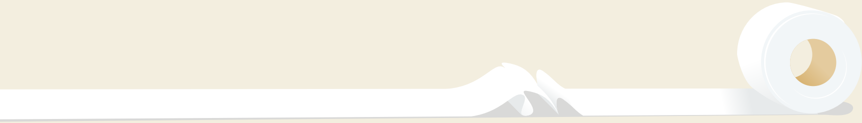
Complex networks & sparsity
Part I: Introduction



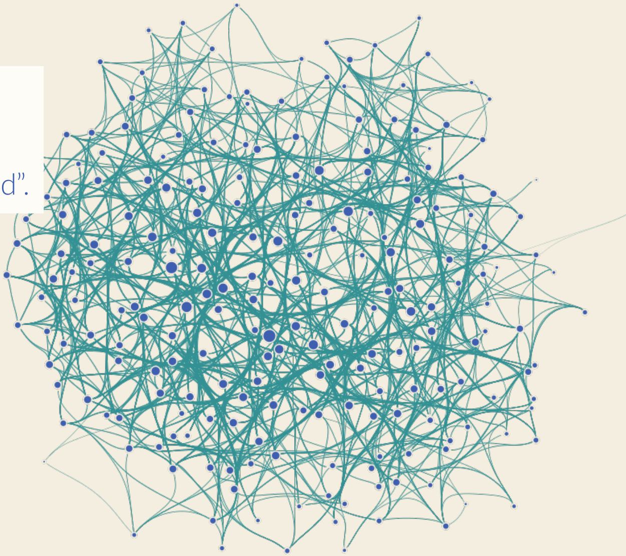
Felix Reidl
Blair D. Sullivan
DOCCOURSE '18

A Grand Tour

Residence hall

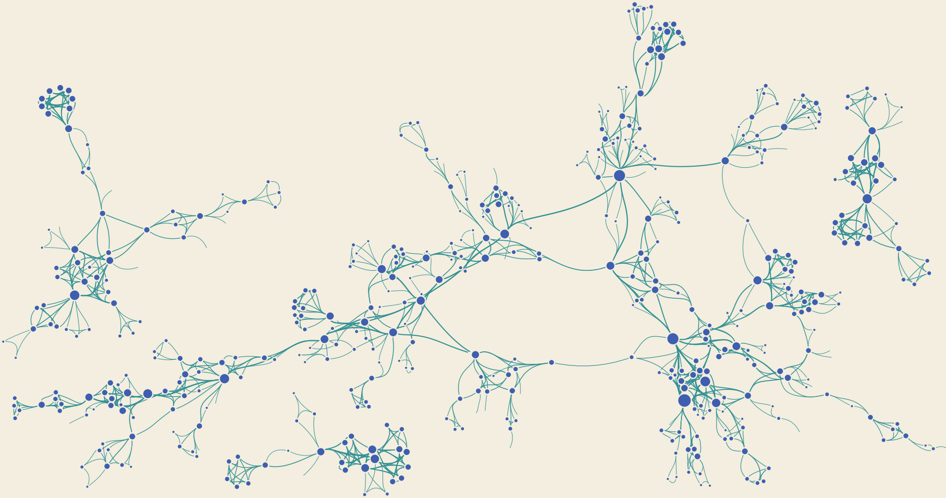
- Student in ANU Hall
- Friendship

Collected via interviews by Cynthia Webster, ranked as “best friend”, “close friend”, and “friend”.



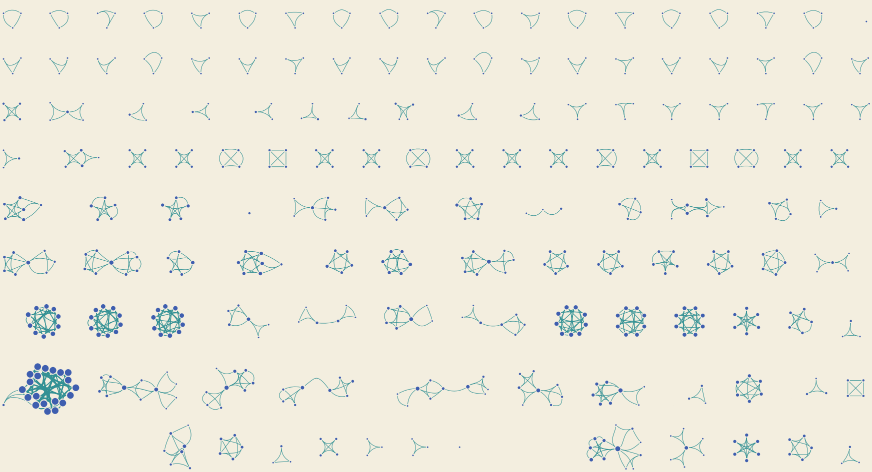
Netscience

- Researchers
- Coauthorship

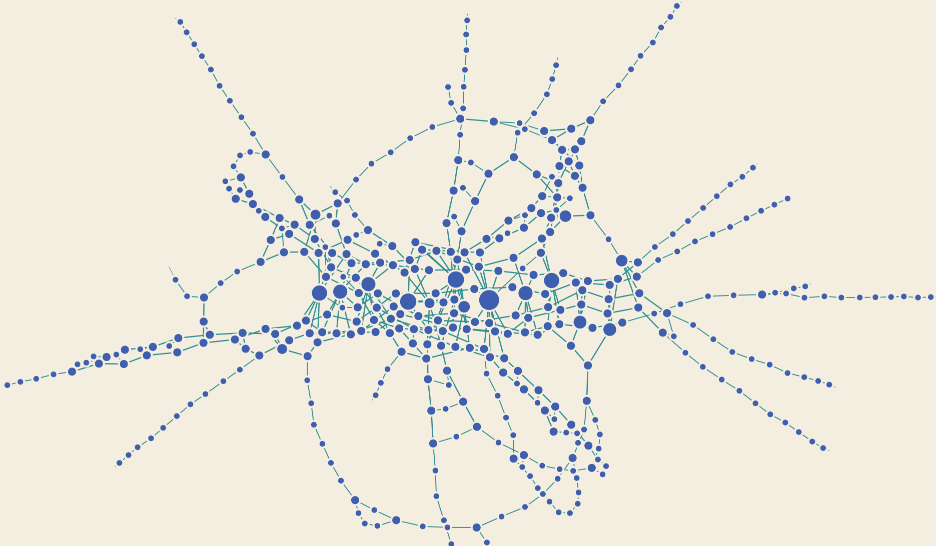


Netscience (cont'd)

