A Connected World

Data Analysis for Real World Network Data

Introduction

08.12.2022

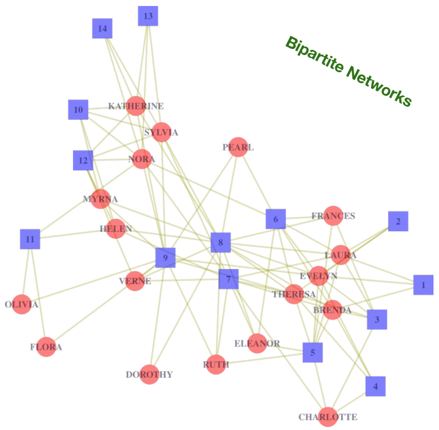
BERD Academy

Networks ≠ Networks

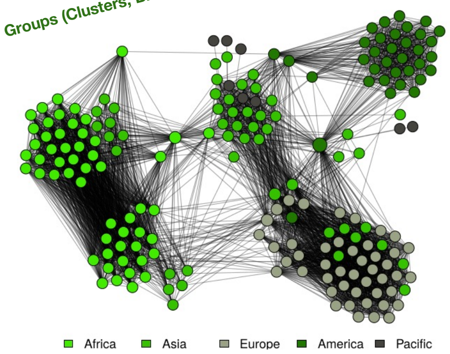
Networks ≠ Networks

Meetings and alliances …

Bipartite Networks



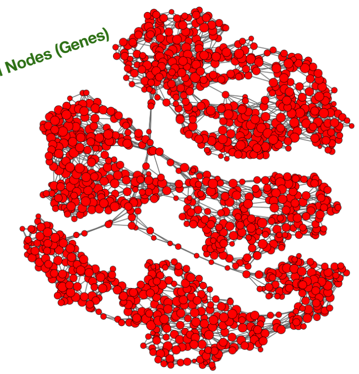
Groups (Clusters, Blocks, Communities)



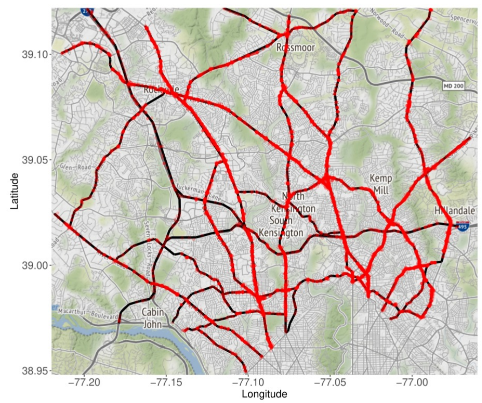
Networks ≠ Networks

… Ig interactions and road grids

Central Nodes (Genes)

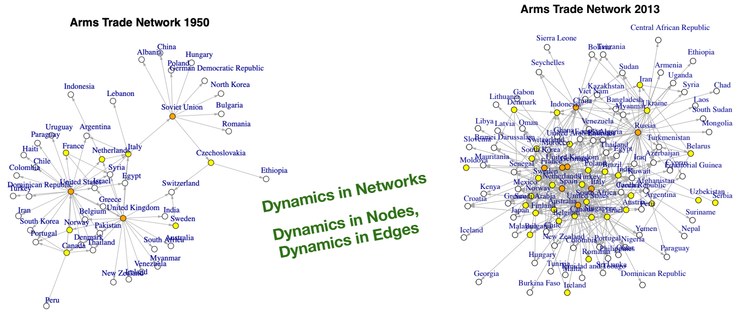


Observations ON Networks



Networks ≠ Networks

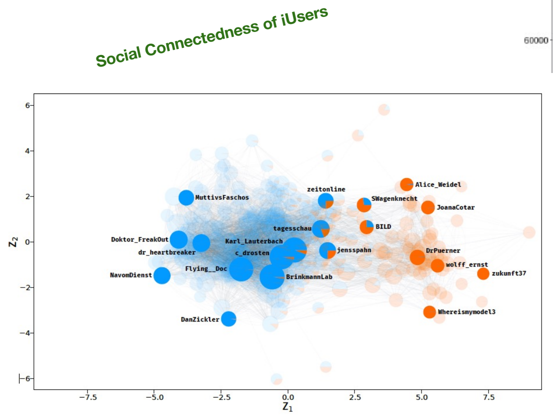
Arms Trading Networks

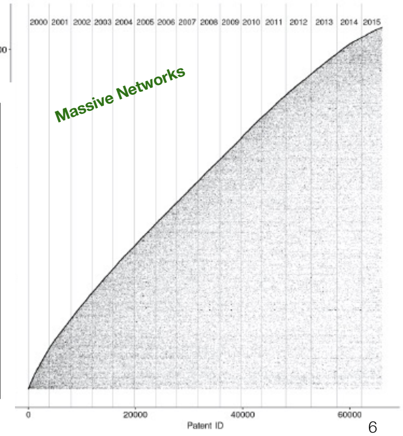


Data Source SIPRI

Networks ≠ Networks

… Twitter Network and patent collaboration





Networks ≠ Networks

Ubiquity of networks?

