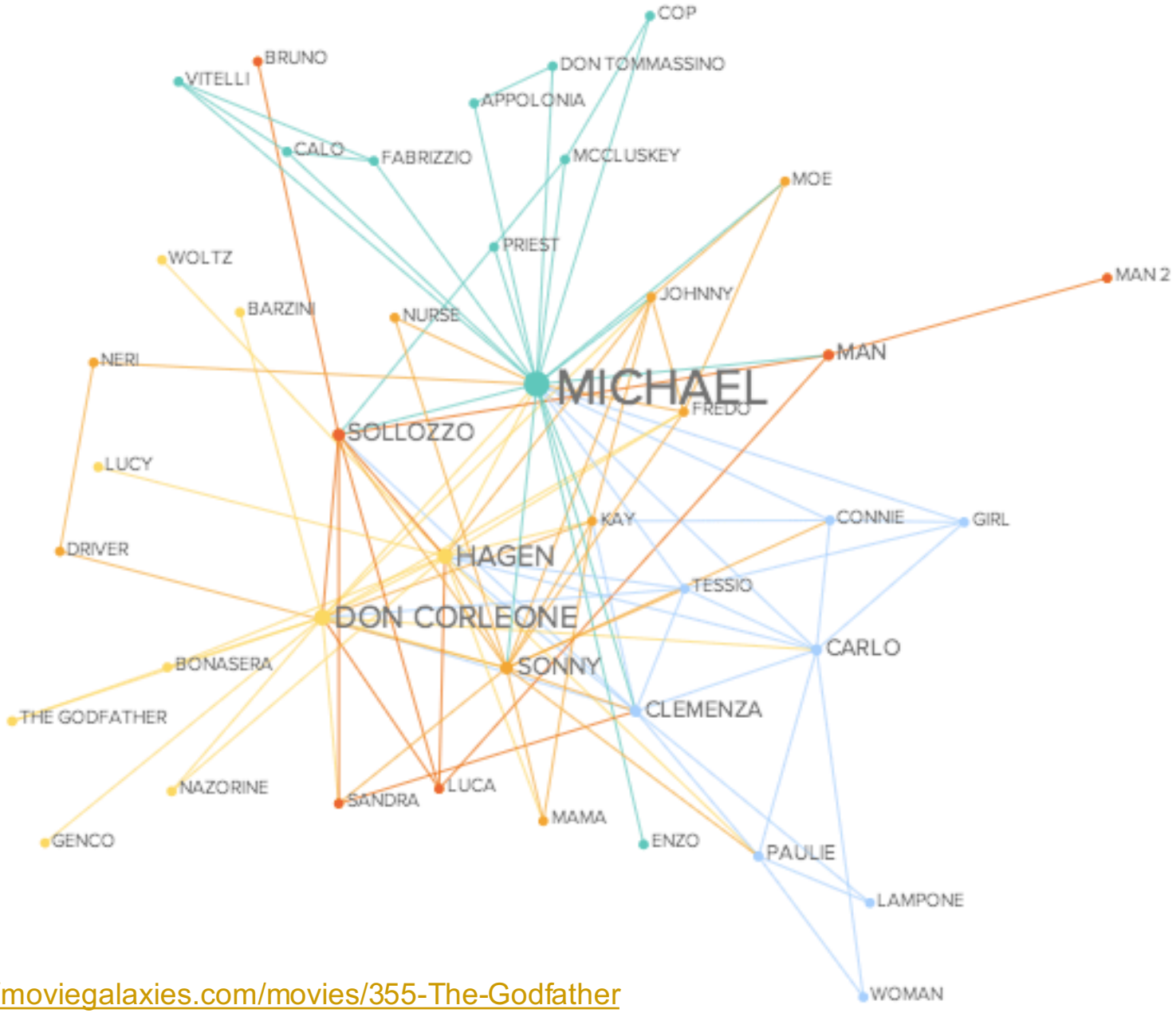
Email: jeljov@gmail.com

Web: http://jelenajovanovic.net

SOCIAL NETWORK ANALYSIS

- Data about members of a social network / group / community and their mutual relationships are presented in the form of a (social) graph
  - each member of the network is represented as a node (vertex) in the graph
  - relationships between the members are edges in the graph
- The analysis is done by applying different measures and algorithms to the social graph