- Researchers
- Coauthorship

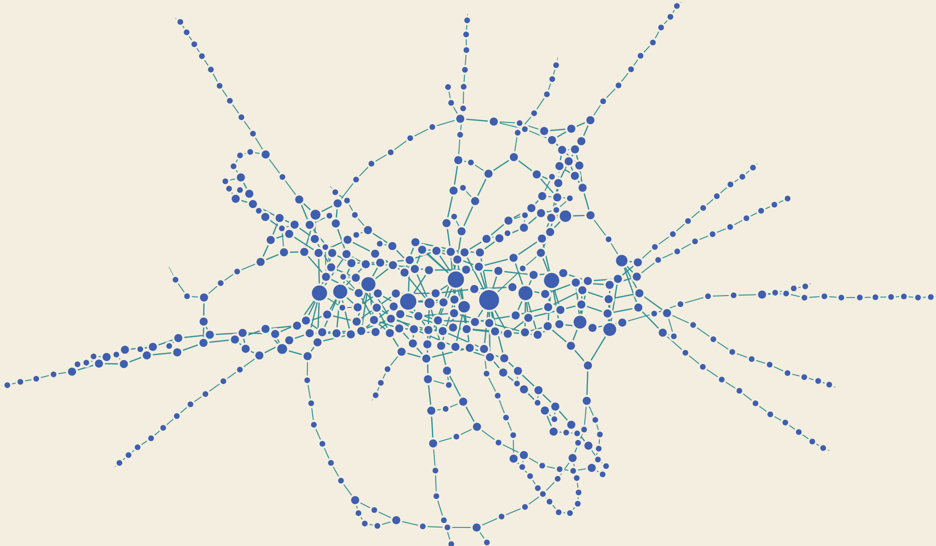


Munich

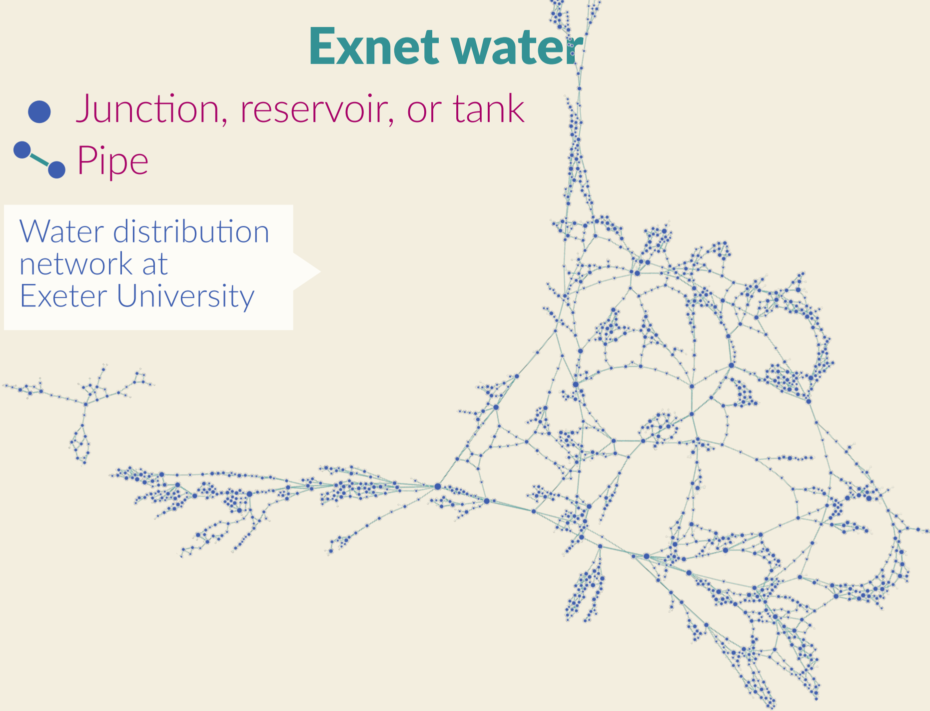


Munich

- Tram and S-bahn stops
- Rail connections



Exnet water

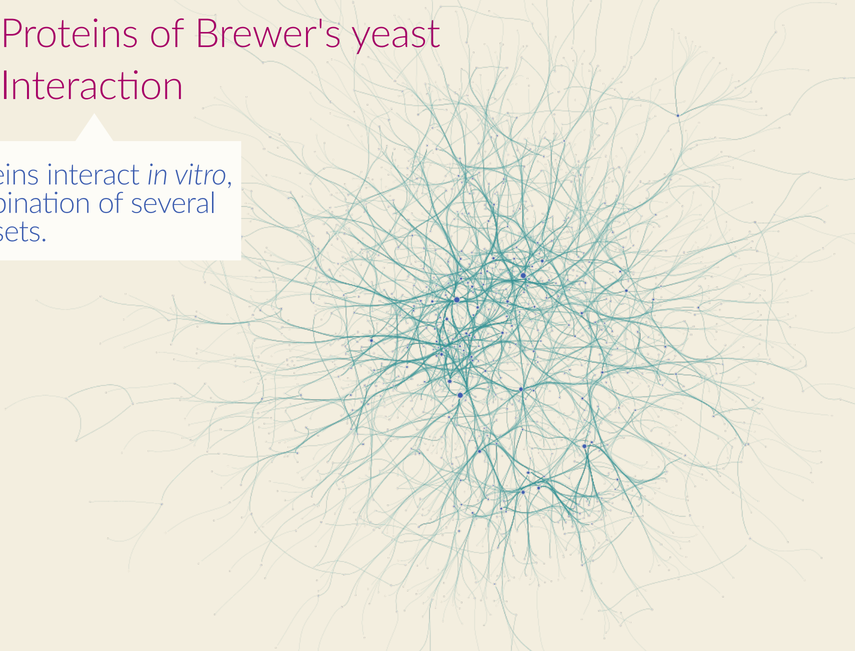


Junction, reservoir, or tank
Pipe
Water distribution network at Exeter University

Y2H union (yeast)

- Proteins of Brewer's yeast
- Interaction

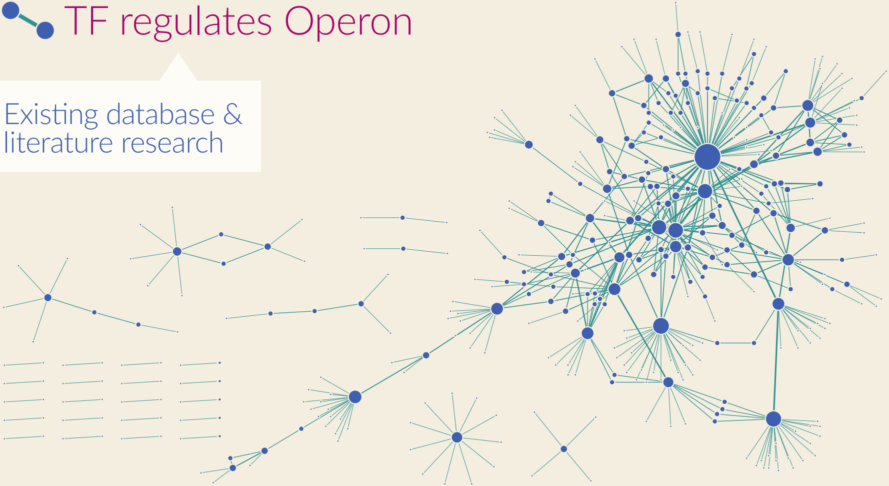
Proteins interact *in vitro*, combination of several datasets.



E-coli

- Operons & Transcription factors
- TF regulates Operon

Existing database & literature research



Diseasome

- Gene / Genetic disorder
- Correlation: expression/disease

Extracted from OMIM database.

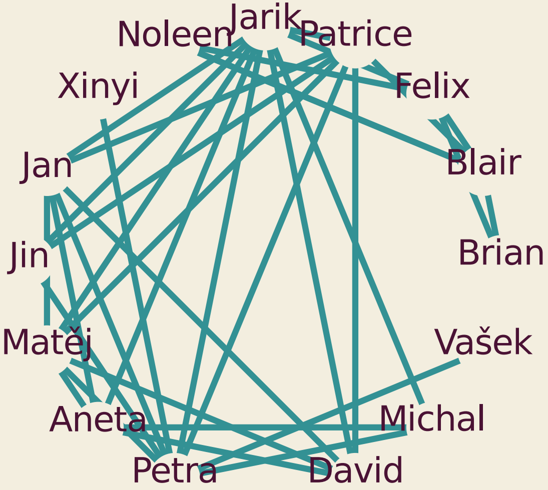
The DOCCOURSE network

- Participants of DOCCOURSE '18
- Had a beer together last week

As queried at the Český Krumlov excursion.

