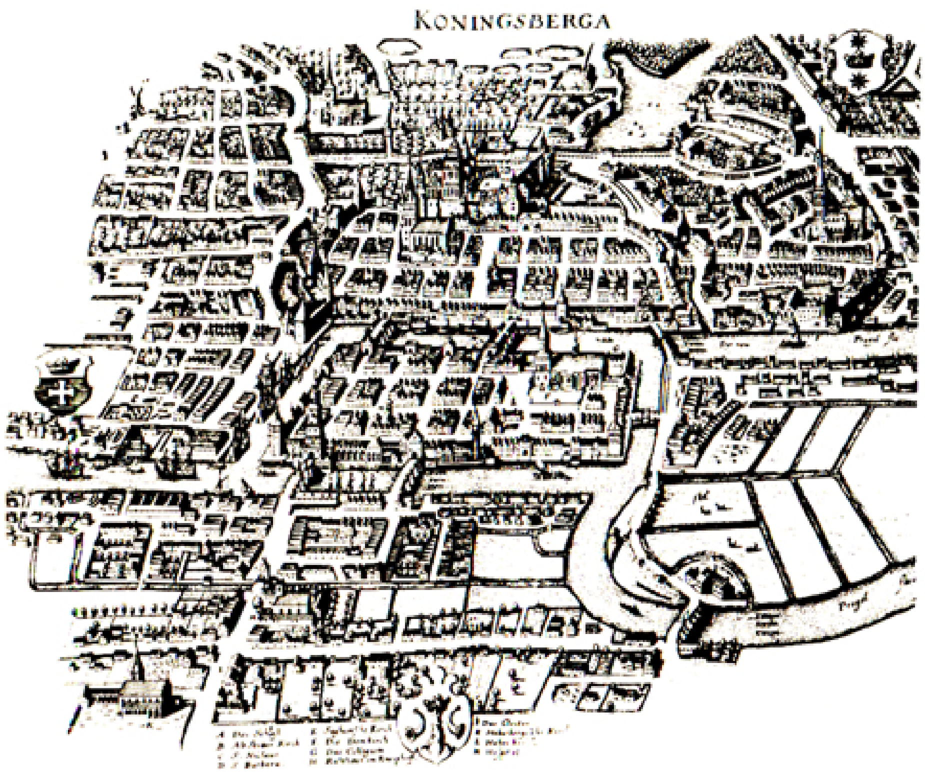


Königsberg bridges

Structure of Networks I

Vladimir Batagelj

University of Ljubljana

Networks Workshop

NICTA, Sydney, June 2005

Outline

1 Approaches to large networks 1
2 Degrees 2
6 Homomorphisms of graphs 6
7 Clusters, clusterings, partitions, hierarchies 7
10 Subgraph 10
12 Cuts 12
15 Walks 15
18 Equivalence relations and Partitions 18
20 Important vertices in network 20
29 Connectivity 29
37 k-connectivity 37
38 Triangular and short cycle connectivities 38

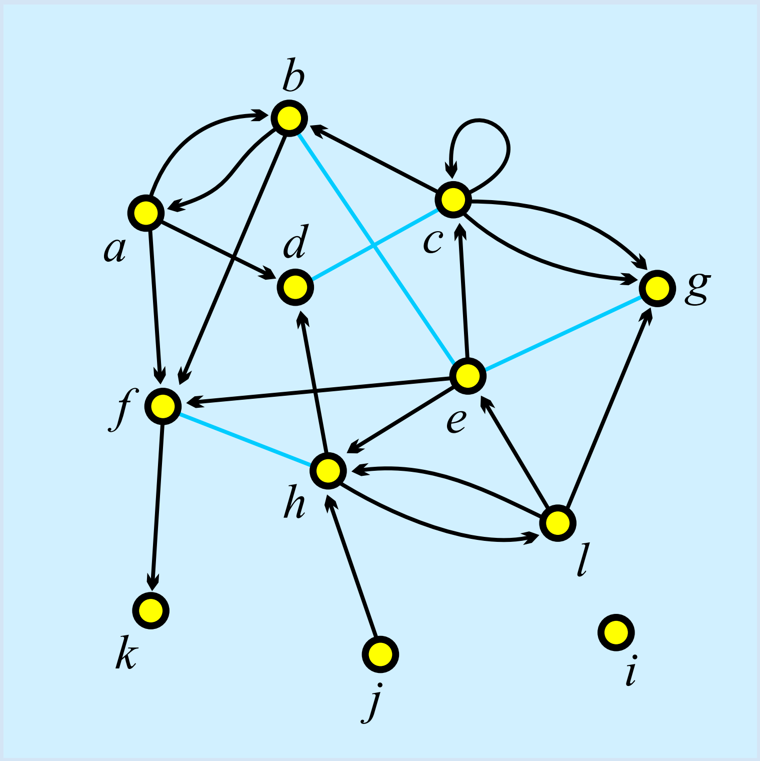
43 Islands 43
57 Dense groups 57
58 Cores and generalized cores 58

Approaches to large networks

In analysis of a *large* network (several thousands or millions of vertices, the network can be stored in computer memory) we can't display it in its totality; also there are only few algorithms available.