Social network of characters in The Godfather (1972) movie



Source: http://moviegalaxies.com/movies/355-The-Godfather
More about social networks from the MovieGalaxies: http://goo.gl/NTdg1D

MODELING A SOCIAL NETWORK

Edges of a (social) network typically represent:

- Some form of social relation (e.g., friendship, family relations, business connections, …), or
- Some form of social interaction (e.g., exchange of chat messages, email exchange, …), or
- Sharing a common feature (e.g., preference for the same type of movies / books / food; graduation from the same school, …)

MODELING A SOCIAL NETWORK

Edges can have weights that quantify the intensity or strength of the relationship between network actors

Weights can be based on:

- Frequency of interaction (quantified via e.g., the number of messages exchanged in a certain time period)
- Subjective assessment of the strength of connection between actors
- Physical proximity or distance between the actors
- Certain combination of the above mentioned elements
- …

WHAT QUESTIONS CAN BE ANSWERED WITH SNA?

- Who are the central / most influential members of the network?
- Which groups can be detected in the network? To what extent is the network clustered into smaller, well connected groups?
- How is the network evolving? Will it grow or diminish in size and eventually cease to exist?
- How are ideas/knowledge/viruses/… spreading through the network?
- …

SNA APPLICATION EXAMPLE: DETECTION OF THE MOST INFLUENTIAL NETWORK MEMBERS

WHO ARE THE MOST INFLUENTIAL NETWORK MEMBERS?

Influence (and power) of an actor in a network originates in his/her *advantageous network position*

Key indicators of one's advantageous network position are high values of the following SNA metrics:

- *Degree centrality*
- *Betweenness centrality*
- *Closeness centrality*

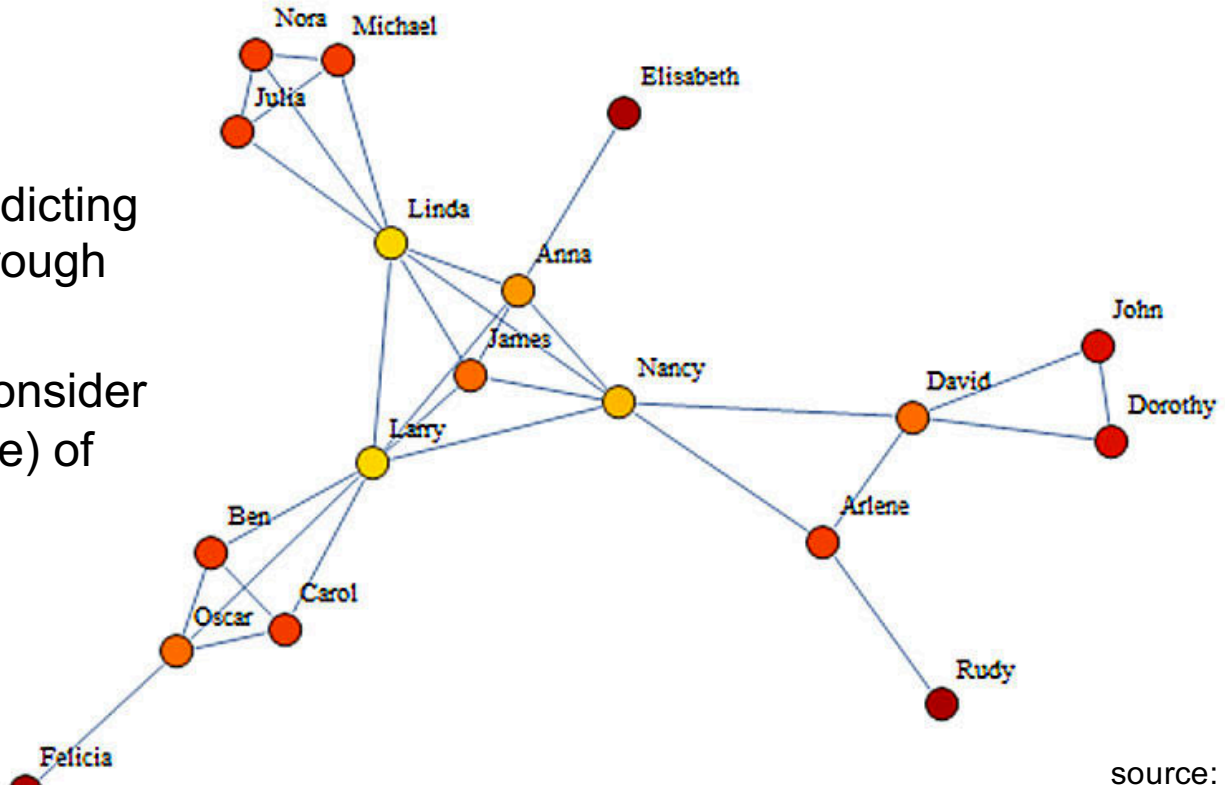
DEGREE CENTRALITY

It is the ratio of the actor's degree (= number of immediate neighbors) and the total number of actors in the network

Example (figure): the lighter the color of a node, the higher its degree centrality is

Useful metric for estimating/predicting the spreading of information through the network

Potential drawback: does not consider the position (influence/relevance) of the neighboring nodes

