Graph mining - lesson 1
Introduction to graphs and networks

Nathalie Vialaneix

nathalie.vialaneix@inrae.fr
http://www.nathalievialaneix.eu
INRAE

M2 Statistics & Econometrics
January 8th, 2020

A brief overview for this class...

Who am I? Statistician working in biostatistics at INRAE Toulouse
My research interests are: data mining, network inference and mining, machine learning

Purpose of this talk: presenting a few statistical tools for graph mining (graph structure, important vertices) and clustering

Outline

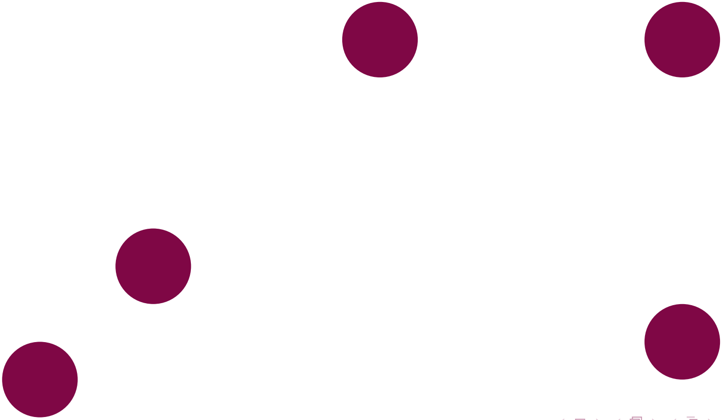
What is a network/graph?

Mathematical object used to model relational data between entities.

What is a network/graph?

Mathematical object used to model relational data between entities.

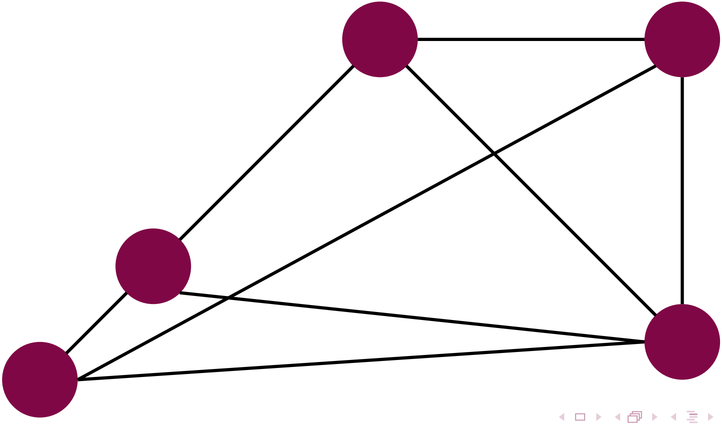
The entities are called the nodes or the vertices



What is a network/graph?

Mathematical object used to model relational data between entities.

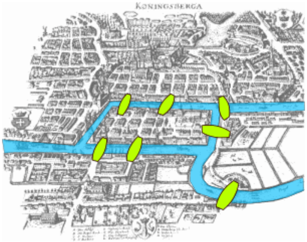
A relation between two entities is modeled by an edge



Where does graph theory come from?

Seven Bridges of Königsberg: notable problem in mathematics. Königsberg set on both sides of the Pregel River and included two large islands.

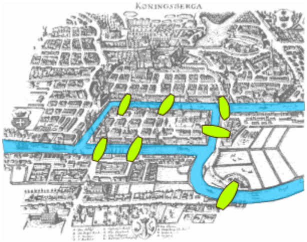
Question: Is there a walk through the city that crosses each bridge once and only once?



Where does graph theory come from?

Seven Bridges of Königsberg: notable problem in mathematics. Königsberg set on both sides of the Pregel River and included two large islands.

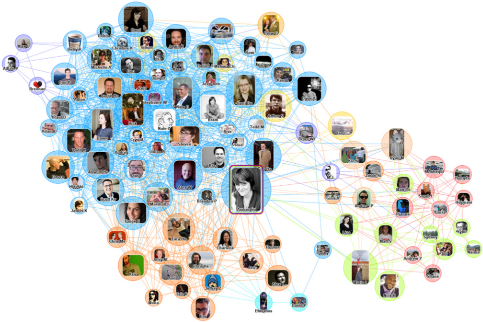
Question: Is there a walk through the city that crosses each bridge once and only once?



Leonhard Euler proved that the problem has no solution using a mathematical proof which was the starting point of graph theory.

Examples of networks

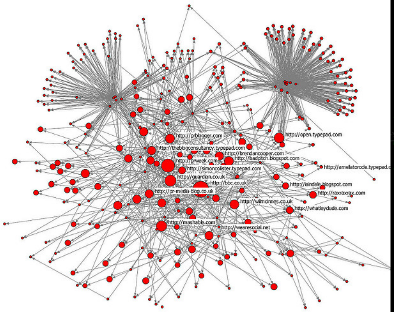
Social networks:

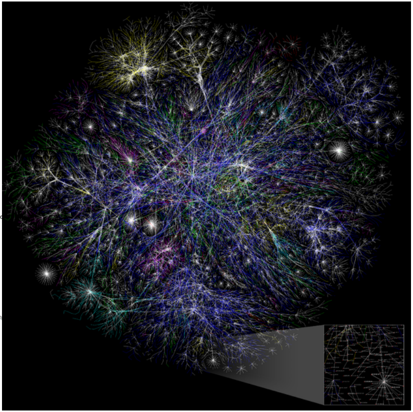


Credits: Frauhoelle (CC BY-SA 2.0) and Caseorganic (CC BY-NC 2.0) on flickr

Examples of networks

Blog co-citations and internet routes:

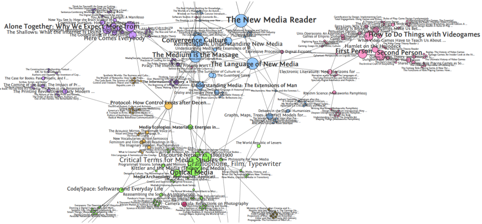




Credits: Porternovelli on flikr (CC BY-SA 2.0) and Matt Britt on wikimedia commons (CC BY 2.5)

Examples of networks

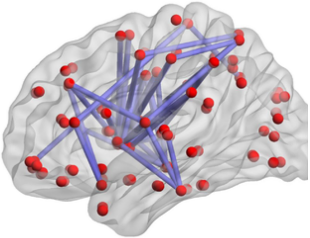
Consumers/products graphs or co-purchase networks:

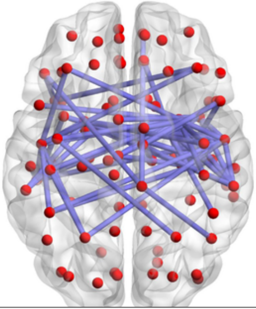


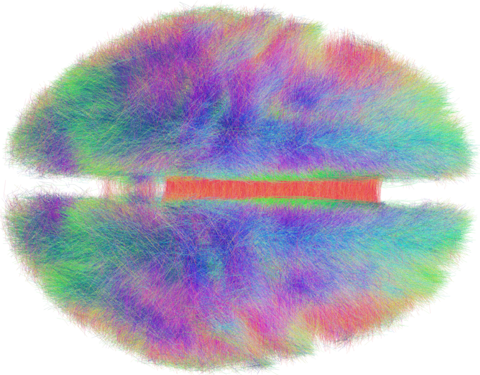
Credits: Loop© and http://www.annehelmond.nl

Examples of networks

(biological) Neural networks:

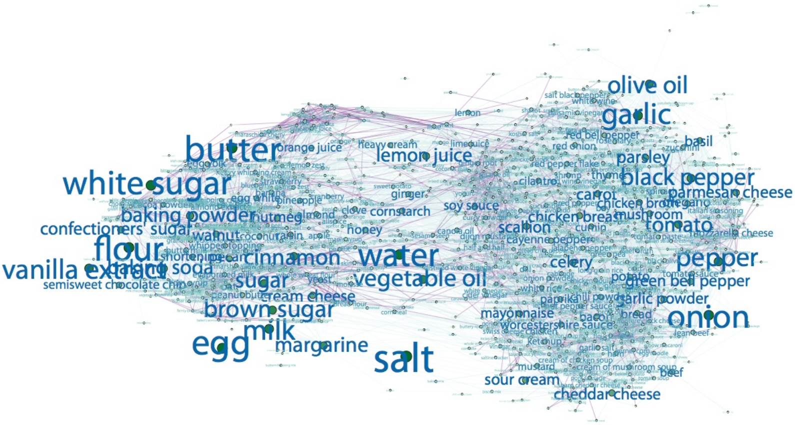






Examples of networks

Ingredient networks... and many others!



Credits: http://www.ladamic.com

More complex relational models

Vertices...

can be labelled with a factor or a numeric variables or several variables (caracteristics attached to the entities in relation)

More complex relational models

Vertices...

can be labelled with a factor or a numeric variables or several variables (caracteristics attached to the entities in relation)

Edges...

- can be oriented
- can be weighted
- can be described by numerical attributes or factors (caracteristics attached to the relation)

Many applications...

- Viral marketing: find a way to efficiently spread the information about a new product using social network informations
- Recommandation systems: recommand a product to someone based on his/her previous purchase and co-purchase information
- Biological network: acquire knowledge about biological networks (genes, metabolomic pathway...) in order to understand diseases associated with disfunctionning
- ...

Standard issues associated with networks

Inference

Giving data, how to build a graph whose edges represent the direct links between variables?

Example: co-expression networks built from microarray data (vertices = genes; edges = significant “direct links” between expressions of two genes)

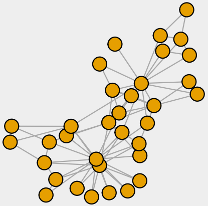
Standard issues associated with networks

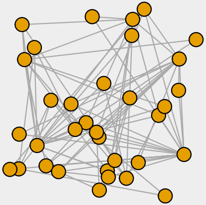
Inference

Giving data, how to build a graph whose edges represent the direct links between variables?

Graph mining (examples)

1. Network visualization: vertices are not a priori associated to a given position. How to represent the network in a meaningful way?





Random positions or positions aiming at representing connected vertices closer.

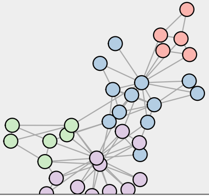
Standard issues associated with networks

Inference

Giving data, how to build a graph whose edges represent the direct links between variables?

Graph mining (examples)

1. Network visualization: vertices are not a priori associated to a given position. How to represent the network in a meaningful way?
2. Network clustering: identify "communities" (groups of vertices that are densely connected and share a few links with the other groups)



Notations for this class

Notations

In the following, a graph $\mathcal{G} = (V, E, W)$ with:

- V: set of vertices $\{x_1, \ldots, x_n\}$;
- E: set of (undirected) edges. $m = |E|$;
- W: weights on edges s.t. $W_{ij} \geq 0$, $W_{ij} = W_{ji}$ and $W_{ii} = 0$.

Notations for this class

Notations

In the following, a graph $\mathcal{G} = (V, E, W)$ with:

- V: set of vertices $\{x_1, \ldots, x_n\}$;
- E: set of (undirected) edges. $m = |E|$;
- W: weights on edges s.t. $W_{ij} \geq 0$, $W_{ij} = W_{ji}$ and $W_{ii} = 0$.

If needed, attributes for the vertices will be denoted by $f_j(x_i)$ (jth attribute for vertex i) and attributes for the edges (other than the weights) by $g_j(x_i, x_{i'})$ (jth attribute for the edge $(x_i, x_{i'})$).

Online graph datasets and ressources

- Mark Newman's collection:
  http://www-personal.umich.edu/~mejn/netdata
- Stanford Large Network Dataset Collection (SNAP):
  http://snap.stanford.edu/data
- KONECT collection (Koblenz university):
  http://konect.uni-koblenz.de/networks
- Colorado Index of Complex Networks (ICON):
  https://icon.colorado.edu

Online course: http://barabasi.com/networksciencebook
(Alberto Barabasi)

Mining graphs/networks

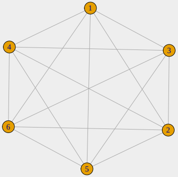
- Visualizing and manipulating graphs in an interactive way: Gephi https://gephi.org, Tulip http://tulip.labri.fr or Cytoscape http://cytoscape.org;

- Packages/librairies in data mining languages:
  - for Python: igraph, NetworkX and graph-tool
  - for R: **igraph**, **statnet**, **bipartite** and **tnet**. See also the CRAN task view: https://cran.r-project.org/web/views/gR.html (graphical models)

Special graphs

Full graphs

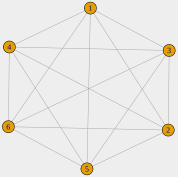
A full graph with vertices $V = \{x_1, \ldots, x_n\}$ is the graph with edge list $E = \{(x_i, x_j) : \ x_i, \ x_j \in V\}$



Special graphs

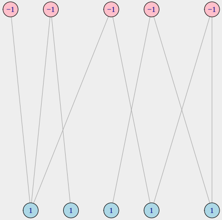
Full graphs

A full graph with vertices $V = \{x_1, \ldots, x_n\}$ is the graph with edge list $E = \{(x_i, x_j) : \; x_i, \, x_j \in V\}$