To analyze a large network we can use statistical approach or we can identify smaller (sub) networks that can be analyzed further using more sophisticated methods.

Degrees



degree of vertex v, $\deg(v)$ = number of lines with v as end-vertex;

indegree of vertex v, $\text{indeg}(v)$ = number of lines with v as terminal vertex (end-vertex is both initial and terminal);

outdegree of vertex v, $\text{outdeg}(v)$ = number of lines with v as initial vertex.

$$n = 12,\ m = 23,\ \text{indeg}(e) = 3,\ \text{outdeg}(e) = 5,\ \deg(e) = 6$$

$$\sum_{v\in\mathcal{V}} \text{indeg}(v) = \sum_{v\in\mathcal{V}} \text{outdeg}(v) = |\mathcal{A}| + 2|\mathcal{E}|,\ \sum_{v\in\mathcal{V}} \deg(v) = 2|\mathcal{L}| - |\mathcal{E}_0|$$

Pajek and R

Pajek 0.89 (and later) supports the use of external programs (menu `Tools`). It provides a special support for statistical program R.

In **Pajek** we determine the degrees of vertices and submit them to R

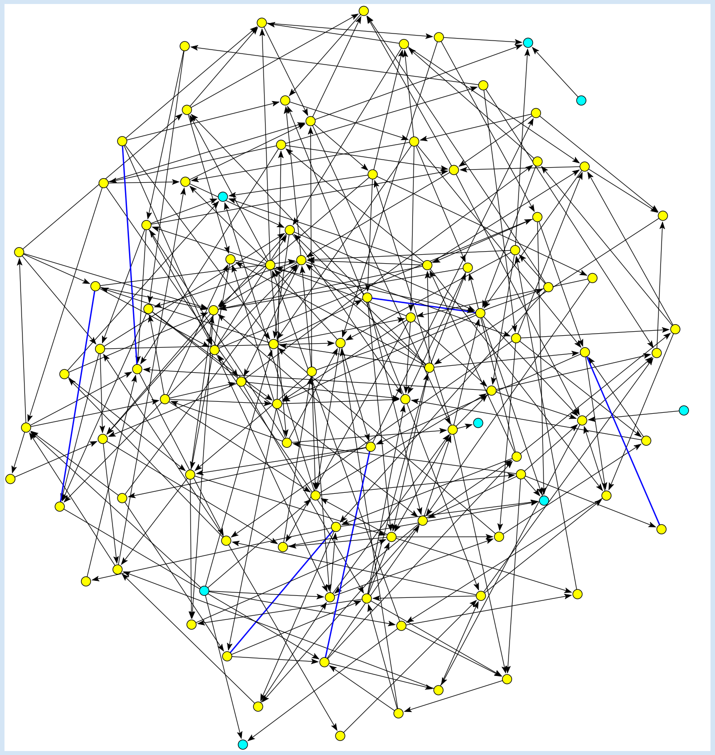
```
info/network/general
Net/Partitions/Degree/All
Partition/Make Vector
Tools/Program R/Send to R/Current Vector
```

In R we determine their distribution and plot it

```
summary(v2)
t <- tabulate(v2)
c <- t[t>0]
i <- (1:length(t))[t>0]
plot(i,c,log='xy',main='degree distribution',
  xlab='deg',ylab='freq')
```

Attention! The vertices of degree 0 are not considered by `tabulate`.

Erdős and Renyi's random graphs



Erdős and Renyi defined a *random graph* as follows: every possible line is included in a graph with a given probabilty p.