In recent years there has been an explosion of network data *— that is, measurements that are either of or from a system conceptualized as a network — from seemingly all corners of science.* (Kolaczyk [106])

Empirical studies and theoretical modeling of networks have been the subject of a large body of recent research in statistical physics and applied mathematics. (Newman and Girvan [83])

Networks have in recent years emerged as an invaluable tool for describing and quantifying complex systems in many branches of science. (Clauset, Moore and Newman [38])

Prompted by the increasing interest in networks in many fields […]. (Bickel and Chen [19])

Networks are fast becoming part of the modern statistical landscape. (Wolfe and Olhede [155])

The rapid increase in the availability and importance of network data […]. (Caron and Fox [32])

Network analysis is becoming one of the most active research areas in statistics. (Gao, Lu and Zhou [79])

Networks are ubiquitous in science. (Fienberg [74])

Networks are ubiquitous in science and have become a focal point for discussion in everyday life. (Goldenberg, Zheng, Fienberg, and Airoldi [84])

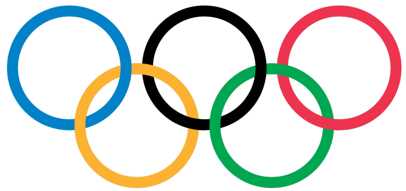
Networks ≠ Graphs

What do those networks have in common?

- Units, actors, agents, or nodes $\Rightarrow \mathcal{N} = \{X_1, \ldots, X_n\}$
- Ties between them $\Rightarrow \mathcal{E} = \{(i,j); i,j \in \mathcal{N}\}$

$$\left.\begin{array}{l} \mathcal{N} \\ \mathcal{E} \end{array}\right\} \mathcal{G} = (\mathcal{N}, \mathcal{E})$$

- Edges are generally directed or undirected (focus now lies on the undirected case)
- Formalization of networks:
  - We use graphs to represent networks as a mathematical object
  - Graphs are a natural way to represent networks graphically



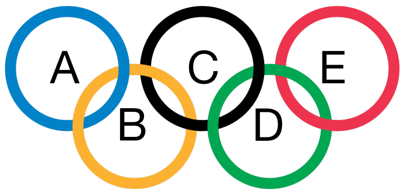
Networks ≠ Graphs

What do those networks have in common?

- Units, actors, agents, or nodes $\Rightarrow \mathcal{N} = \{X_1, \dots, X_n\}$
- Ties between them $\Rightarrow \mathcal{E} = \{(i,j); i,j \in \mathcal{N}\}$

$\Big\}\ \mathcal{G} = (\mathcal{N}, \mathcal{E})$

- Edges are generally directed or undirected (focus now lies on the undirected case)
- Formalization of networks:
  - We use graphs to represent networks as a mathematical object
  - Graphs are a natural way to represent networks graphically



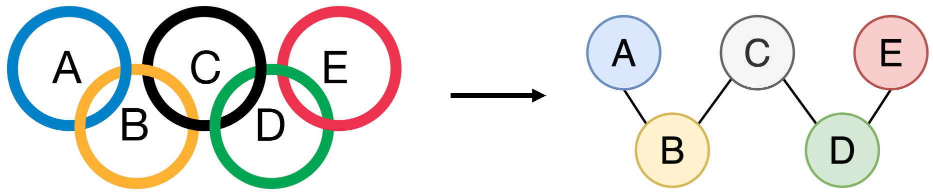
Networks ≠ Graphs

What do those networks have in common?

- Units, actors, agents, or nodes $\Rightarrow \mathcal{N} = \{X_1, \dots, X_n\}$
- Ties between them $\Rightarrow \mathcal{E} = \{(i,j); i,j \in \mathcal{N}\}$

$\Big\}\ \mathcal{G} = (\mathcal{N}, \mathcal{E})$

- Edges are generally directed or undirected (focus now lies on the undirected case)
- Formalization of networks:
  - We use graphs to represent networks as a mathematical object
  - Graphs are a natural way to represent networks graphically



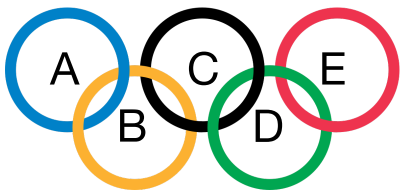
Networks ≠ Graphs

What do those networks have in common?