Bipartite graphs

A graph with vertices $V = \{x_1, \ldots, x_n\}$ partitionned into two groups $\{x_i : \; f(x_i) = 1\}$ and $\{x_i : \; f(x_i) = -1\}$ and such that edges are a subset of $\{(x_i, x_j) : f(x_i) = 1 \text{ and } f(x_j) = -1\}$ (*e.g.*, purchase network)



Most standard ways to record a graph

- adjacency matrix: matrix W if the network is weighted or $A_{ij} = \begin{cases} 1 & \text{if } (x_i, x_j) \in E \\ 0 & \text{otherwise} \end{cases}$ if it is unweighted.
  requires to store n^2 values

Most standard ways to record a graph

- adjacency matrix: matrix W if the network is weighted or $A_{ij} = \begin{cases} 1 & \text{if } (x_i, x_j) \in E \\ 0 & \text{otherwise} \end{cases}$ if it is unweighted.
  requires to store n^2 values

- edge list: matrix B of dimension $m \times 2$ (unweighted network) or $m \times 3$ (weighted network), $B_{k.} = (x_i, x_j, W_{ij})$ for a $(x_i, x_j) \in E$.
  requires to store $3m$ values

Most standard ways to record a graph

- adjacency matrix: matrix W if the network is weighted or $A_{ij} = \begin{cases} 1 & \text{if } (x_i, x_j) \in E \\ 0 & \text{otherwise} \end{cases}$ if it is unweighted.
  requires to store n^2 values

- edge list: matrix B of dimension $m \times 2$ (unweighted network) or $m \times 3$ (weighted network), $B_{k.} = (x_i, x_j, W_{ij})$ for a $(x_i, x_j) \in E$.
  requires to store $3m$ values

Since usually $m \ll n^2$, the second solution is often prefered.

Most standard ways to record a graph

- adjacency matrix: matrix W if the network is weighted or $A_{ij} = \begin{cases} 1 & \text{if } (x_i, x_j) \in E \\ 0 & \text{otherwise} \end{cases}$ if it is unweighted.
  requires to store n^2 values

- edge list: matrix B of dimension $m \times 2$ (unweighted network) or $m \times 3$ (weighted network), $B_{k.} = (x_i, x_j, W_{ij})$ for a $(x_i, x_j) \in E$.
  requires to store $3m$ values

Since usually $m \ll n^2$, the second solution is often prefered.

Other standard formats (readable by interactive software and allowing metadata) such as graphml (a graph version of XML)
http://graphml.graphdrawing.org

Running example 1 “GOT”

“Game of Thrones” coappearances network: weighted and undirected network with 107 vertices corresponding to unique characters and 353 edges weighted by the number of times the two characters’ names appeared within 15 words of each other in the Game of Thrones series by George R.R. Martin.

Running example 1 “GOT”

“Game of Thrones” coappearances network: weighted and undirected network with 107 vertices corresponding to unique characters and 353 edges weighted by the number of times the two characters’ names appeared within 15 words of each other in the Game of Thrones series by George R.R. Martin.

Reference: **[Beveridge and Shan, 2016]**

http://www.jstor.org/stable/10.4169

Running example 1 “GOT”

“Game of Thrones” coappearances network: weighted and undirected network with 107 vertices corresponding to unique characters and 353 edges weighted by the number of times the two characters’ names appeared within 15 words of each other in the Game of Thrones series by George R.R. Martin.

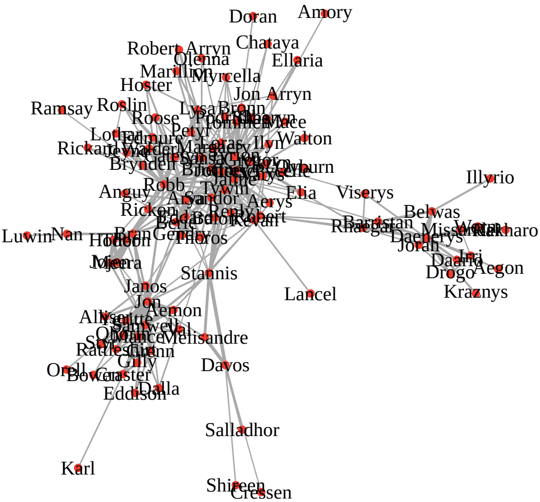
Reference: **[Beveridge and Shan, 2016]**
http://www.jstor.org/stable/10.4169

Dataset available at: http://www.macalester.edu/~abeverid/data/stormofswords.csv (edgelist format)

```
Source,Target,Weight
Aemon,Grenn,5
Aemon,Samwell,31
Aerys,Jaime,18
Aerys,Robert,6
Aerys,Tyrion,5
Aerys,Tywin,8
```

Running example 1 “GOT”

“Game of Thrones” coappearances network



Running example 2 “NVV”

my facebook network (extracted from facebook in 2015) with 152 vertices (my friends on facebook) and 551 edges (mutual friendship between my friends)

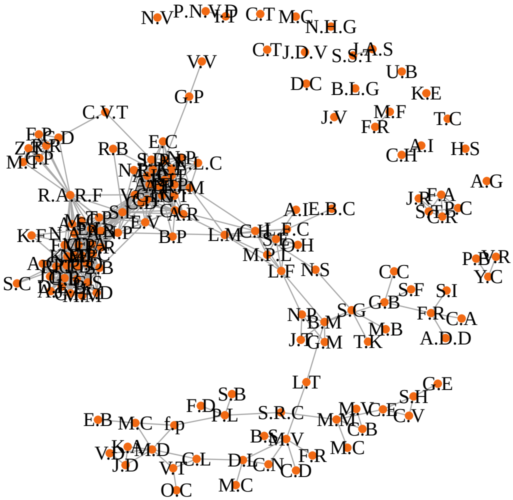
Running example 2 "NVV"

my facebook network (extracted from facebook in 2015) with 152 vertices (my friends on facebook) and 551 edges (mutual friendship between my friends)

Dataset available at: http://www.nathalievialaneix.eu/doc/txt/fbnet-el-2015.txt (edge list) and http://www.nathalievialaneix.eu/doc/txt/fbnet-name-2015.txt (metadata -initials- for the vertices)

Running example 2 “NVV”

my facebook network (extracted from facebook in 2015)



Running example 3 "FB"

Amherst College `https://www.amherst.edu` facebook network : Snapshots of within-college social networks of the first 100 colleges and universities admitted to `thefacebook.com`, in September 2005. Vertices are annotated with metadata giving the type of account (student, faculty, alumni, etc.), dorm, major, gender, and graduation year (2,235 vertices and 90,954 edges).