In **Pajek**'s

`Net/Random Network/Erdos-Renyi`

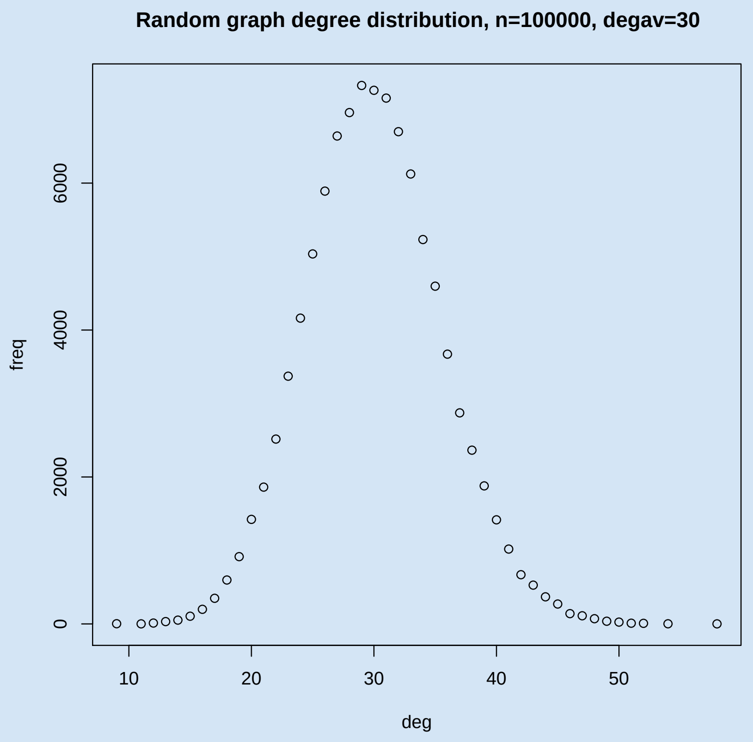
instead of probability p a more intuitive average degree is used

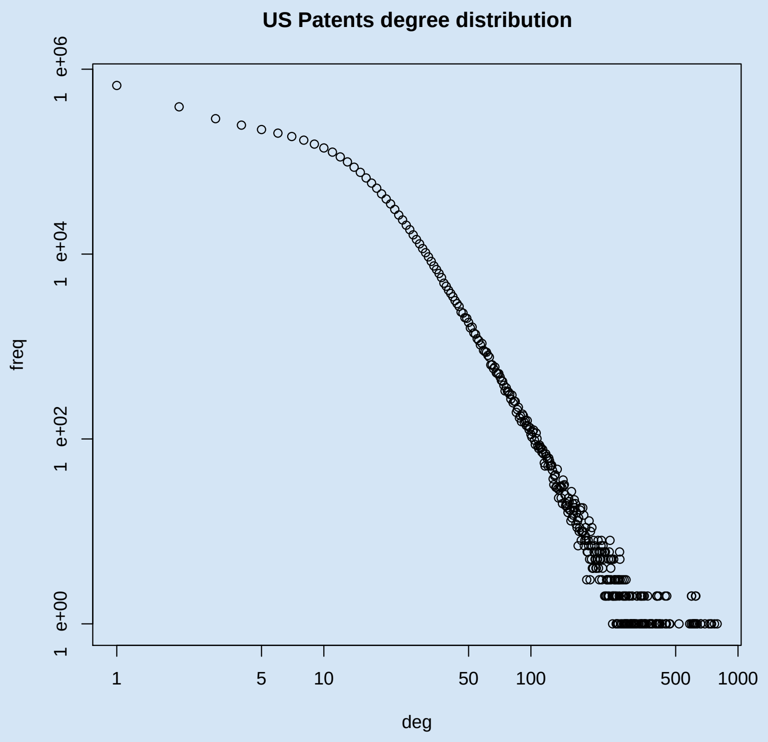
$$\overline{\deg} = \frac{1}{n} \sum_{v \in \mathcal{V}} \deg(v)$$

It holds $p = \frac{m}{m_{max}}$ and, for simple graphs, also $\overline{\deg} = \frac{2m}{n}$.

Random graph in picture has 100 vertices and average degree 3.

Degree distribution





Real-life networks are usually not random in the Erdős/Renyi sense. The analysis of their distributions gave a new view about their structure – Watts (Small worlds), Barabási (nd/networks, Linked).

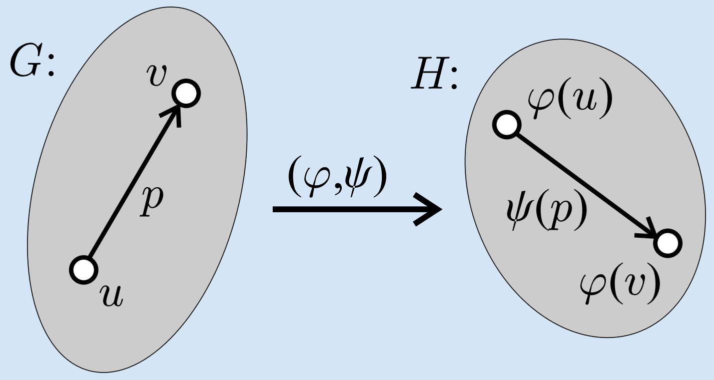
Homomorphisms of graphs

Functions (φ, ψ), $\varphi\colon \mathcal{V} \to \mathcal{V}'$ and $\psi\colon \mathcal{L} \to \mathcal{L}'$ determine a *weak homomorphism* of graph $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ in graph $\mathcal{H} = (\mathcal{V}', \mathcal{L}')$ iff:

$$\forall u, v \in \mathcal{V}\, \forall p \in \mathcal{L} : (p(u : v) \Rightarrow \psi(p)(\varphi(u) : \varphi(v)))$$

and they determine a *(strong) homomorphism* of graph $\mathcal{G}$ in graph $\mathcal{H}$ iff:

$$\forall u, v \in \mathcal{V}\, \forall p \in \mathcal{L} : (p(u, v) \Rightarrow \psi(p)(\varphi(u), \varphi(v)))$$



If φ and ψ are bijections and the condition hold in both direction we get an *isomorphism* of graphs $\mathcal{G}$ and $\mathcal{H}$. We denote the weak isomorphism by $\mathcal{G} \sim \mathcal{H}$; and the (strong) isomorphism by $\mathcal{G} \approx \mathcal{H}$. Itholds $\approx \subset \sim$.

An *invariant* of graph is called each graph characteristic that has the same value for all isomorphic graphs.

EulerGT

Clusters, clusterings, partitions, hierarchies

A nonempty subset $C \subseteq \mathcal{V}$ is called a *cluster* (group). A nonempty set of clusters $\mathbf{C} = \{C_i\}$ forms a *clustering*.

Clustering $\mathbf{C} = \{C_i\}$ is a *partition* iff

$$\cup\mathbf{C} = \bigcup_i C_i = \mathcal{V} \quad \text{in} \quad i \neq j \Rightarrow C_i \cap C_j = \emptyset$$

Clustering $\mathbf{C} = \{C_i\}$ is a *hierarchy* iff

$$C_i \cap C_j \in \{\emptyset, C_i, C_j\}$$

Hierarchy $\mathbf{C} = \{C_i\}$ is *complete*, iff $\cup\mathbf{C} = \mathcal{V}$; and is *basic* if for all $v \in \cup\mathbf{C}$ also $\{v\} \in \mathbf{C}$.