15 vertices
33 edges

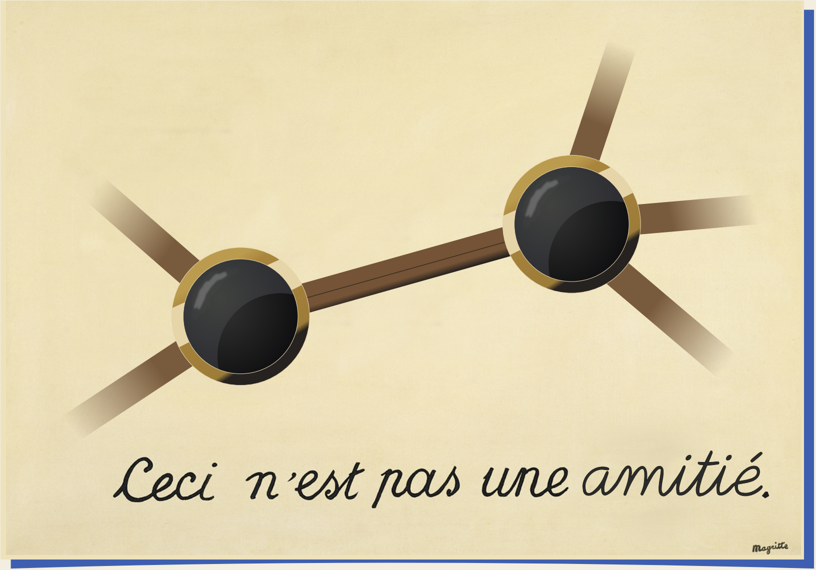
min degree 1
avg. degree 4.4
max degree 9