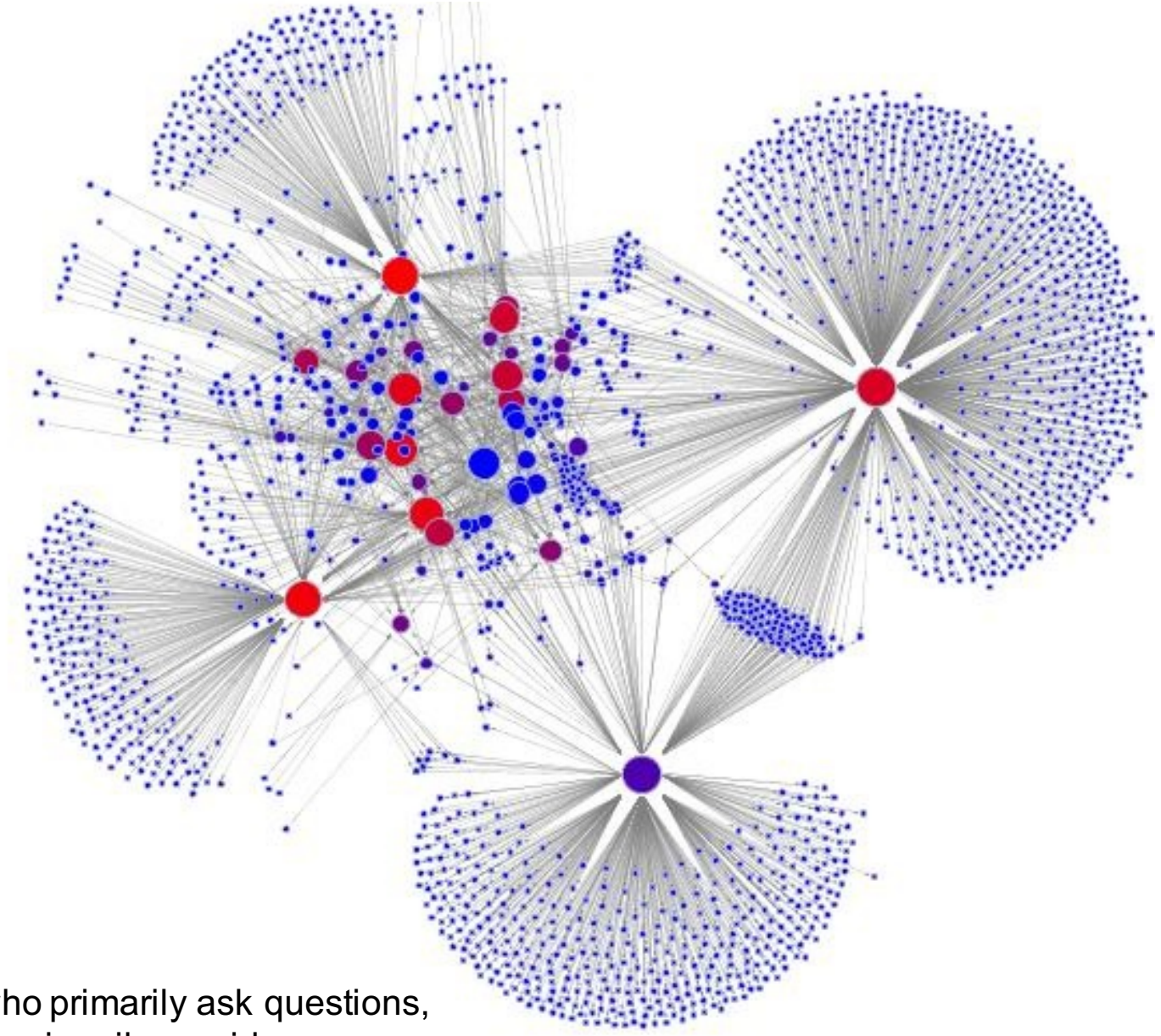


source: http://blog.postbit.com/photos/social-graphs/degree-centrality-in-social-networks.html

Network formed by the members of Java discussion forum

Example of a network where *degree centrality* identifies dominant actors (hubs)

More precisely, since this is a directed network, we distinguish: *indegree* (number c incoming links) and *outdegree* (number of outgoing connections)