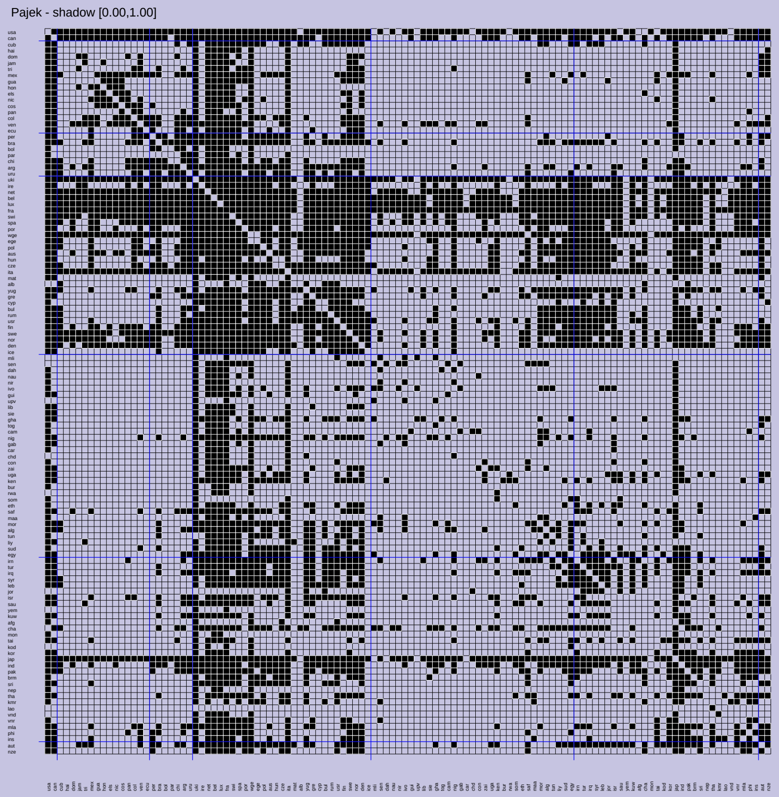
Contraction of cluster

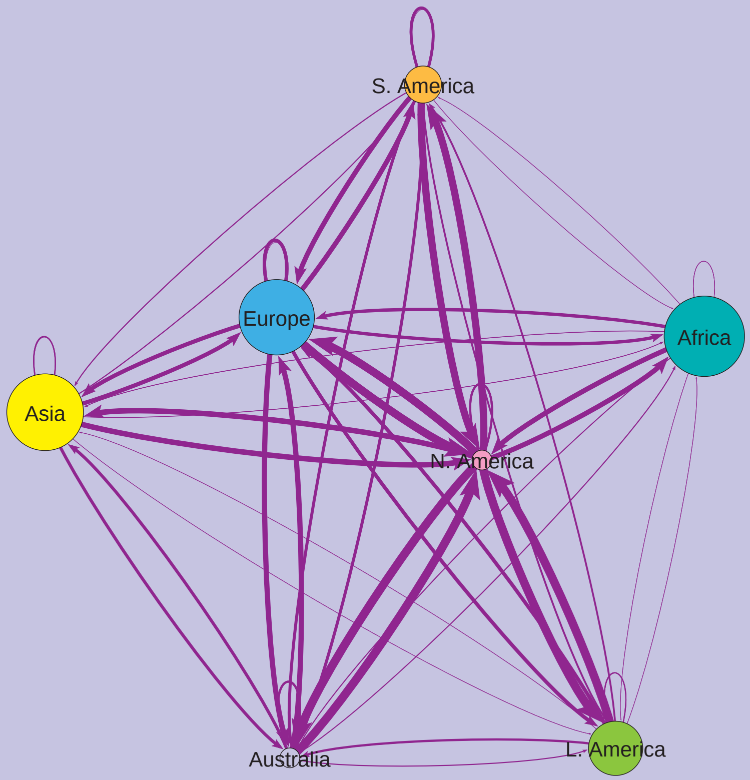
Contraction of cluster C is called a graph $\mathcal{G}/C$, in which all vertices of the cluster C are replaced by a single vertex, say c. More precisely:

$\mathcal{G}/C = (\mathcal{V}', \mathcal{L}')$, where $\mathcal{V}' = (\mathcal{V} \setminus C) \cup \{c\}$ and $\mathcal{L}'$ consists of lines from $\mathcal{L}$ that have both end-vertices in $\mathcal{V} \setminus C$. Beside these it contains also a 'star' with the center c and: arc (v, c), if $\exists p \in \mathcal{L}, u \in C : p(v, u)$; or arc (c, v), if $\exists p \in \mathcal{L}, u \in C : p(u, v)$. There is a loop (c, c) in c if $\exists p \in \mathcal{L}, u, v \in C : p(u, v)$.

In a network over graph $\mathcal{G}$ we have also to specify how are determined the values/weights in the shrunk part of the network. Usually as the sum or maksimum/minimum of the original values.