Running example 3 “FB”

Amherst College https://www.amherst.edu facebook network : Snapshots of within-college social networks of the first 100 colleges and universities admitted to thefacebook.com, in September 2005. Vertices are annotated with metadata giving the type of account (student, faculty, alumni, etc.), dorm, major, gender, and graduation year (2,235 vertices and 90,954 edges).

Reference: **[Traud et al., 2012]** http://arxiv.org/abs/1102.2166

Running example 3 "FB"

Amherst College https://www.amherst.edu facebook network : Snapshots of within-college social networks of the first 100 colleges and universities admitted to thefacebook.com, in September 2005. Vertices are annotated with metadata giving the type of account (student, faculty, alumni, etc.), dorm, major, gender, and graduation year (2,235 vertices and 90,954 edges).

Reference: [Traud et al., 2012] http://arxiv.org/abs/1102.2166

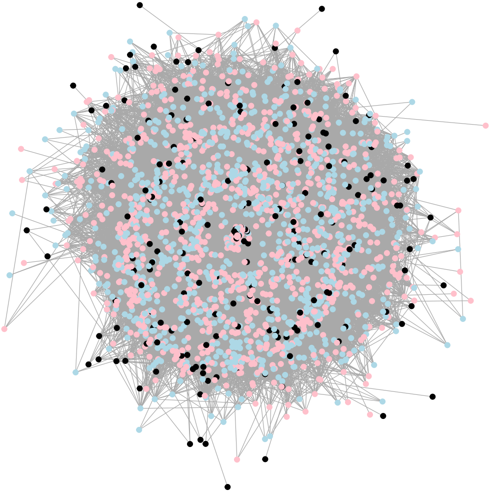
Dataset available at: https://escience.rpi.edu/data/DA/fb100 (Matlab© format; adjacency matrix + data frame with information on vertices: 7 columns)

0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0

1	2	0	0	359	2008	50112
1	1	106	103	340	2007	11279
1	1	114	0	0	2007	16202
1	1	99	0	0	2006	9076
1	2	111	109	347	2007	17773
1	1	0	0	0	2009	3576
1	1	0	0	360	2009	9414
1	2	99	117	340	2006	0
1	1	103	102	0	0	0
1	2	0	0	358	2009	50468
1	1	114	0	360	2008	1355
2	1	113	102	328	2004	0
1	2	99	117	335	2006	50460
1	2	0	0	0	2009	24694
1	1	0	0	0	2009	51059

Running example 3 “FB”

Amherst College `https://www.amherst.edu` facebook network



Connected component

The graph is said to be connected if any vertex can be reached from any other vertex by a path along the edges.

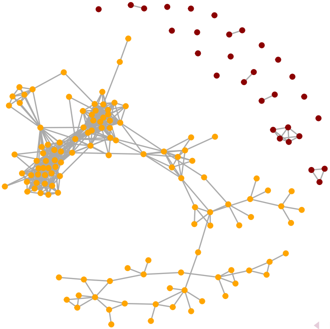
The connected components of a graph are all its connected subgraphs with maximum sizes.

Connected component

The graph is said to be connected if any vertex can be reached from any other vertex by a path along the edges.

The connected components of a graph are all its connected subgraphs with maximum sizes.

Examples: GOT and FB are connected graphs. NVV is not connected and contains 21 connected components, among which the largest has 122 vertices.



Outline

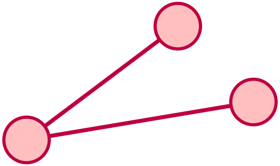
Visualization tools help understand the graph macro-structure

Purpose: How to display the vertices in a meaningful and aesthetic way?

Visualization tools help understand the graph macro-structure

Purpose: How to display the vertices in a meaningful and aesthetic way?

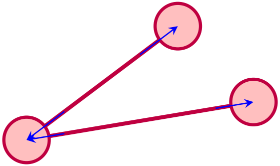
Standard approach: force directed placement algorithms (FDP) (e.g., **[Fruchterman and Reingold, 1991]**)



Visualization tools help understand the graph macro-structure

Purpose: How to display the vertices in a meaningful and aesthetic way?

Standard approach: force directed placement algorithms (FDP) (e.g., **[Fruchterman and Reingold, 1991]**)

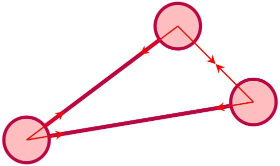


- attractive forces: similar to springs along the edges

Visualization tools help understand the graph macro-structure

Purpose: How to display the vertices in a meaningful and aesthetic way?

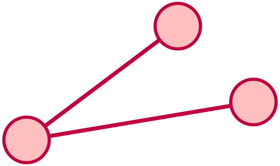
Standard approach: force directed placement algorithms (FDP) (e.g., **[Fruchterman and Reingold, 1991]**)



- attractive forces: similar to springs along the edges
- repulsive forces: similar to electric forces between all pairs of vertices

Visualization tools help understand the graph macro-structure

Purpose: How to display the vertices in a meaningful and aesthetic way?
Standard approach: force directed placement algorithms (FDP) (e.g., [Fruchterman and Reingold, 1991])



- attractive forces: similar to springs along the edges
- repulsive forces: similar to electric forces between all pairs of vertices

iterative algorithm until stabilization of the vertex positions.

Visualization software

- R package `igraph`[1] [Csardi and Nepusz, 2006] (static representation with useful tools for graph mining)

[1] http://igraph.sourceforge.net/
[2] http://gephi.org

Visualization software

- R package `igraph`[1] [Csardi and Nepusz, 2006] (static representation with useful tools for graph mining)
- free software Gephi[2] (interactive software, supports zooming and panning)

1 http://igraph.sourceforge.net/
2 http://gephi.org

Outline

Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

Examples

Example 1: GOT

- 107 vertices, 352 edges $\Rightarrow$ density $= \frac{352}{107\times 106/2} \simeq 6.2\%$.

Example 2: NVV

- 152 vertices, 551 edges $\Rightarrow$ density $\simeq 4.8\%$;
- largest connected component: 122 vertices, 535 edges $\Rightarrow$ density $\simeq 7.2\%$.

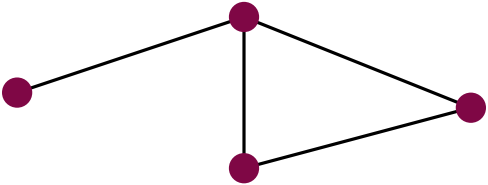
Example 3: FB

- 2235 vertices, 9.0954×10^4 edges $\Rightarrow$ density $\simeq 3.6\%$.

Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

Transitivity (sometimes called clustering coefficient): Number of triangles divided by the number of triplets connected by at least two edges. *What is the probability that two people with a common friend are also friends?*

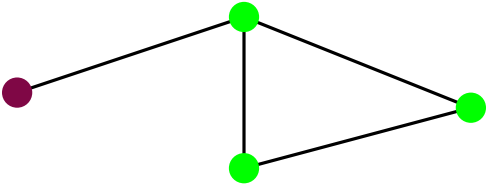


Density is equal to $\frac{4}{4\times 3/2} = 2/3$; Transitivity is equal to $1/3$.

Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

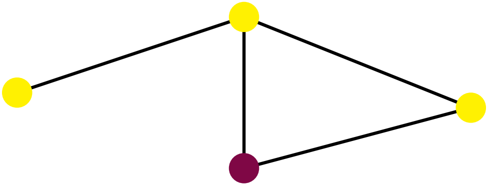
Transitivity (sometimes called clustering coefficient): Number of triangles divided by the number of triplets connected by at least two edges. *What is the probability that two people with a common friend are also friends?*



Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

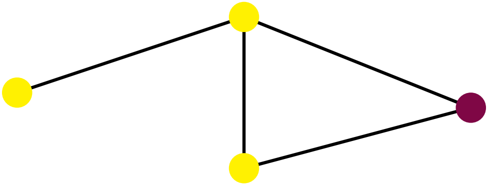
Transitivity (sometimes called clustering coefficient): Number of triangles divided by the number of triplets connected by at least two edges. *What is the probability that two people with a common friend are also friends?*



Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

Transitivity (sometimes called clustering coefficient): Number of triangles divided by the number of triplets connected by at least two edges. *What is the probability that two people with a common friend are also friends?*



Density / Transitivity

Density: Number of edges divided by the number of pairs of vertices. *Is the network densely connected?*

Transitivity (sometimes called clustering coefficient): Number of triangles divided by the number of triplets connected by at least two edges. *What is the probability that two people with a common friend are also friends?*

Examples

Example 1: GOT

- density $\simeq$ 6.2%, transitivity $\simeq$ 32.9%.

Example 2: NVV

- density $\simeq$ 4.8%, transitivity $\simeq$ 56.2%;
- LCC: density $\simeq$ 7.2%, transitivity $\simeq$ 56%.

Example 3: FB

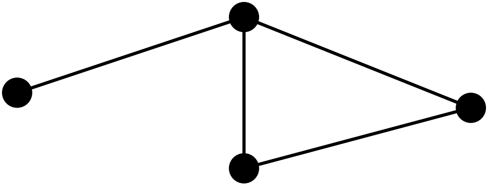
- density $\simeq$ 3.6%, transitivity $\simeq$ 23.3%.

Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.

Diameter and 6 degrees of separation

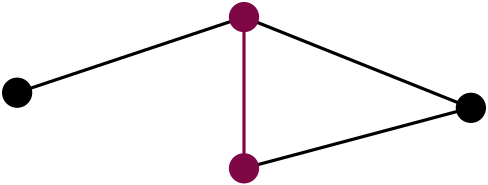
Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

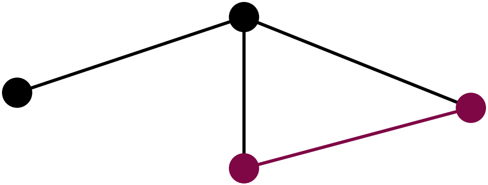
Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

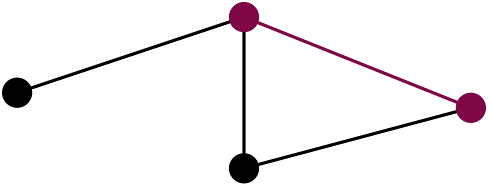
Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

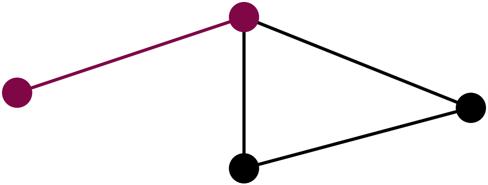
Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

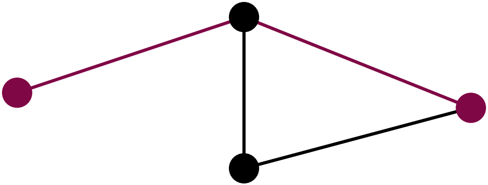
Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

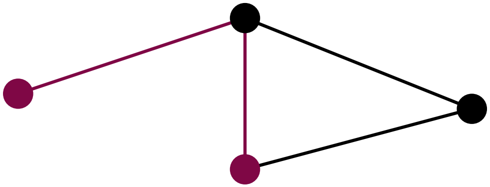
Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



The diameter of this graph is 2.

Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.

6 degrees of separation

From a volume of novels by Frigyes Karinthy: all living things and everything else in the world is six or fewer steps away from each other (*i.e.*, a chain of “a friend of a friend” statements can be made to connect any two people in a maximum of six steps).
Hypothesis tested by Milgram with a letter chain (1967): $\simeq 2 - 10$ intermediates to reach a target person from any starting people (known as the “small world experiment”).

Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.

Examples of diameter

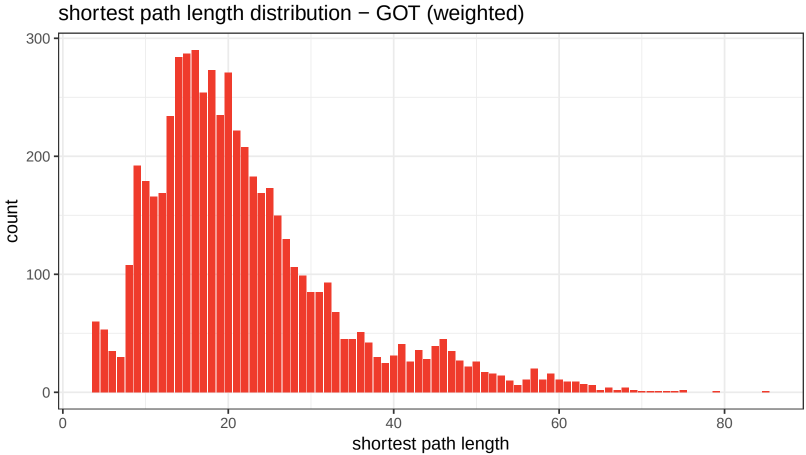
Example 1: GOT diameter = 6 (unweighted) and 85 (weighted)
Example 2: NVV diameter in LCC: 18
Example 3: FB diameter = 7

Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.