Beware of network

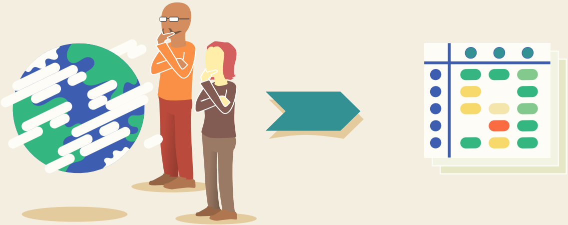
The treachery of images

The treachery of images

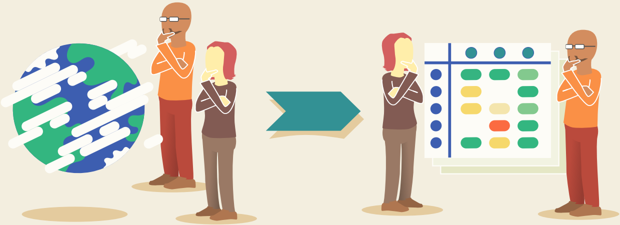


How data is ~~collected~~ constructed

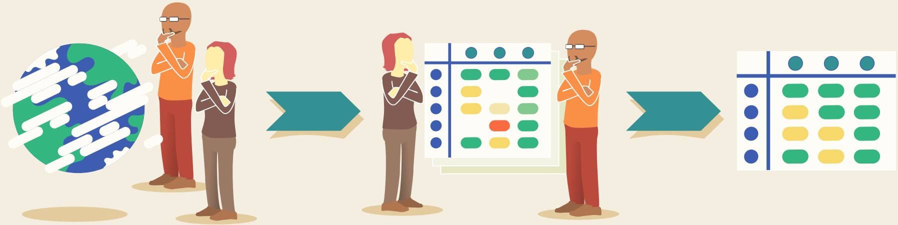
How data is ~~collected~~ constructed



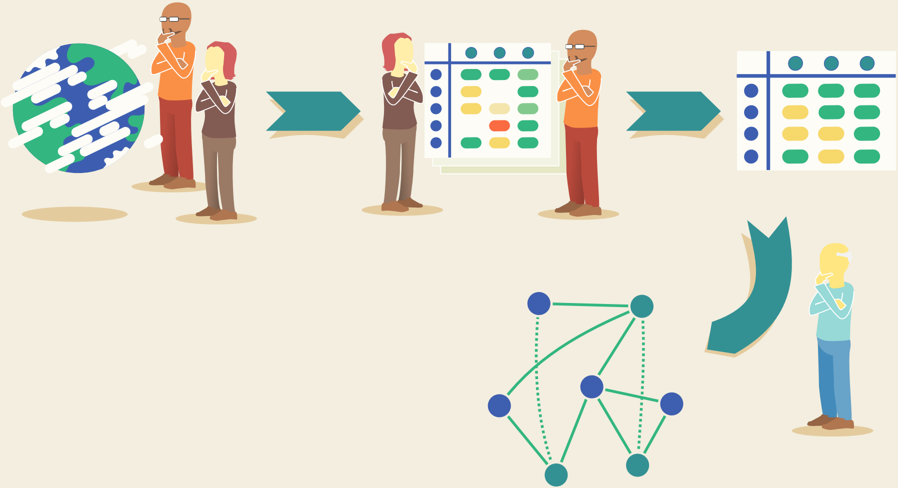
How data is ~~collected~~ constructed



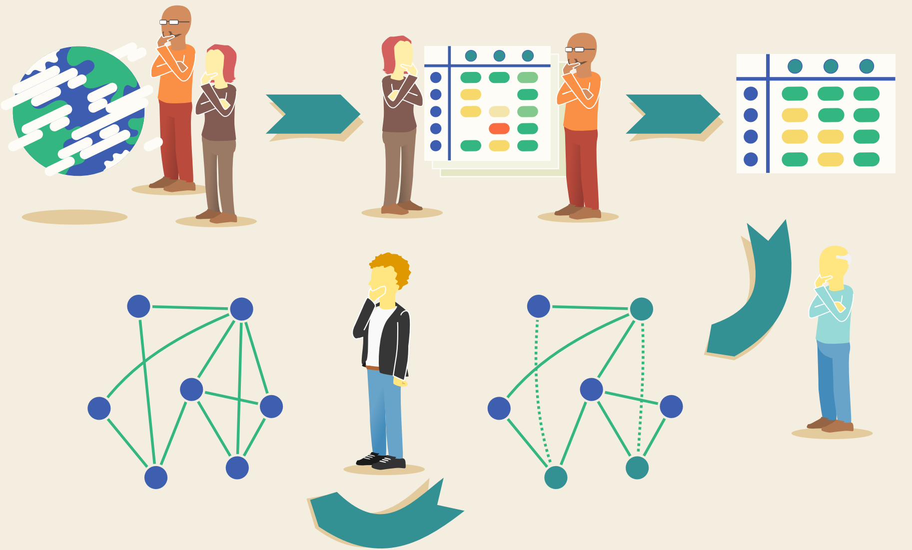
How data is ~~collected~~ constructed



How data is ~~collected~~ constructed



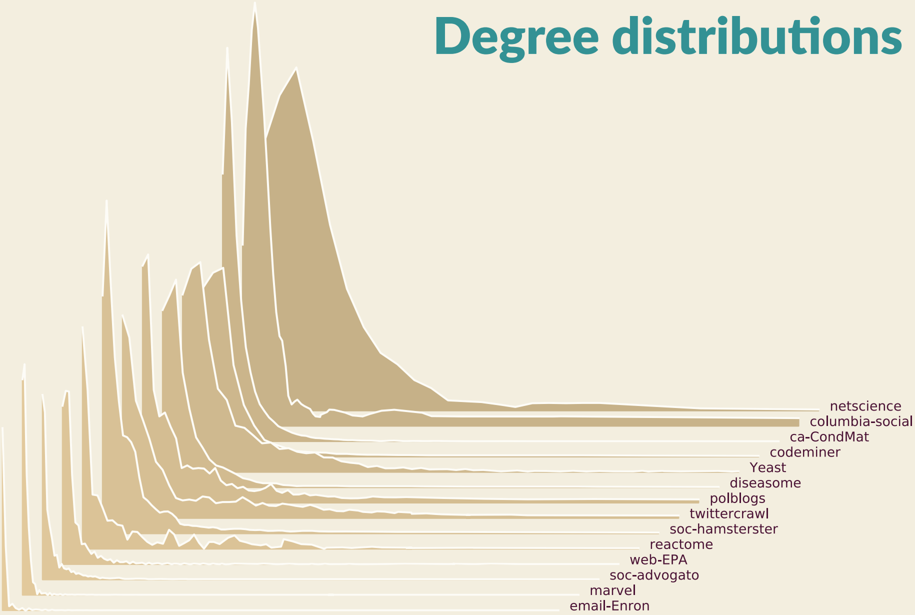
How data is ~~collected~~ constructed



Key Characteristics

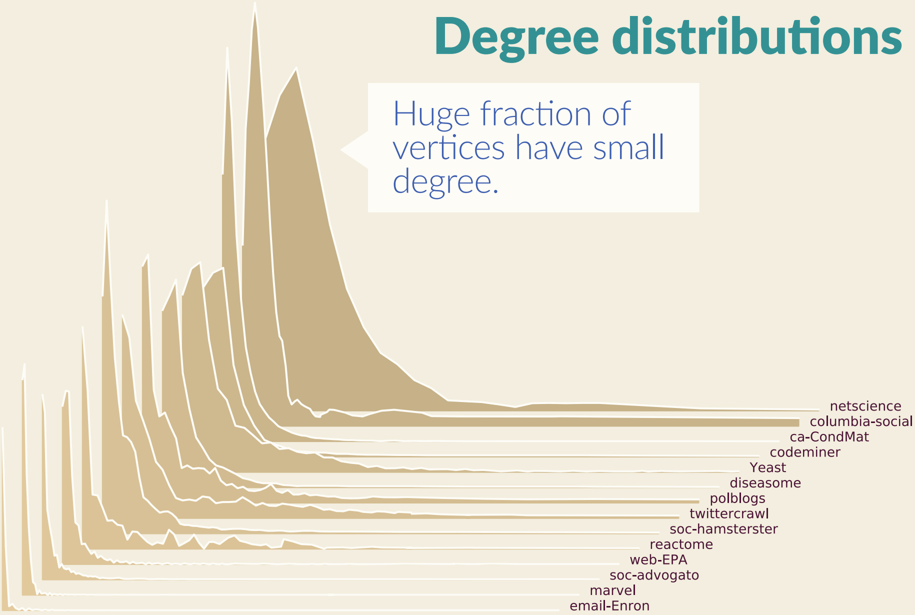
degree distribution

Degree distributions

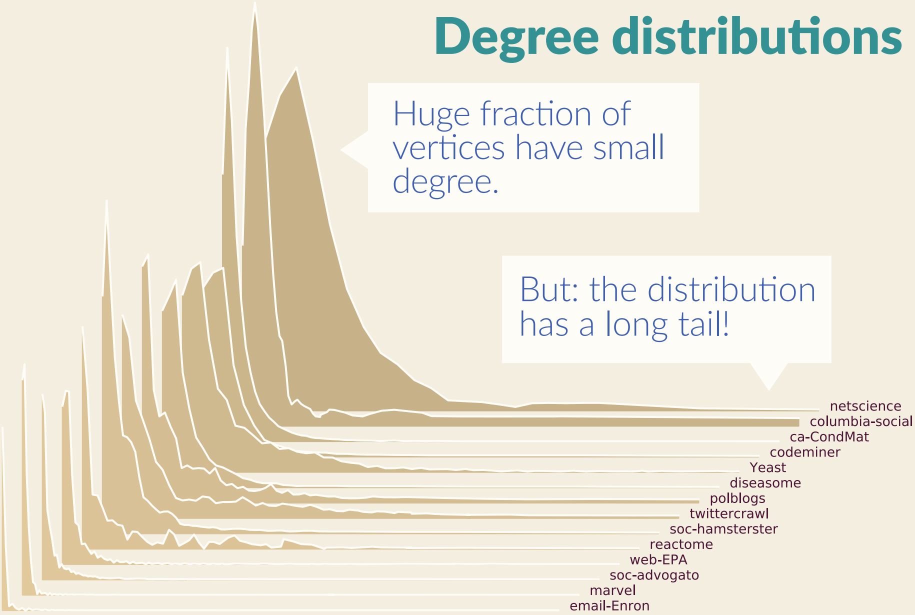


Degree distributions

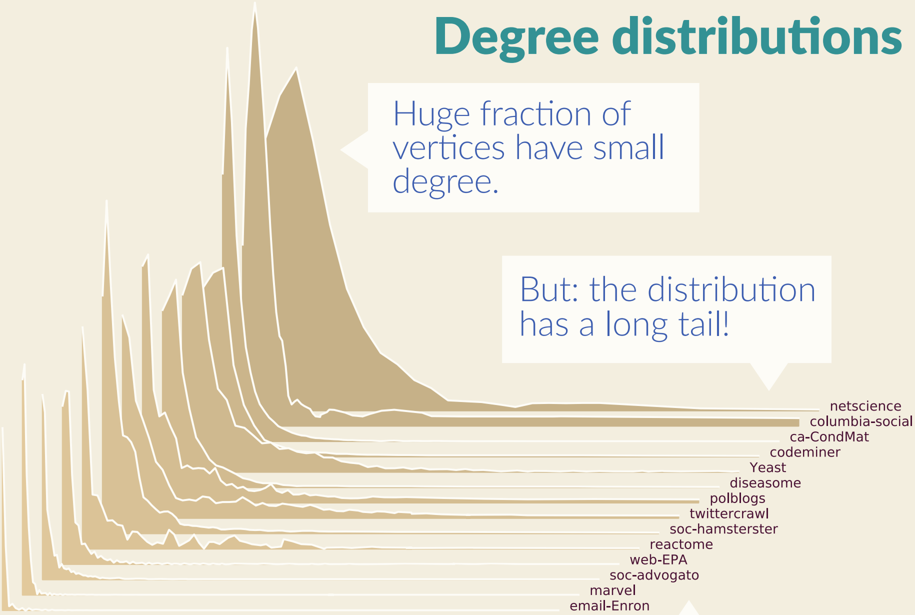
Huge fraction of vertices have small degree.



Degree distributions



Degree distributions



Huge fraction of vertices have small degree.

But: the distribution has a long tail!

More extreme in large networks!

Powerlaws...?

A *powerlaw distribution* has the form

$$f(d) = \frac{\lambda}{d^\gamma}.$$

Networks with powerlaw degree distributions are sometimes called *scale-free**.

* There are diverging opinions on what *scale-free* means.

Powerlaws...?

A *powerlaw distribution* has the form

$$f(d) = \frac{\lambda}{d^\gamma}.$$

Networks with powerlaw degree distributions are sometimes called *scale-free**.

It is often clamed that "complex networks have powerlaw degree distributions" or that they are "scale-free".

* There are diverging opinions on what *scale-free* means.

Powerlaws...?

A *powerlaw distribution* has the form

$$f(d) = \frac{\lambda}{d^\gamma}.$$

Networks with powerlaw degree distributions are sometimes called *scale-free**.

It is often clamed that “complex networks have powerlaw degree distributions” or that they are “scale-free”.

It is safe to say that the initial claims of ubiquitious powerlaws where overstated and based on very shoddy methodology.

* There are diverging opinions on what *scale-free* means.

The 'scale free' panic

GADGETS

Who Protects The Internet?

Pull up the wrong undersea cable, and the Internet goes dark in Berlin or Dubai. See our animated infographics of how the web works!