- Units, actors, agents, or nodes $\Rightarrow \mathcal{N} = \{X_1, \dots, X_n\}$
- Ties between them $\Rightarrow \mathcal{E} = \{(i,j); i,j \in \mathcal{N}\}$

$\Big\}\ \mathcal{G} = (\mathcal{N}, \mathcal{E})$

- Edges are generally directed or undirected (focus now lies on the undirected case)
- Formalization of networks:
  - We use graphs to represent networks as a mathematical object
  - Graphs are a natural way to represent networks graphically

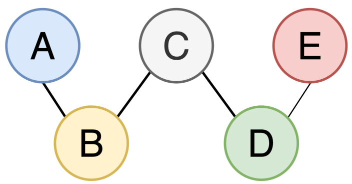


$$\begin{aligned} \mathcal{N} &= \{A, B, C, D, E\} \\ \mathcal{E} &= \{(A,B), (B,C), \\ &\quad (C,D), (D,E)\} \end{aligned}$$

Adjacency Matrix

Alternative representation of networks?

1. Graphs as tuples are not handy
2. Matrices are easier to handle



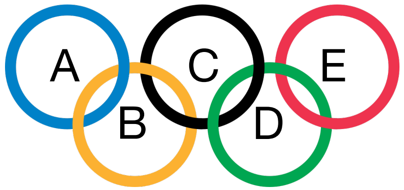
Graph $\mathcal{G}$

$$\begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} - & 1 & 0 & 0 & 0 \\ 1 & - & 1 & 0 & 0 \\ 0 & 1 & - & 1 & 0 \\ 0 & 0 & 1 & - & 1 \\ 0 & 0 & 0 & 1 & - \end{pmatrix} \end{array}$$

Adjacency Matrix y

$$y_{ij} = \begin{cases} 1, \mathit{if}(i,j) \in \mathcal{E} \\ 0, \mathit{else} \end{cases}$$

Networks ⇒ Graphs ⇒ Adjacency Matrix



Network

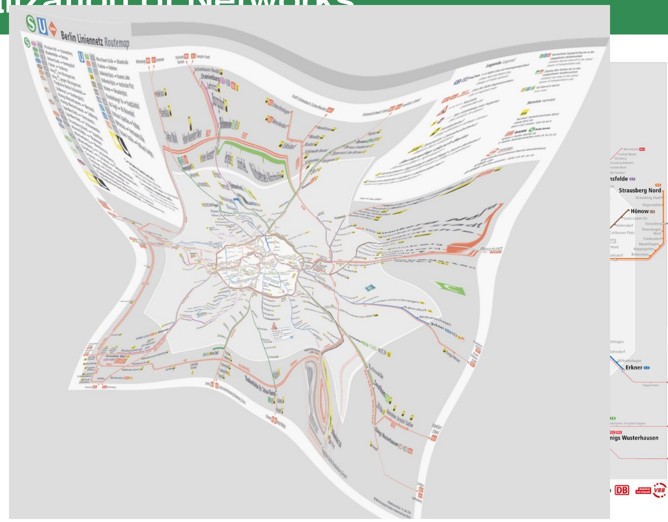
$$\mathcal{N} = \{A, B, C, D, E\}$$
$$\mathcal{E} = \{(A,B), (B,C), (C,D), (D,E)\}$$

Graph $\mathcal{G}$

$$\begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array} \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} - & 1 & 0 & 0 & 0 \\ 1 & - & 1 & 0 & 0 \\ 0 & 1 & - & 1 & 0 \\ 0 & 0 & 1 & - & 1 \\ 0 & 0 & 0 & 1 & - \end{pmatrix} \end{array}$$

Adjacency Matrix y

Visualization of Networks



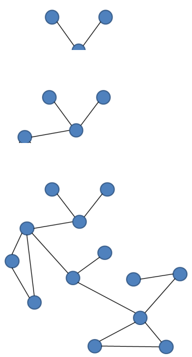
Quelle:
Hans Hack
www.hanshack.com/

Research Questions in/on Networks

Research Questions

Selected Network Topics:

- Systemic Risks in Networks,
  - Spreading Diseases
  - Bank Failure
- Changing Actors in a Network
  - Traffic System
  - International Trade
- Changing Links in a Network,
  - Social Networks
  - Employment Networks
- Growing Network Size
- Network Traffic,
- Etc.