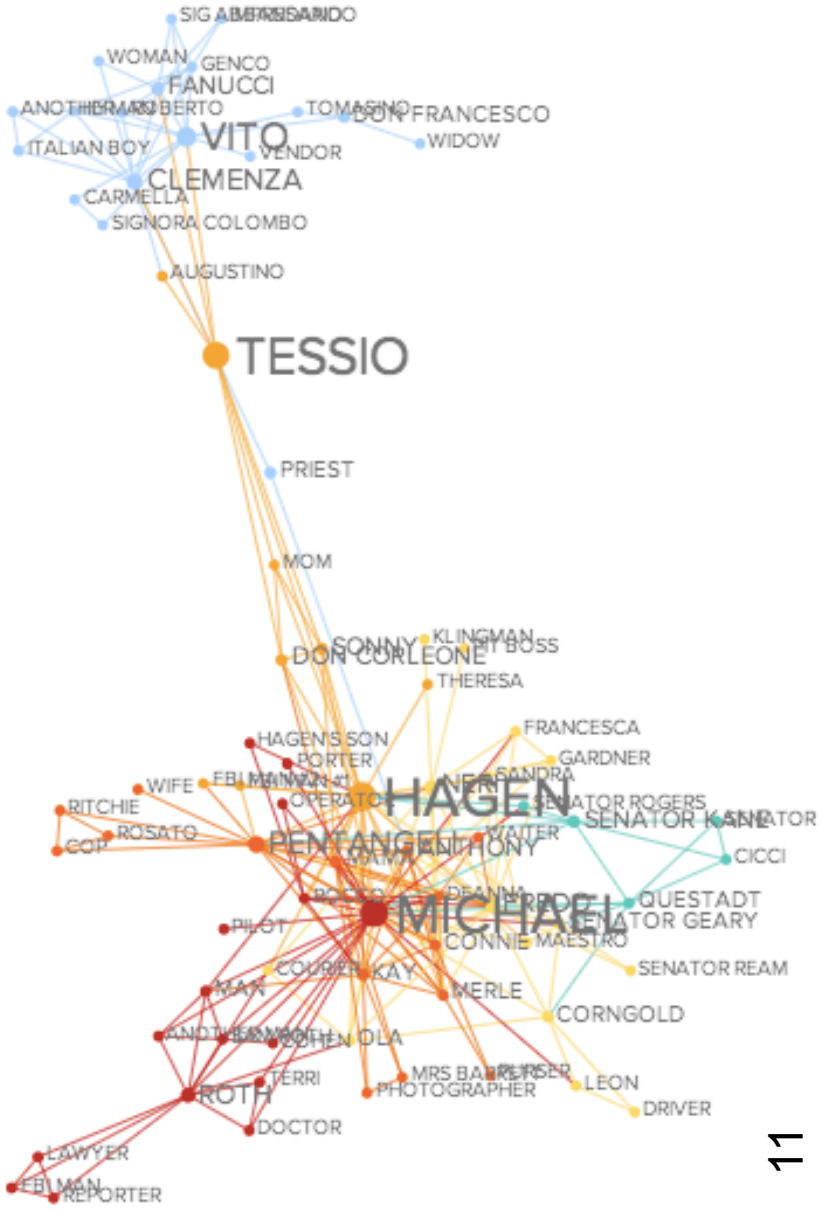


Blue nodes are those who primarily ask questions, while red are those who primarily provide answers

Image source: http://www-personal.umich.edu/~ladamic/img/JavaForumFragmentWhite.jpg

BETWEENNESS CENTRALITY

- Enables the detection of actors who often appear as intermediaries / middlemen connecting other network members
- In addition, it allows for the detection of weak points in the network, where the network might break