By James Geary, World Map Courtesy TeleGeography   *March 13, 2009*

Scale-Free Networks

Terremark and the other exchanges scattered across the country (Chicago, New York and Los Angeles are just a few of the other locations) are so vital because the Internet is a "scale-free network." In a scale-free network, connections are not randomly or evenly distributed. Some points have relatively few connections to other points (a single server in the basement of a small business, for example), and some points—known as hubs—have a relatively huge number of connections to other points (Terremark). This ratio of very connected hubs to less-connected

MailOnline Science & Tech

Home | News | U.S. | Sport | TV&Showbiz | Australia | Femail | Health | Science | Money | Video | Travel | DailyMailTV

Latest Headlines | Science | Pictures | Discounts   Login

The 'Kevin Bacon' effect online: Researchers reveal how EVERYTHING on the web is connected by just 19 clicks

- Hungarian physicist claims that because of huge 'superhubs' like Google and Facebook, web pages are more connected to each other than expected
- Warns that this effect could be used to attack the web - with attacks of the 'superconnecters' being the achilles heel of the online world

Hidden Vulnerability Discovered in the World's Airline Network

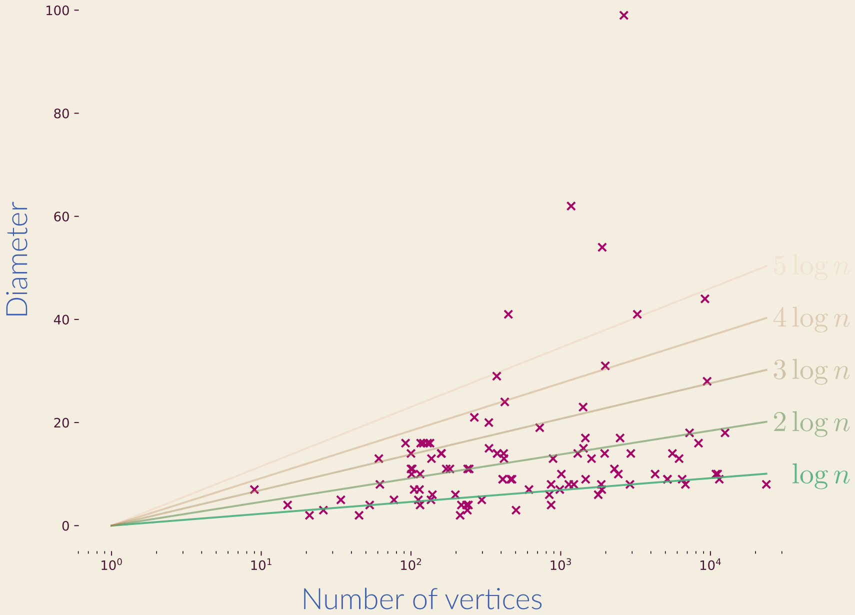
The global network of links between the world's airports looks robust but contains a hidden weakness that could lead to entire regions of the planet being cut off.

April 16, 2014

Key Characteristics

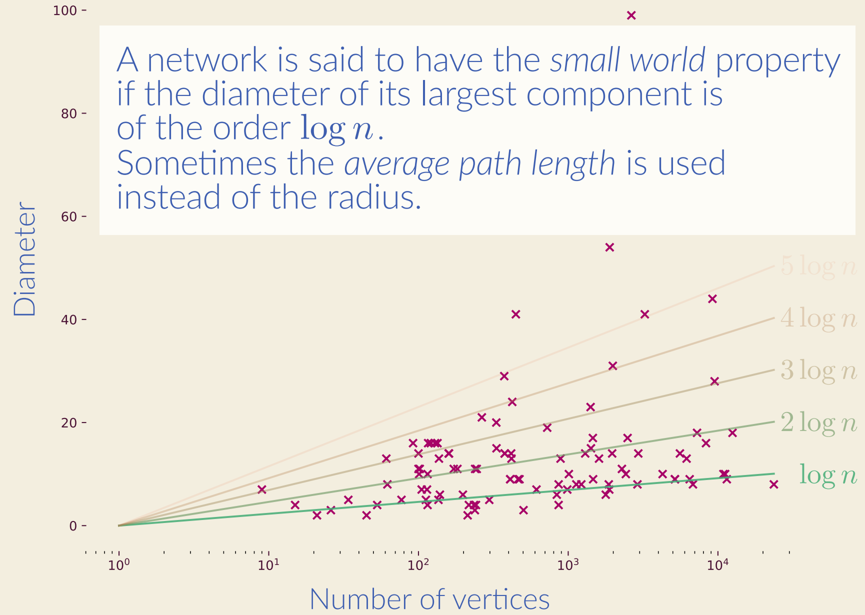
small diameter

Small world property



Small world property

A network is said to have the *small world* property if the diameter of its largest component is of the order $\log n$.
Sometimes the *average path length* is used instead of the radius.



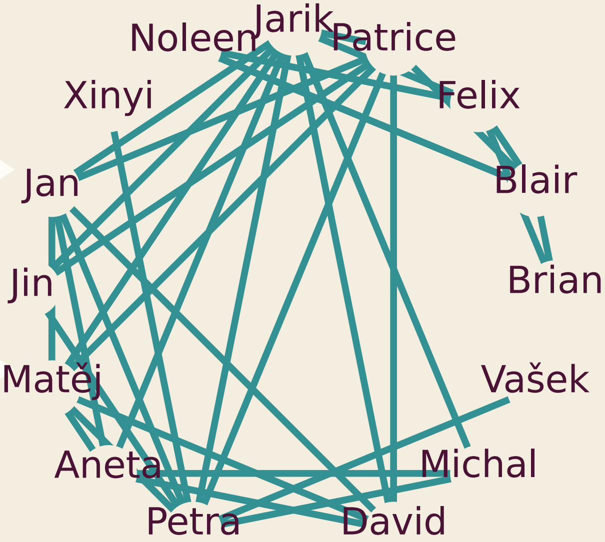
Krumlov is a small world

The diameter of our Krumlov network is

4

Patrice

is a center of the network.



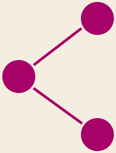
Key Characteristics

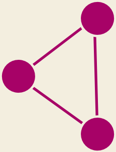
community structure

Clustering in social networks

We observe that two people with a common friend are more likely to be friends as well:

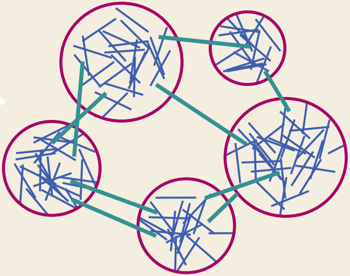
'unstable'





'stable'

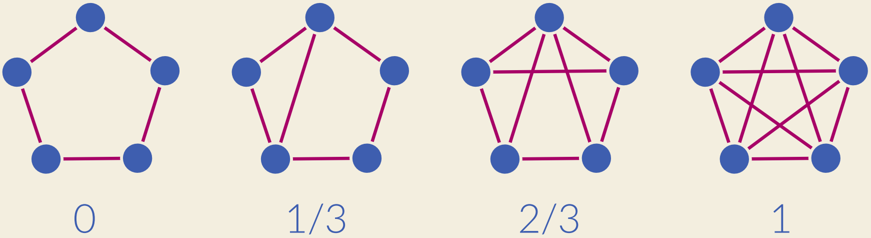
On a larger scale, we see that social networks are composed of *communities*