Research Questions

- Reseach Questions are heterogeneous
- We focus in this course on a very few questions, namely
  - ”Mutual dependence of edges”
    - *Is the friend of a friend my friend?*
  - “Latent Space of edges”
    - *Are two friends lying together in a social space?*

Visualzing Networks

Visualizing Networks

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori
Acciaiuoli	0	0	0	0	0	0
Albizzi	0	0	0	0	0	0
Barbadori	0	0	0	0	1	1
Bischeri	0	0	1	0	0	0
Castellani	0	0	1	0	0	0
Ginori	0	0	1	0	0	0

- How can we visualize the network?
- What structural information does this matrix give us?

Visualizing Networks

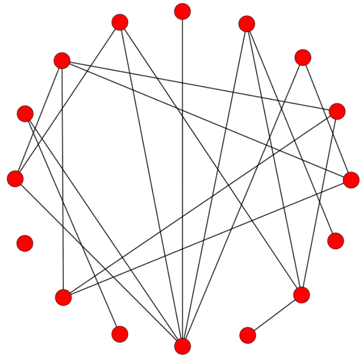
	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori
Acciaiuoli	0	0	0	0	0	0
Albizzi	0	0	0	0	0	0
Barbadori	0	0	0	0	1	1
Bischeri	0	0	1	0	0	0
Castellani	0	0	1	0	0	0
Ginori	0	0	1	0	0	0

- **How can we visualize the network?**
- What structural information does this matrix give us?

Visualizing Networks

How to place nodes and edges in space?

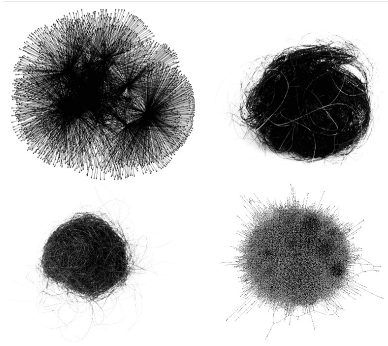
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties



Visualizing Networks

How to place nodes and edges in space?

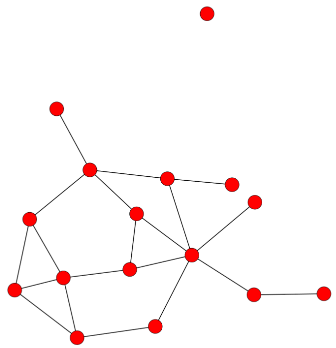
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties



Visualizing Networks

How to place nodes and edges in space?

1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - **Kamada-Kawai** (MDS-based)
   - Fruchterman-Reingold (Analogy to physical systems)



Visualizing Networks

	Munich	Berlin	Frankfurt	Düsseldorf
Munich	0	517	304	486
Berlin	517	0	424	477
Frankfurt	304	424	0	182
Düsseldorf	486	477	182	0

Distance matrix of cities

Haversine formula ←

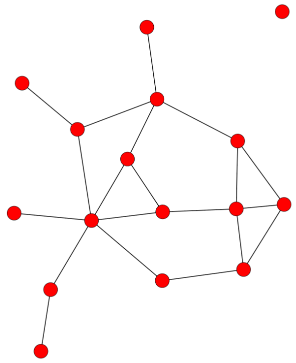
→ MDS

Coordinates

Visualizing Networks

How to place nodes and edges in space?

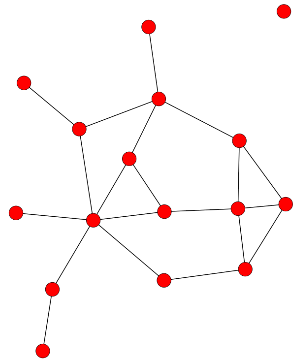
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - **Kamada-Kawai** (MDS-based)
   - Fruchterman-Reingold (Analogy to physical systems)



Visualizing Networks

How to place nodes and edges in space?

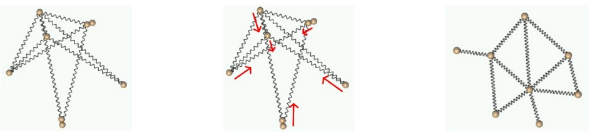
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - Kamada-Kawai (MDS-based)
   - **Fruchterman-Reingold** (Analogy to physical systems)



Visualizing Networks

How to place nodes and edges in space?