Contracted clusters – international trade

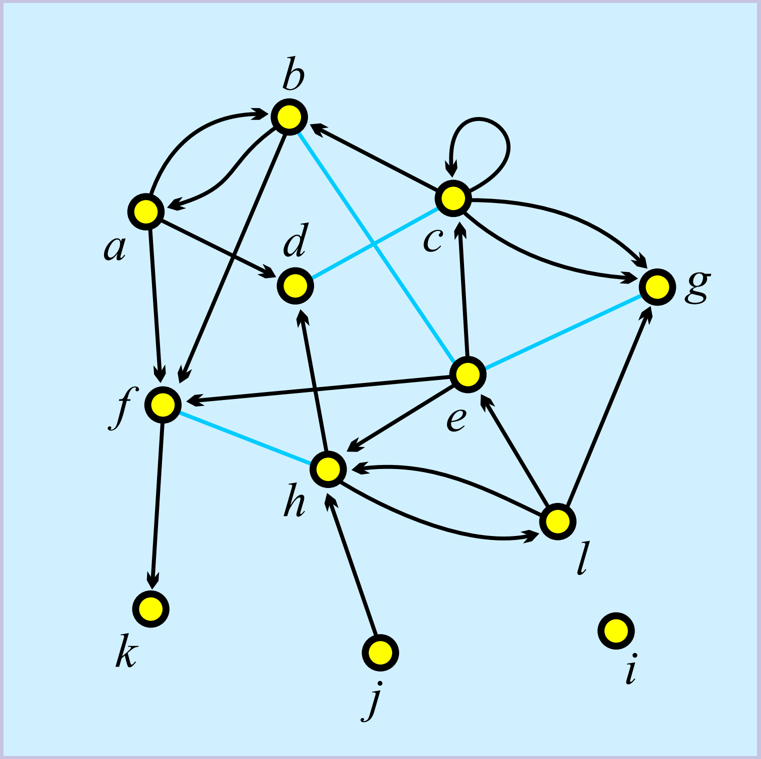


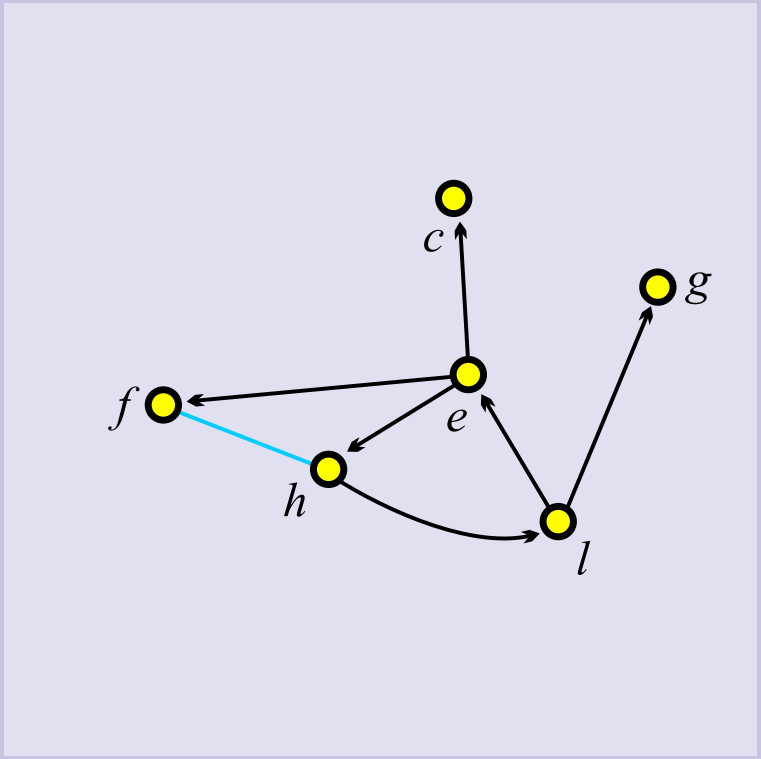


Snyder and Kick's international trade. Matrix display of dense networks.