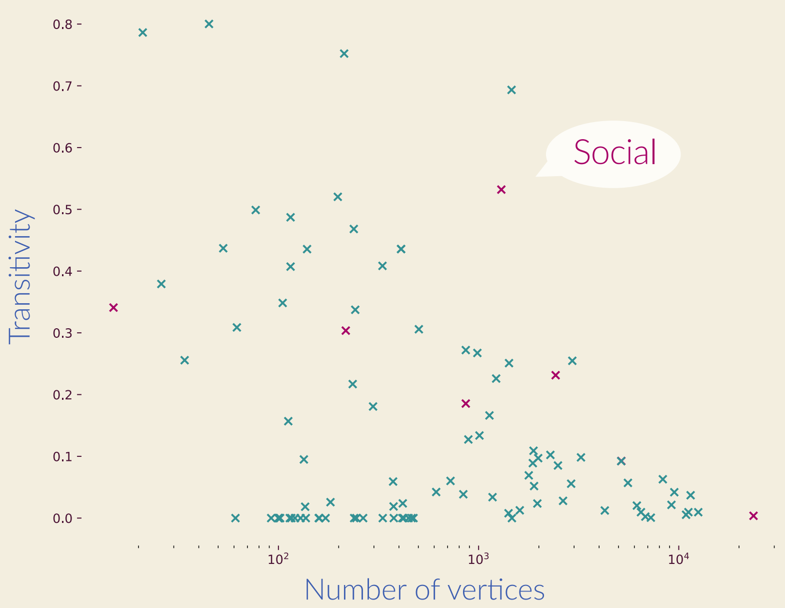
Transitivity

The transitivity of a graph G is defined as the quotient of the number of triangles over the number of (non-induced) P_3s:

$$T(G) = 3\frac{\#\triangle(G)}{\#\wedge(G)}$$

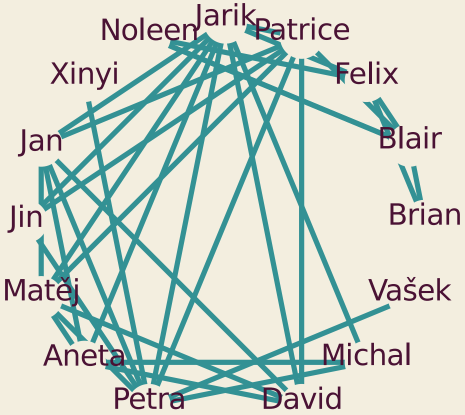


Transitivity



Transitivity

$$3\,\frac{\#\triangle(G)}{\#\wedge(G)} = \frac{87}{155} = 0.56$$



Average clustering coefficient

The *clustering coefficient* of a vertex v is defined as

$$C(v) = \|G[N(v)]\| \binom{|N(v)|}{2}^{-1}.$$

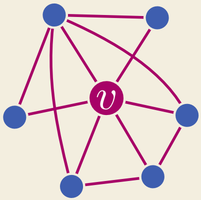
It measures the density of v's neighbourhood.

Average clustering coefficient

The *clustering coefficient* of a vertex v is defined as

$$C(v) = \|G[N(v)]\|\binom{|N(v)|}{2}^{-1}.$$

It measures the density of v's neighbourhood.



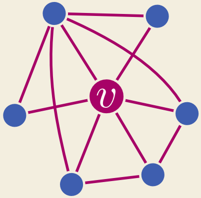
$$C(v) = 6 \cdot \frac{2}{6 \cdot 5} = 0.4$$

Average clustering coefficient

The *clustering coefficient* of a vertex v is defined as

$$C(v) = \|G[N(v)]\| \binom{|N(v)|}{2}^{-1}.$$

It measures the density of v's neighbourhood.

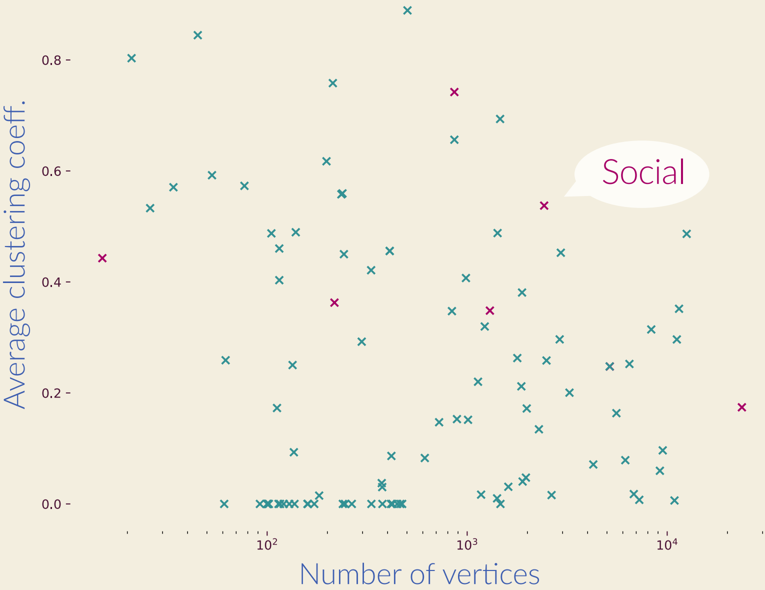


$$C(v) = 6 \cdot \frac{2}{6 \cdot 5} = 0.4$$

The *clustering coefficient* of a graph G is defined as

$$C(G) = \frac{1}{|G|} \sum_{v \in G} C(v).$$

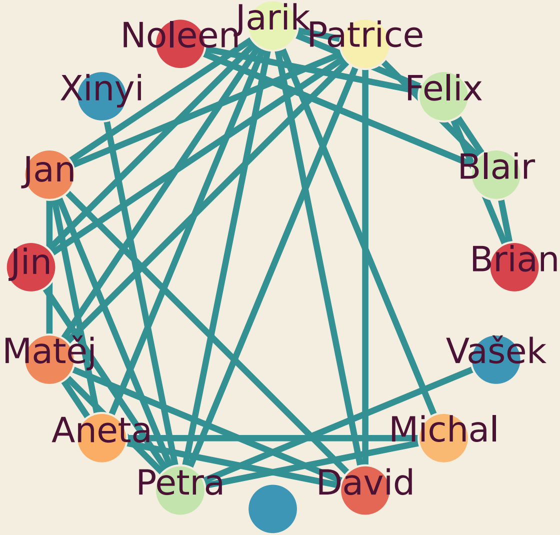
Average clustering coefficient



Clustering coefficient

$\frac{1}{15}$ (0.33+0.80+0.52+1.00+0.32+0.67+1.00+0.70+0.80+0.44+1.00+0

$= 0.59$



"Vašek"
has the largest
clustering coefficient.

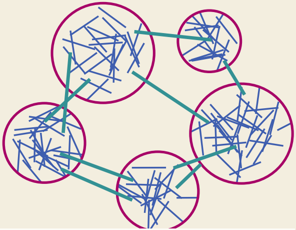
Complex networks: summary

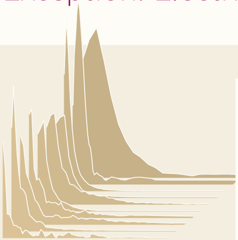
Most real world networks have a **very small** diameter.

Exception: Infrastructure networks

Many networks exhibit **clustering** and a **community structure**.

Exception: Electrical networks





Most networks have a highly **left-skewed** degree distribution or even a **long tail**.

Exception: de-Bruijn graphs

What's important?

Measure of centrality

The structuralist school of thought in sociology seeks to understand society through the relationship of its members (and not the members themselves).

Very simplified, we might ask:

> Which are the important/influencial nodes of a given network?

Measure of centrality

The structuralist school of thought in sociology seeks to understand society through the relationship of its members (and not the members themselves). Very simplified, we might ask:

Which are the important/influencial nodes of a given network?

A *centrality measure* is a function that assigns numbers to vertices or edges of network, where higher values indicate *central* nodes and lower values *peripheral* nodes.

Measure of centrality

The structuralist school of thought in sociology seeks to understand society through the relationship of its members (and not the members themselves).

Very simplified, we might ask:

Which are the important/influencial nodes of a given network?

Highly domain-dependant!

A *centrality measure* is a function that assigns numbers to vertices or edges of network, where higher values indicate *central* nodes and lower values *peripheral* nodes.