1. Step: Cyc
   - Define
   - Drawin
     - Only
   - Aesthe
     - No ir
2. Step: Algorithms to find “good” visualizations
   - Kamada-Kawai (MDS-based)
   - **Fruchterman-Reingold** (Analogy to physical systems)

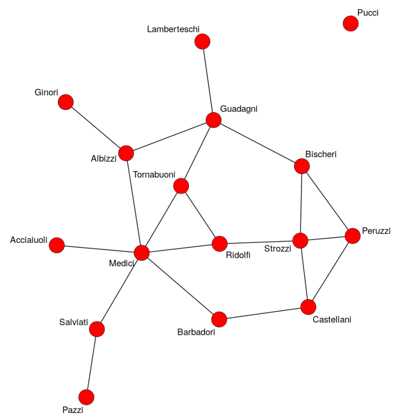


Source: Kobourov, 2012

Visualizing Networks

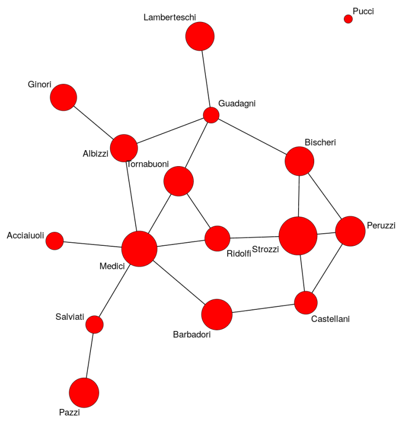
How to place nodes and edges in space?

1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - Kamada-Kawai (MDS-based)
   - Fruchterman-Reingold (Analogy to physical systems)
3. Step: Additional data can be represented
   - Label, size, shape, color of the nodes



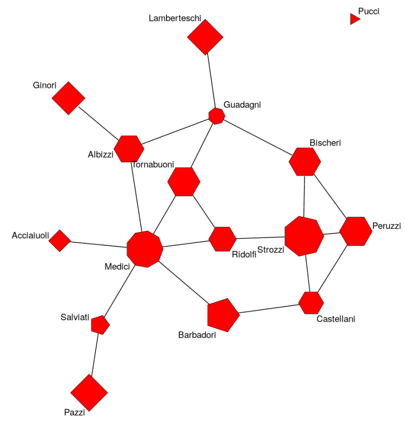
Visualizing Networks

How to place nodes and edges in space?
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - Kamada-Kawai (MDS-based)
   - Fruchterman-Reingold (Analogy to physical systems)
3. Step: Additional data can be represented
   - Label, size, shape, color of the nodes



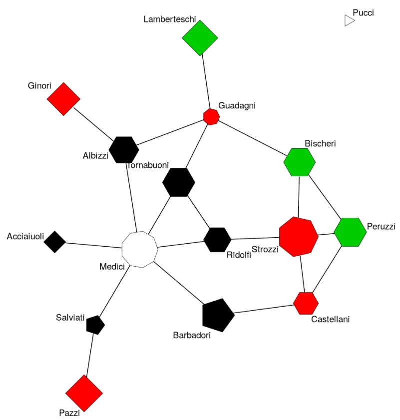
Visualizing Networks

How to place nodes and edges in space?
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - Kamada-Kawai (MDS-based)
   - Fruchterman-Reingold (Analogy to physical systems)
3. Step: Additional data can be represented
   - Label, size, shape, color of the nodes



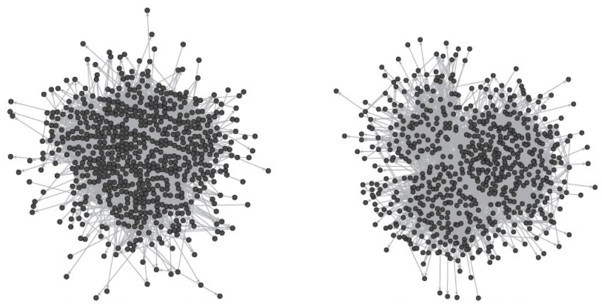
Visualizing Networks

How to place nodes and edges in space?
1. Step: Cycle Layout (“Hair-Ball Effect”)
   - Define requirements for “nice graphs”
   - Drawing Conventions: Hard Constraints
     - Only use Straight line
   - Aesthetics: Soft Constraints
     - No intersecting ties
2. Step: Algorithms to find “good” visualizations
   - Kamada-Kawai (MDS-based)
   - Fruchterman-Reingold (Analogy to physical systems)
3. Step: Additional data can be represented
   - Label, size, shape, color of the nodes



Limitations in Visualizing Networks

- Large Networks are difficult to draw
- Dependent of the Algorithm we “see” the one ore other thing