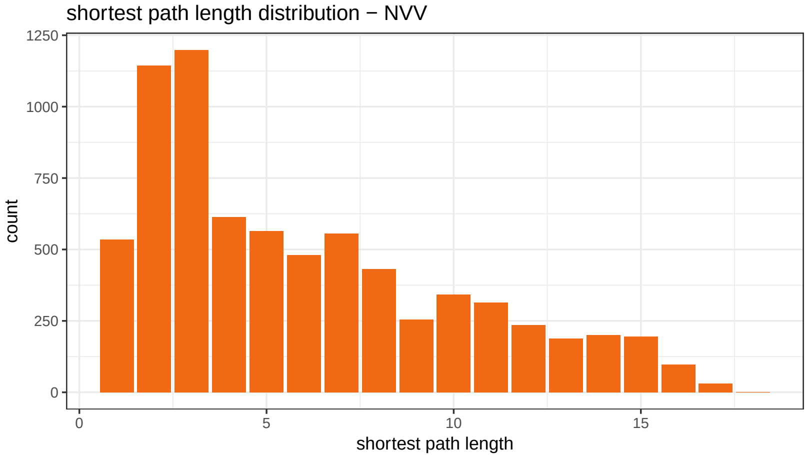
Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



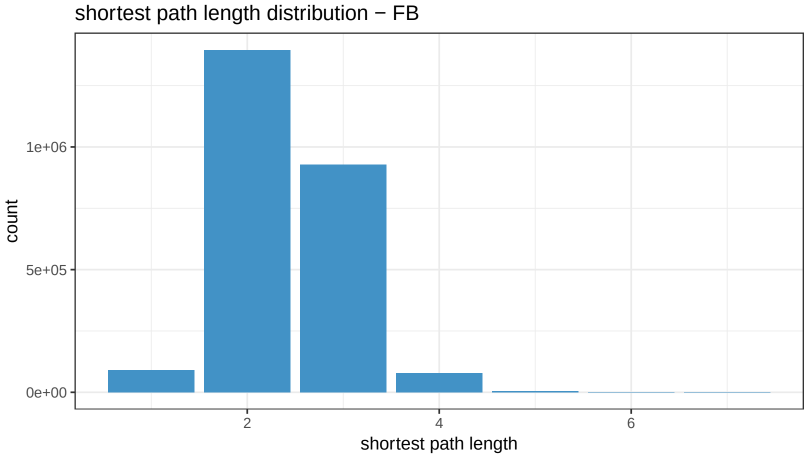
Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



Diameter and 6 degrees of separation

Diameter (of a connected graph): length of the longest shortest path between two vertices in the graph.



girth, cohesion

- girth: number of vertices in the shortest circle (equal to 3 if transitivity is not equal to 0)
- cohesion (of a connected graph): minimum number of vertices to remove in order to disconnect the graph

girth, cohesion

- girth: number of vertices in the shortest circle (equal to 3 if transitivity is not equal to 0)
- cohesion (of a connected graph): minimum number of vertices to remove in order to disconnect the graph

Examples

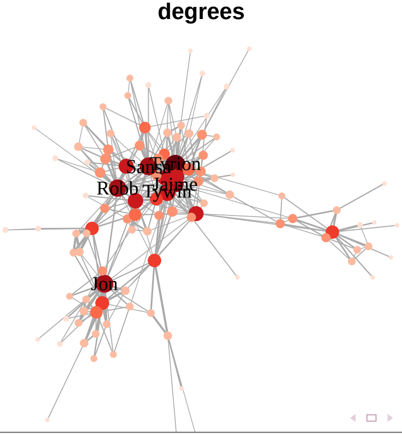
Example 1: GOT girth = 3 and cohesion = 1
Example 2: NVV girth = 3 and cohesion = 1
Example 3: FB girth = 3 and cohesion = 1

Outline

Extracting important vertices: hubs

vertex degree: number of edges adjacent to the vertex $\left|\{x_j : (x_i, x_j) \in E, j \neq i\}\right|$. The weighted version $\sum_{j \neq i} W_{ij}$ is called the strength.

Vertices with a high degree are called hubs: measure of the vertex popularity.



Extracting important vertices: hubs

vertex degree: number of edges adjacent to the vertex $\left|\{x_j : (x_i, x_j) \in E, j \neq i\}\right|$. The weighted version $\sum_{j\neq i} W_{ij}$ is called the strength.

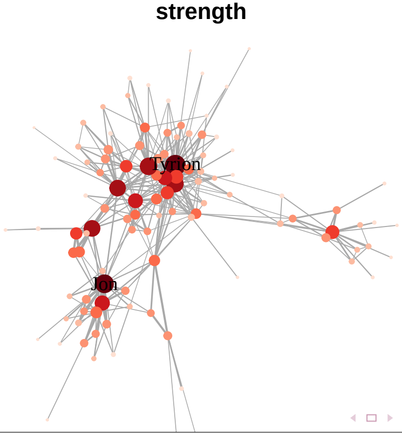
Vertices with a high degree are called hubs: measure of the vertex popularity.

	degrees
Jaime	24
Tyrion	36
Tywin	22
Jon	26
Robb	25
Sansa	26

(degree larger than 20)

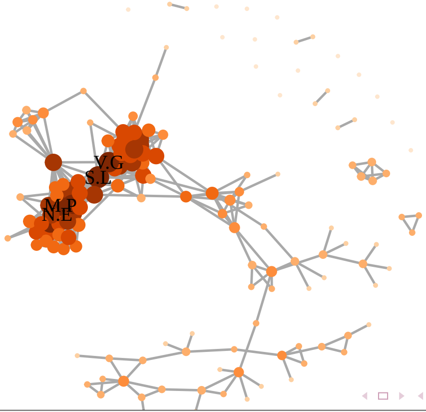
Extracting important vertices: hubs

- vertex degree: number of edges adjacent to the vertex $\left|\{x_j :\ (x_i, x_j) \in E,\ j \neq i\}\right|$. The weighted version $\sum_{j \neq i} W_{ij}$ is called the strength.
- Vertices with a high degree are called hubs: measure of the vertex popularity.



Extracting important vertices: hubs

- vertex degree: number of edges adjacent to the vertex $\left|\{x_j : (x_i, x_j) \in E, j \neq i\}\right|$. The weighted version $\sum_{j \neq i} W_{ij}$ is called the strength.
- Vertices with a high degree are called hubs: measure of the vertex popularity.



Extracting important vertices: hubs

vertex degree: number of edges adjacent to the vertex $\left|\{x_j : (x_i, x_j) \in E, j \neq i\}\right|$. The weighted version $\sum_{j \neq i} W_{ij}$ is called the strength.

Vertices with a high degree are called hubs: measure of the vertex popularity.