Centralities: a selection

$C(v)$

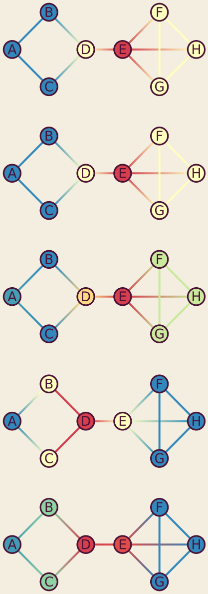
	$C(v)$
Degree	$\deg(v)$
Closeness	$\left(\sum_{u \in G} \mathrm{dist}(u,v)\right)^{-1}$
Pagerank	$\bar{v} = (\frac{1-\alpha}{n} E + \alpha M)\bar{v}$
Eccentricity	$\max_{u \in G} \left(\mathrm{dist}(u,v)\right)$
Betweenness	$\sum_{s,t \in G - v} \frac{\sigma_{st}(v)}{\sigma_{st}}$

> M Row-normalized adjacency matrix
> α Fudge factor
> E All-one matrix of size n x n

> σ_{st} # of shortest s-t-paths
> $\sigma_{st}(v)$ # of shortest s-t-paths via v

Centralities: a toy example

$$C(D) > C(E)$$



Degree

Closeness

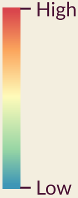
Pagerank

$$C(D) < C(E)$$

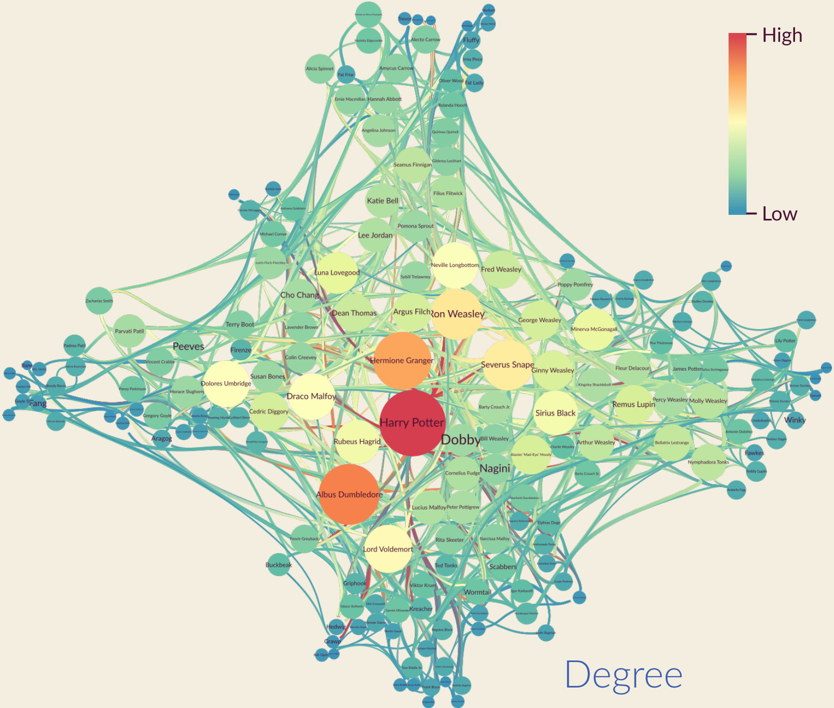
Eccentricity

$$C(D) \approx C(E)$$