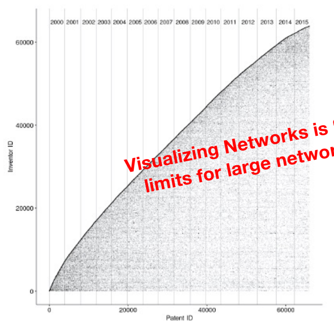


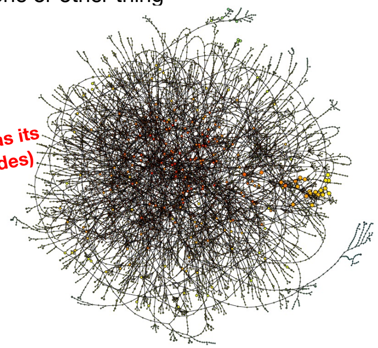
(a) standard spring embedder (b) stress minimization

Source: Network Science, Brandes & Sedlmair. Springer Verlag

Limitations in Visualizing Networks

- Large Networks are difficult to draw
- Dependent of the Algorithm we “see” the one or other thing





Visualizing Networks is “art” and has its limits for large networks (> 50 nodes)

Describing Networks

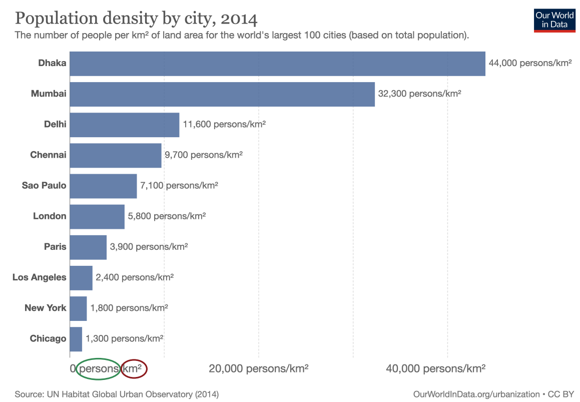
Descriptives of Networks: Global

	Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani	Ginori
Acciaiuoli	0	0	0	0	0	0
Albizzi	0	0	0	0	0	0
Barbadori	0	0	0	0	1	1
Bischeri	0	0	1	0	0	0
Castellani	0	0	1	0	0	0
Ginori	0	0	1	0	0	0

- How can we visualize the network?
- **What structural information does this matrix give us?**

Descriptives of Networks: Global

How to compare more or less dense cities?



How many?
Per space?

Descriptives of Networks: Global

How can we translate this to networks?

How many? ⇒ How many ties are observed? ⇒ Number of edges: $\sum_{i<j} y_{ij}$

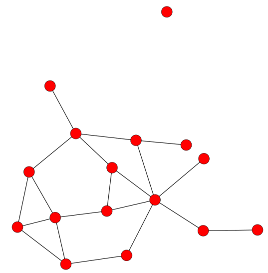
Per space? ⇒ How many ties are possible? ⇒ Possible edges: $\binom{n}{2} = \frac{n(n-1)}{2}$

Density of graph $\mathcal{G}$

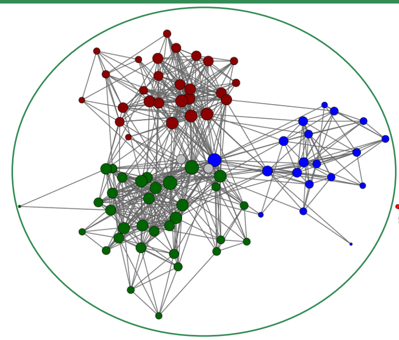
The density of a undirected graph is the frequency of realized edges relative to potential edges

$$den(\mathcal{G}) = \frac{\sum_{i<j} y_{ij}}{\binom{n}{2}}$$

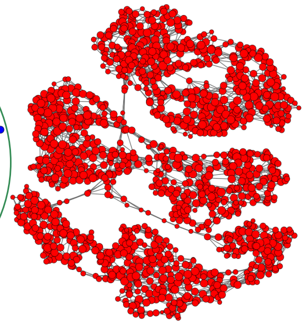
Descriptives of Networks: Global



Marriage (0.166)



Friendships (0.178)

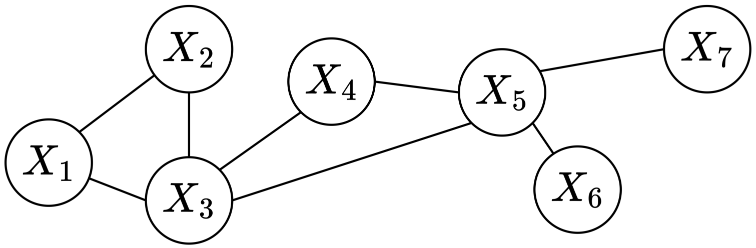


Ig Interaction (0.007)

Which network is most dense?