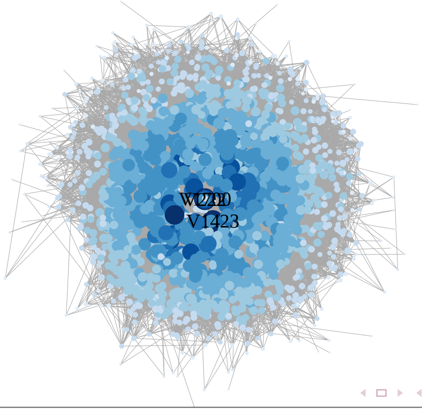
	degrees
S.L	31
M.P	29
V.G	27
N.E	27

(degree larger than 25)

Extracting important vertices: hubs

vertex degree: number of edges adjacent to the vertex $\left|\{x_j : (x_i, x_j) \in E,\ j \neq i\}\right|$. The weighted version $\sum_{j \neq i} W_{ij}$ is called the strength.

Vertices with a high degree are called hubs: measure of the vertex popularity.



Extracting important vertices: hubs

vertex degree: number of edges adjacent to the vertex $\left|\{x_j : (x_i, x_j) \in E, j \neq i\}\right|$. The weighted version $\sum_{j \neq i} W_{ij}$ is called the strength.

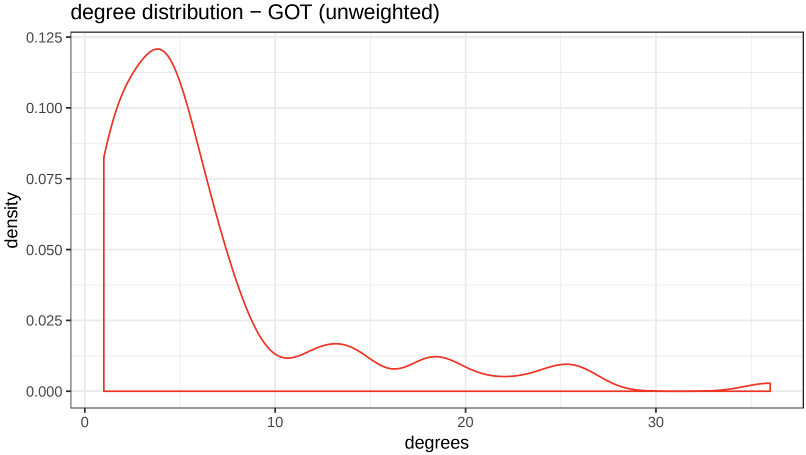
Vertices with a high degree are called hubs: measure of the vertex popularity.

	degrees
V222	456
V1423	467
V1700	467

(degree larger than 400)

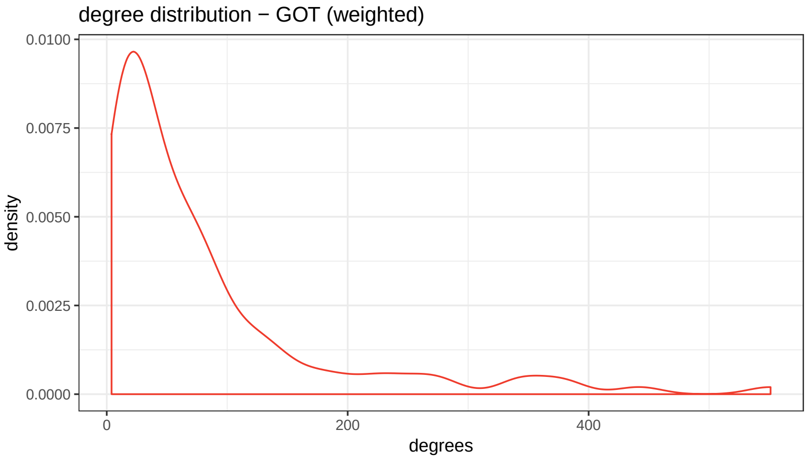
Degree distribution

In real graphs (WWW, social networks...), the degree distribution is often found to fit a power law: $\mathbb{P}(\text{degree} = k) \sim k^{-\gamma}$ for a $\gamma > 0$.



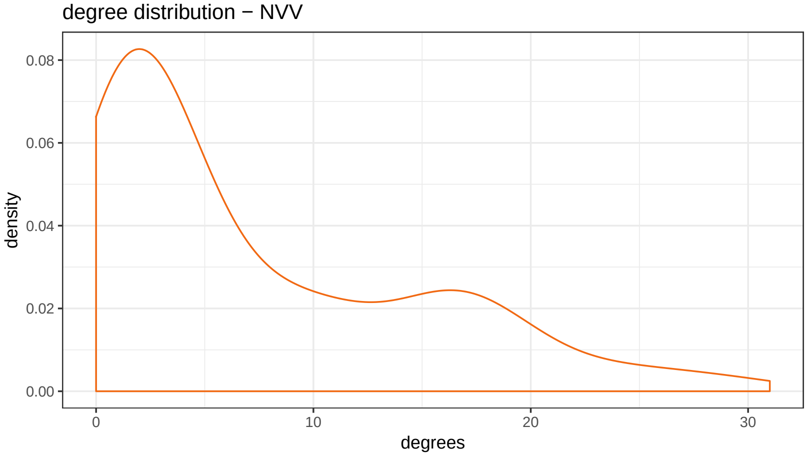
Degree distribution

In real graphs (WWW, social networks...), the degree distribution is often found to fit a power law: $\mathbb{P}(\text{degree} = k) \sim k^{-\gamma}$ for a $\gamma > 0$.



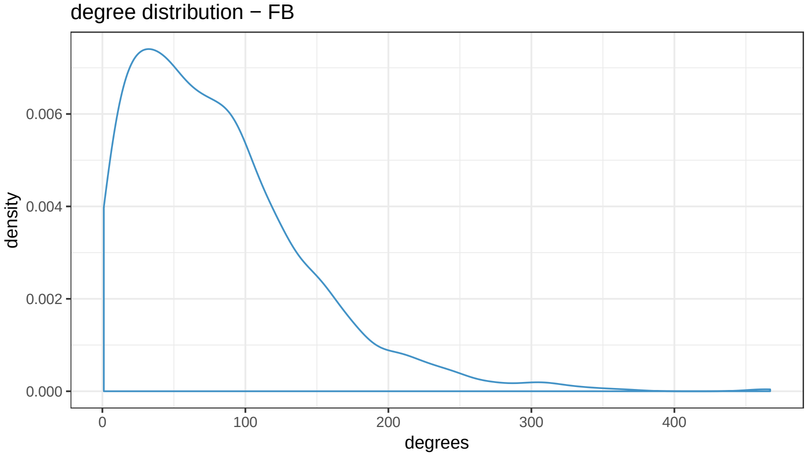
Degree distribution

In real graphs (WWW, social networks...), the degree distribution is often found to fit a power law: $\mathbb{P}(\text{degree} = k) \sim k^{-\gamma}$ for a $\gamma > 0$.