$$w(C_i, C_j) = \frac{n(C_i, C_j)}{n(C_i) \cdot n(C_j)}$$

Subgraph

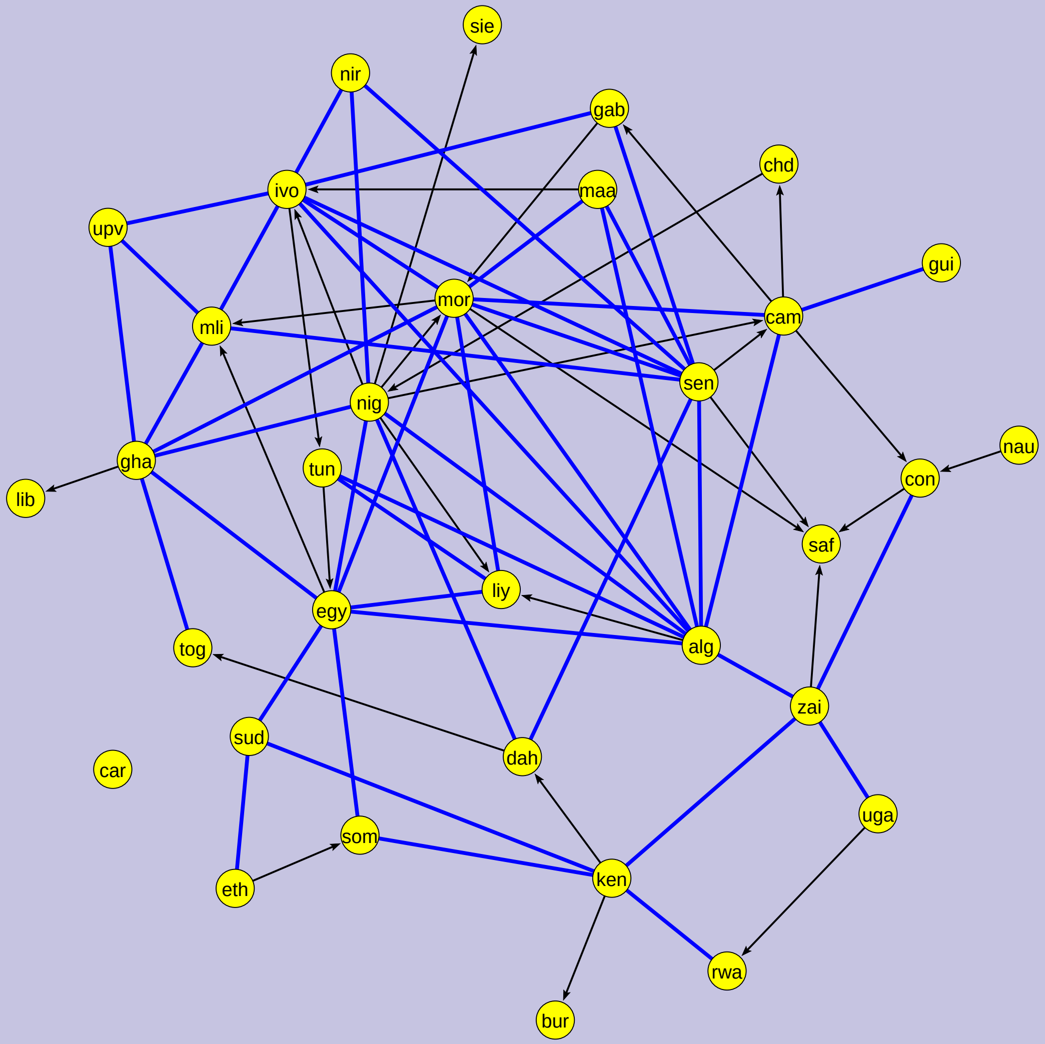




A *subgraph* $\mathcal{H} = (\mathcal{V}', \mathcal{L}')$ of a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ is a graph which set of lines is a subset of set of lines of $\mathcal{G}$, $\mathcal{L}' \subseteq \mathcal{L}$, its vertex set is a subset of set of vertices of $\mathcal{G}$, $\mathcal{V}' \subseteq \mathcal{V}$, and it contains all end-vertices of $\mathcal{L}'$.

A subgraph can be *induced* by a given subset of vertices or lines.

Cut-out – induced subgraph: Snyder and Kick – Africa



Cuts

The standard approach to find interesting groups inside a network was based on properties/weights – they can be *measured* or *computed* from network structure (for example Kleinberg's hubs and authorities).

The *vertex-cut* of a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \to \mathbb{R}$, at selected level t is a subnetwork $\mathcal{N}(t) = (\mathcal{V}', \mathcal{L}(\mathcal{V}'), p)$, determined by the set

$$\mathcal{V}' = \{v \in \mathcal{V} : p(v) \geq t\}$$

and $\mathcal{L}(\mathcal{V}')$ is the set of lines from $\mathcal{L}$ that have both endpoints in $\mathcal{V}'$.

The *line-cut* of a network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \to \mathbb{R}$, at selected level t is a subnetwork $\mathcal{N}(t) = (\mathcal{V}(\mathcal{L}'), \mathcal{L}', w)$, determined by the set

$$\mathcal{L}' = \{e \in \mathcal{L} : w(e) \geq t\}$$

and $\mathcal{V}(\mathcal{L}')$ is the set of all endpoints of the lines from $\mathcal{L}'$.