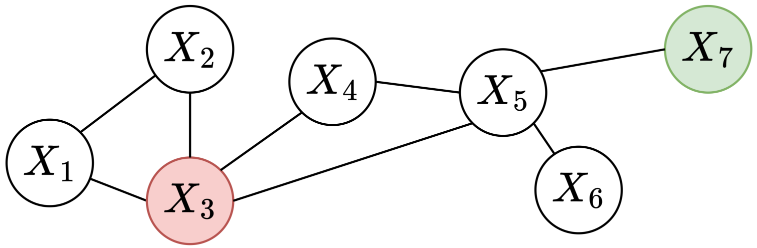
Descriptives of Networks: Degree

How can we describe the local position of a node? (more vs. less central)



Descriptives of Networks: Degree

How can we describe the local position of a node? (more vs. less central)



Descriptives of Networks: Degree

How can we translate this to a measure?

How many? ⇒ How many ties are observed? ⇒ Number of edges: $\sum_{j \neq i} y_{ij}$

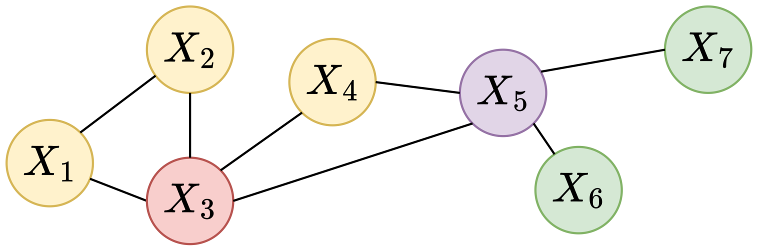
Degree of node X_i in graph $\mathcal{G}$

The degree of node X_i in an undirected graph is the frequency of realized edges

$$deg(X_i) = \sum_{j \neq i} y_{ij}$$

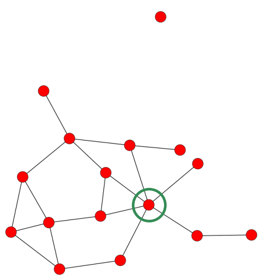
Descriptives of Networks: Degree

How can we describe the local position of a node? (more vs. less central)

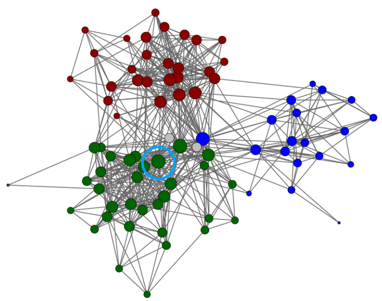


$$deg(X_i) = \sum_{j \neq i} y_{ij}$$

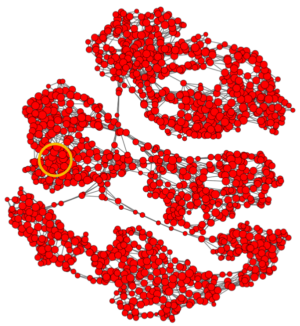
Descriptives of Networks: Degree



Marriage



Friendships



Ig Interaction

Which actors of each networks are the most “central”?

Descriptives of Networks: Centrality

How can we translate this to a global measure of “centralization”?

How much vary the degrees around the maximal degree?

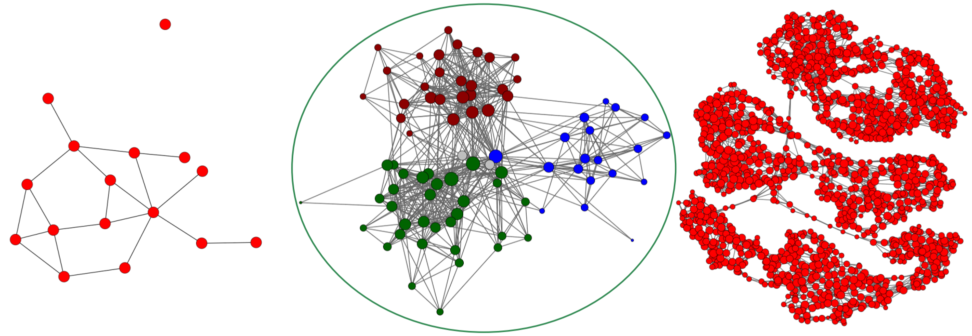
Degree centrality of graph $\mathcal{G}$

The degree centrality of an undirected graph $\mathcal{G}$ is given by:

$$deg(\mathcal{G}) = \frac{\sum_{i=1}^{n} |maxdeg(\mathcal{G}) - deg_{\mathcal{G}}(X_i)|}{(n-1)(n-2)},$$

where the denominator relates to the maximal absolute deviation of the nominator possible.