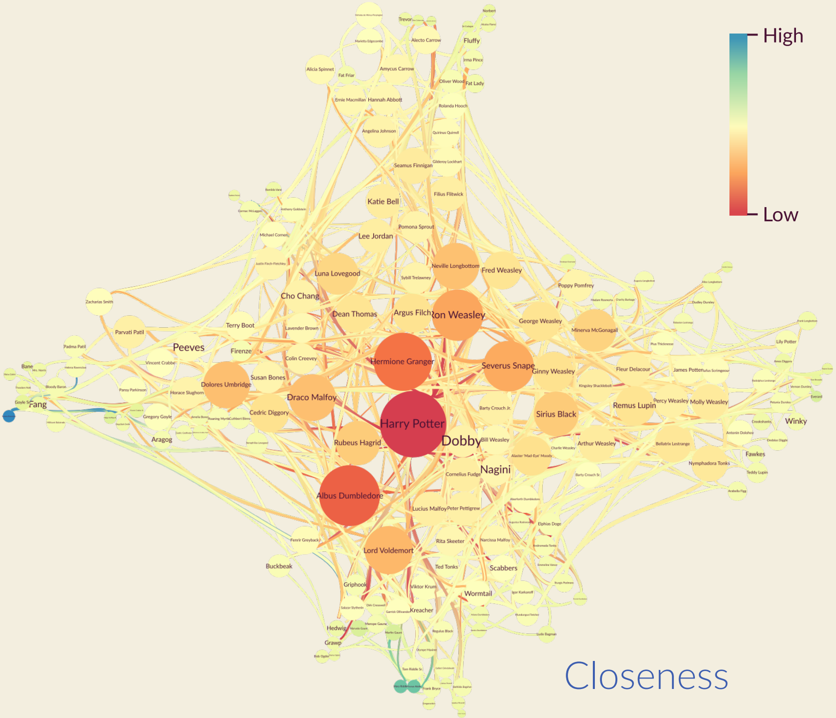
Betweenness



Harry Potter a centralita měřená stupněm uzlu

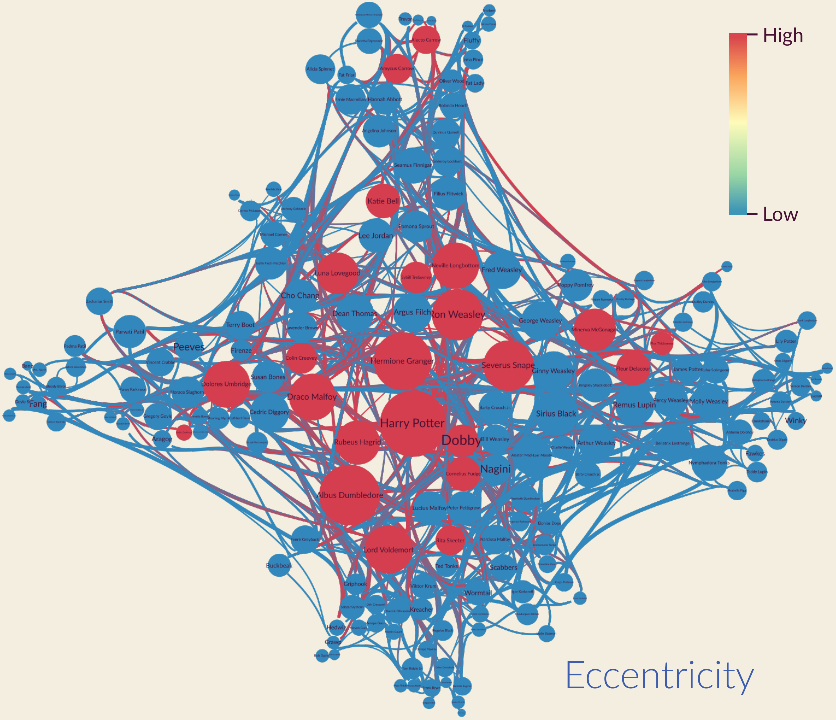


Harry Potter a blízkost polohy ve středu

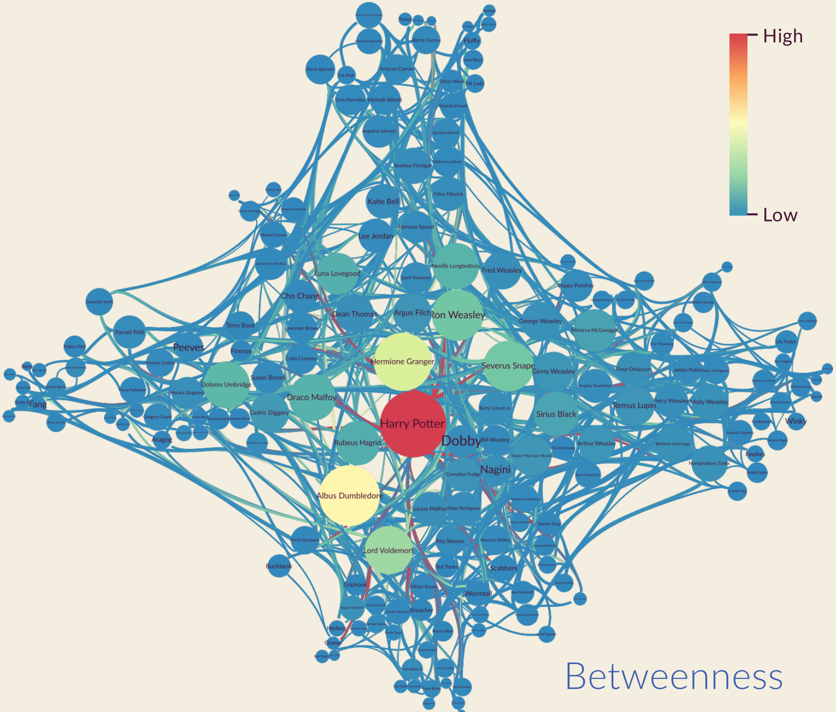


Closeness

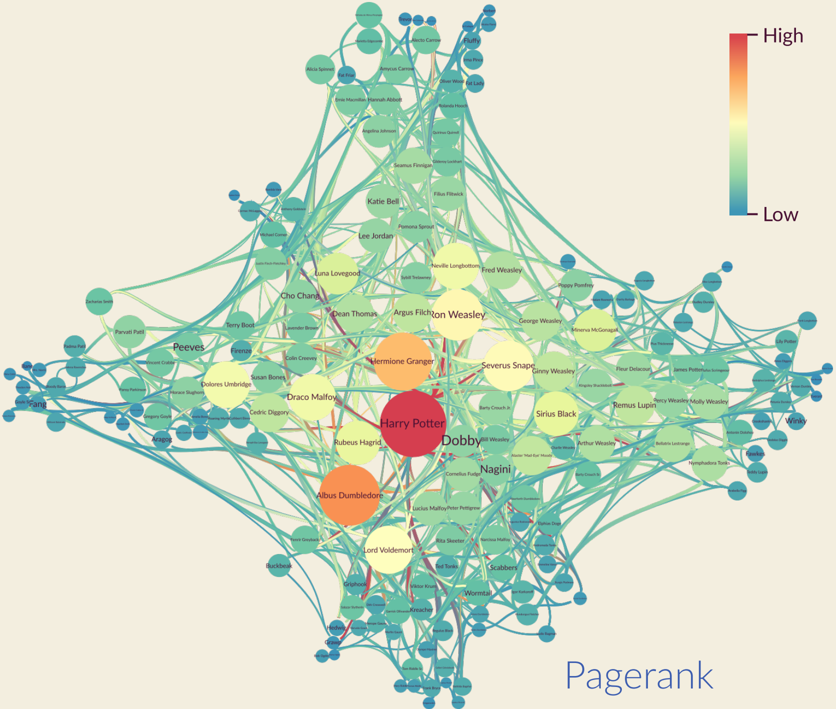
Harry Potter a výstřednost



Harry Potter a blízkost polohy ve středu

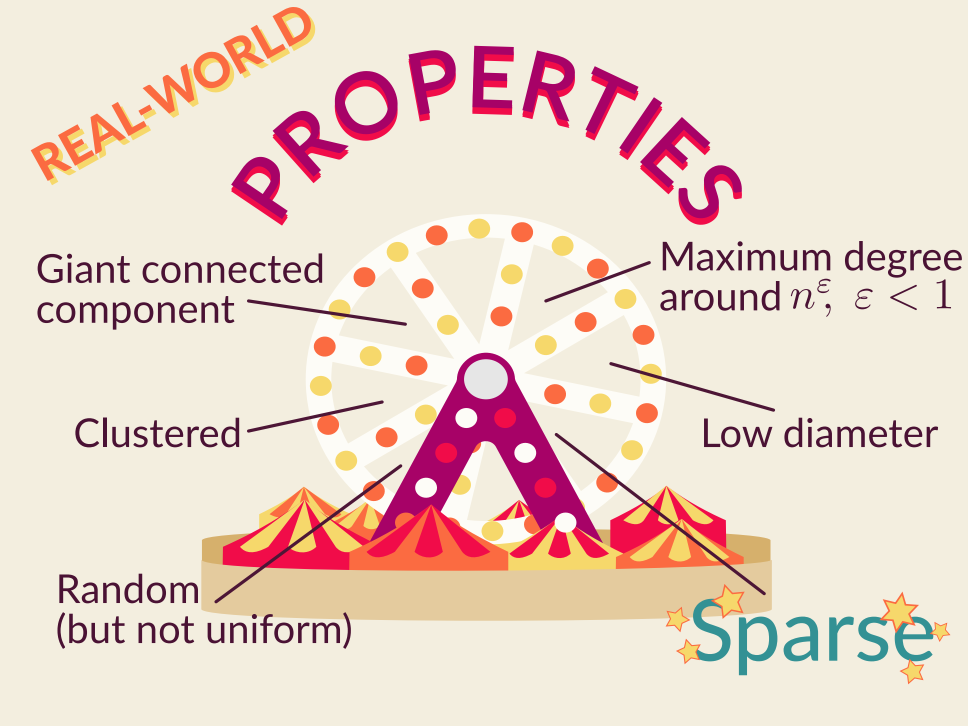


Harry Potter *a pagerank*

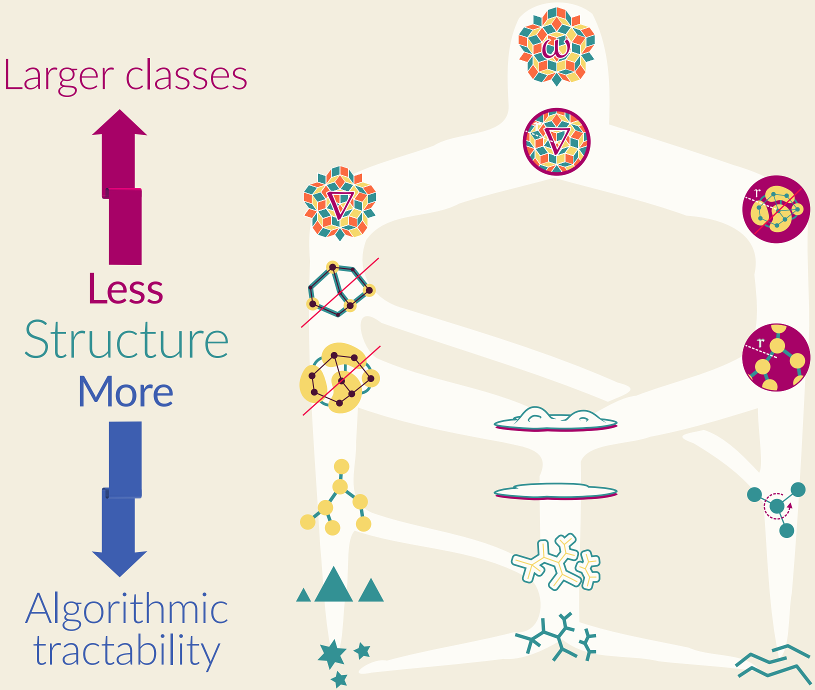


Pagerank

Why should we care?



REAL-WORLD
PROPERTIES
Giant connected component
Maximum degree around $n^{\varepsilon},\ \varepsilon < 1$
Clustered
Low diameter
Random (but not uniform)
Sparse


Larger classes
Less
Structure
More
Algorithmic
tractability