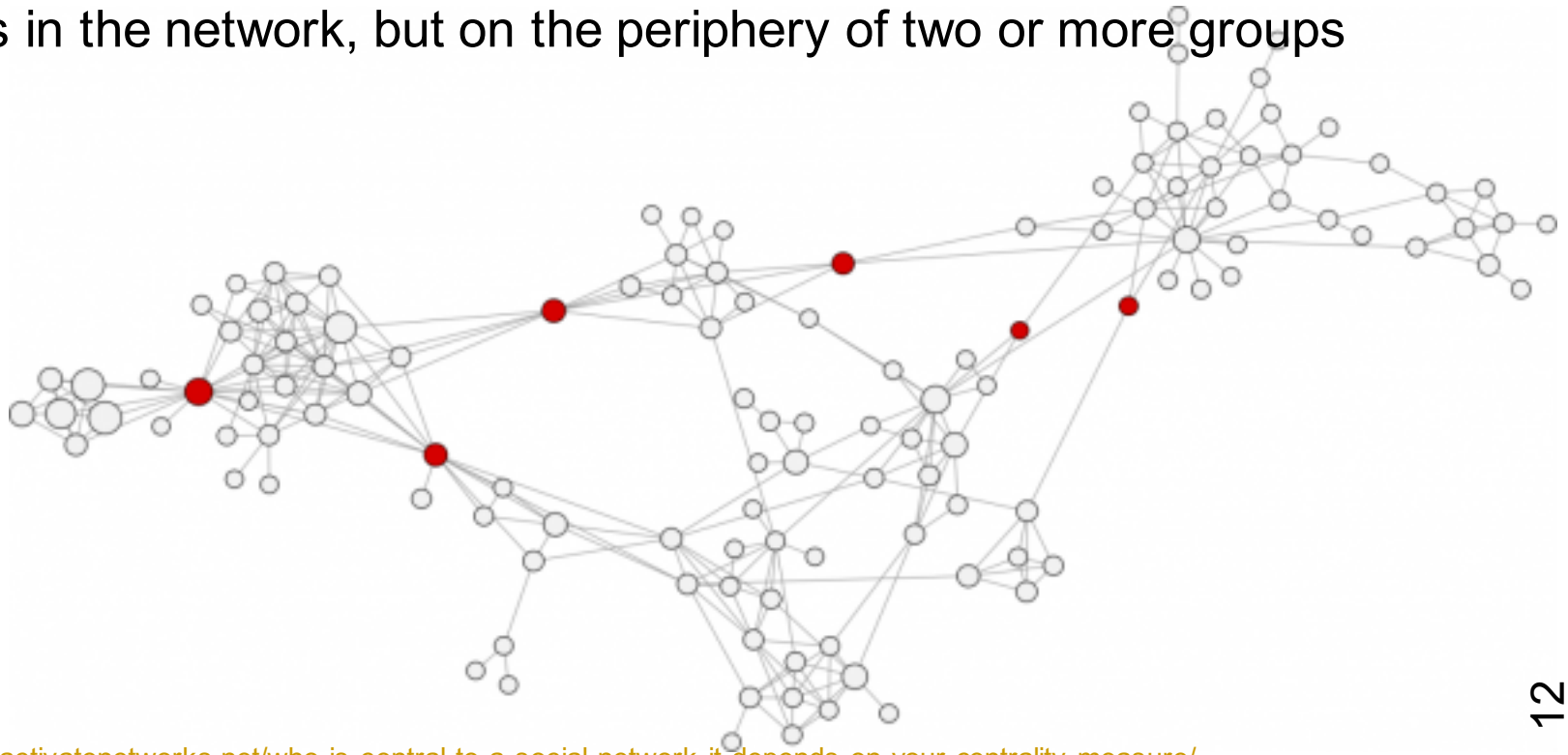


Source: http://moviegalaxies.com/movies/356-The-Godfather:-Part-II

BETWEENNESS CENTRALITY

Even though actors with high betweenness have important role of network *brokers*, typically they are not among the prominent network members (often left unnoticed)

Reason: they are often positioned, not in the center of any of the groups in the network, but on the periphery of two or more groups



source: http://www.activatenetworks.net/who-is-central-to-a-social-network-it-depends-on-your-centrality-measure/

CLOSENESS CENTRALITY

This metric represents average closeness / distance of the given actor from all the other actors in the network

High closeness centrality is typical for actors positioned close to the centers of local clusters (groups) in a larger social network

Source: http://www.activatenetworks.net/who-is-central-to-a-social-network-it-depends-on-your-centrality-measure/

CLOSENESS CENTRALITY

Some characteristics of this metric:

- It indicates how successful an actor can be in spreading information through the network
- Actors with high closeness centrality tend to be
  - Influential members of the local groups they belong to
  - Unknown or weakly known at the level of the network as a whole (in spite of being locally influential)
  - Efficient in spreading information through the part of the network they belong to

Application domains

"TRADITIONAL" APPLICATION DOMAINS

Organizations: improving communication channels within an organization, as well as in the larger network that includes business partners and clients

Marketing: identification of central network actors so that they can be 'employed' to better promote products / services / …

Telecommunications: optimization of the structure and capacity of telecommunication networks

Police and investigation agencies: identification of central members of different criminal / terrorist groups and networks

NEW APPLICATION DOMAINS: ONLINE SOCIAL NETWORKS

Popularity and application of SNA methods and techniques have substantially increased in the last couple of years due to:

- Omnipresence of diverse kinds of online social networks that
  - generate huge quantities of data that can be used for analysis
  - need to understand and manage network features and network dynamics
- Substantial increase in the processing power of computers, enabling today's computers to quickly perform complex and/or computationally demanding SNA algorithms

ONLINE SOCIAL NETWORKS

Depending on the type of online social network, SNA can be applied for:

- Recommendation of potentially interesting / relevant friends / contacts, groups, events
- Detection of influential network members who could help in sharing and/or popularizing certain ideas, news, products, events,…
- Identification of experts in a specific subject area
- …

ENTERPRISE SOCIAL SOFTWARE

- Increasingly present in companies (of various sizes)
  - E.g., Work.com, SocialCast.com, Yammer.com
- Includes diverse kinds of social software
  - Wiki, blog, forum, micro-blog, chat,…
- Acts as a source of huge amounts of data about interactions among the members of an organization, including:
  - direct interaction (e.g., exchange of messages in a discussion forum or chat)
  - indirect interaction (e.g., editing of the same wiki page, commenting the same blog post)
- SNA applied to these data allows for the detection of informal groups, influential individuals and groups, and the like

SOME INTERESTING THINGS TO READ / WATCH

- [news article] Degrees of separation: After all, it is who you know (http://goo.gl/aPWip5)
- "Social Physics: How Good Ideas Spread" by Sandy Pentland, Talks at Googl: https://www.youtube.com/watch?v=HMBl0ttu-Ow
- "The hidden influence of social networks", TED talk by Nicholas Christakis: https://www.youtube.com/watch?v=2U-tOghblfE
- [talk] Duncan Watts on Network Analysis, Small World networks, Big Data and the like: https://www.youtube.com/watch?v=4YoOCLoJn6U