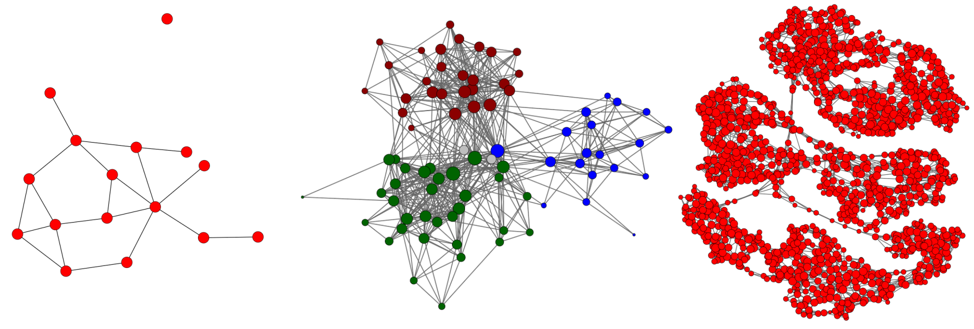
Descriptives of Networks: Centrality



Marriage (0.233) Friendships (0.334) Ig Interaction) (0.006)

Which network is the most “centralized”?

Descriptives of Networks: Centrality



Marriage (0.233) Friendships (0.334) Ig Interaction) (0.006)

There are **HUNDREDS** of different centrality measures
There are HUNDREDS of further measures

Modelling Networks

Modelling Networks

- We denote with $Y \in \{0,1\}^{n \times n}$ the network, represented as adjacency matrix.
- We assume that Y is a **matrix valued random variable** with model $P(Y = y)$
- Sample Size
  - N = 1 ⟹ one network, no repetition
  - N = n ⟹ each of the n actors contributes
  - N = n * (n -1) /2 ⟹ each edge contributes
- This is a mathematical as well as a conceptual question

Modelling Networks

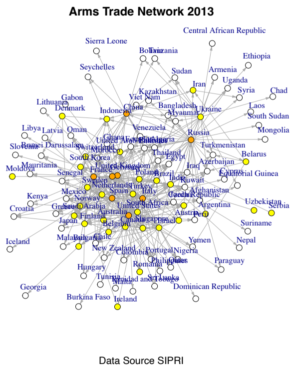
N = 1

The observed network is given.

Why should we apply probability models in the form $P(Y=y)$ if y is NOT a realization of a random process

The approach $P(Y=y)$ allows to uncover the "driving forces" or edges

In what way is the network 2013 different to other years ?



Modelling Networks

N = n

Actor/node focus

What drives an actor to build links/edges to other actors

Who is the global player, and why?

Modelling Networks

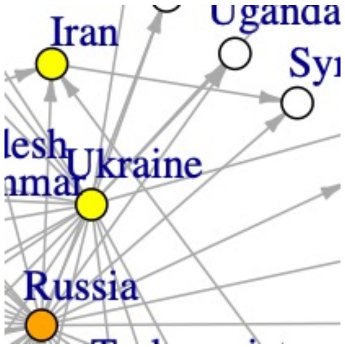
N = n (n -1)/2

Edge focus

What influences the existence of a link

Mutual dependence of links

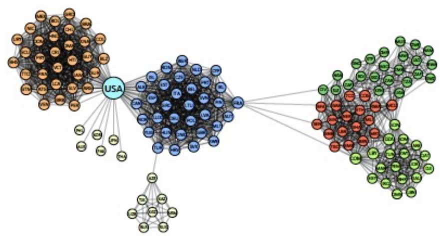
Is the friend of a friend a friend?



Modelling Networks

Latent Space

- Nodes connect due to their local distance
- Nodes connect due to some latent model



Modelling Networks

- The probability model is useful, even though networks do not follow the classical statistical paradigm of i.i.d. data.
- The probability model itself is conceptually very questionable.
- There is no realistic asymptotics in network data analysis.
- Confidence intervals as well as variance estimates are not rigorously justifiable.
- Still: The models are very useful.

Outline of Course

- Models based on Edge Behavior (Exponential Random Graph Models)

  Cornelius Fritz

- Models based on Latent Spaces

  Giacomo de Nicola