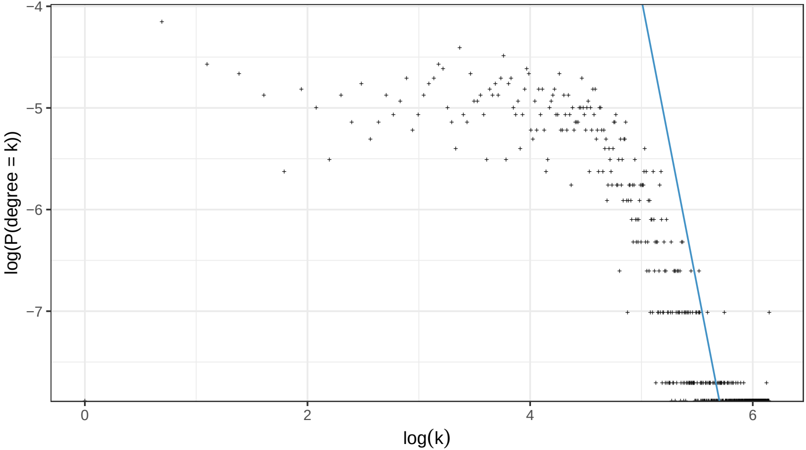
Degree distribution

In real graphs (WWW, social networks...), the degree distribution is often found to fit a power law: $\mathbb{P}(\text{degree} = k) \sim k^{-\gamma}$ for a $\gamma > 0$.



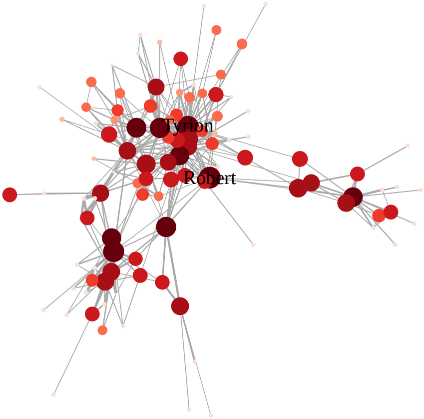
Degree distribution

In real graphs (WWW, social networks...), the degree distribution is often found to fit a power law: $\mathbb{P}(\text{degree} = k) \sim k^{-\gamma}$ for a $\gamma > 0$.



Extracting important vertices: betweenness

vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure indicating which vertices are the most important to connect the network.



Extracting important vertices: betweenness

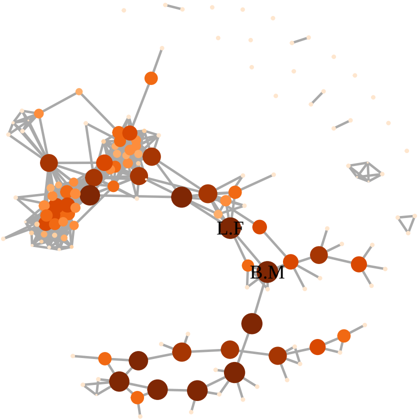
vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure indicating which vertices are the most important to connect the network.

	betweenness	degree
Robert	1166	18
Tyrion	1164	36

(betweenness larger than 1000; hubs had a degree larger than 20)

Extracting important vertices: betweenness

vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure indicating which vertices are the most important to connect the network.



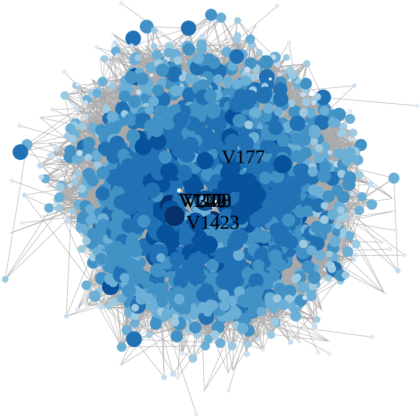
Extracting important vertices: betweenness

vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure indicating which vertices are the most important to connect the network.

	betweenness
B.M	3439
L.F	3146

Extracting important vertices: betweenness

vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure indicating which vertices are the most important to connect the network.



Extracting important vertices: betweenness

vertex betweenness: number of shortest paths between all pairs of vertices that pass through the vertex. Betweenness is a centrality measure indicating which vertices are the most important to connect the network.

	betweenness	degree
V177	30797	253
V222	30649	456
V340	49127	299
V1173	30778	313
V1423	60272	467
V1700	37868	467

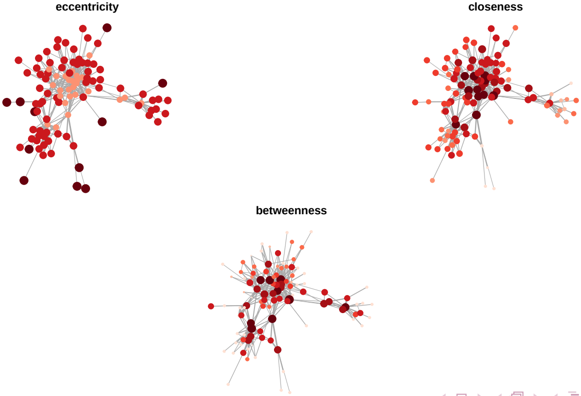
(betweenness larger than 30,000; hubs had a degree larger than 400)

Other centrality measure: eccentricity and closeness

vertex eccentricity: shortest path length from the farthest other vertex in the graph (the smallest eccentricity is the radius)

vertex closeness: inverse of the average length of the shortest paths from this vertex to all the other vertices in the graph:

$$\frac{1}{\sum_{j\neq i} \mathrm{spl}(i,j)}$$



Other centrality measure: eccentricity and closeness

vertex eccentricity: shortest path length from the farthest other vertex in the graph (the smallest eccentricity is the radius)

vertex closeness: inverse of the average length of the shortest paths from this vertex to all the other vertices in the graph: $\frac{1}{\sum_{j \neq i} \text{spl}(i,j)}$

Radius

Example 1: GOT radius = 3

Example 2: NVV radius in LCC: 9

Example 3: FB radius = 4

Beveridge, A. and Shan, J. (2016).
Network of thrones.
Math Horizons, 23(4):18–22.

Csardi, G. and Nepusz, T. (2006).
The igraph software package for complex network research.
InterJournal, Complex Systems.

Fruchterman, T. and Reingold, B. (1991).
Graph drawing by force-directed placement.
Software, Practice and Experience, 21:1129–1164.

Traud, A., Mucha, P., and Porter, M. (2012).
Social structure of facebook networks.
Physica A, 391(16):4165–4180.