Simple analysis using cuts

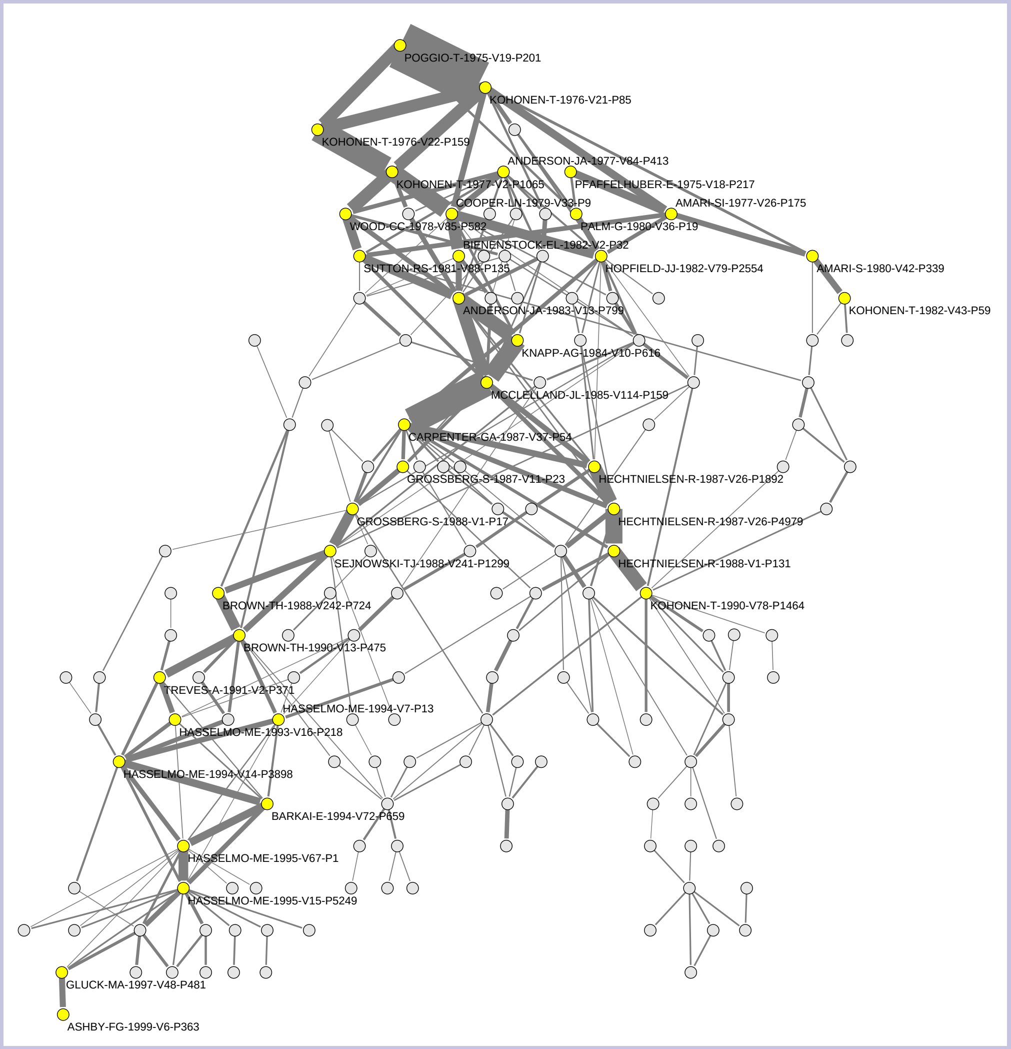
We look at the components of $\mathcal{N}(t)$.

Their number and sizes depend on t. Usually there are many small components. Often we consider only components of size at least k and not exceeding K. The components of size smaller than k are discarded as 'noninteresting'; and the components of size larger than K are cut again at some higher level.

The values of thresholds t, k and K are determined by inspecting the distribution of vertex/arc-values and the distribution of component sizes and considering additional knowledge on the nature of network or goals of analysis.

We developed some new and efficiently computable properties/weights.

Citation weights

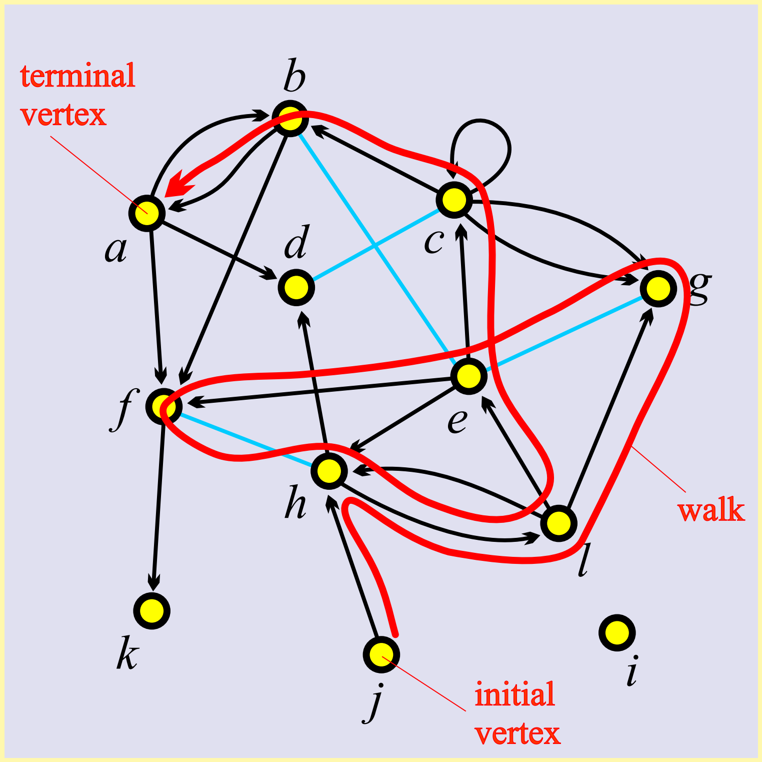


The citation network analysis started in 1964 with the paper of Garfield et al. In 1989 Hummon and Doreian proposed three indices – weights of arcs that are proportional to the number of different source-sink paths passing through the arc. We developed algorithms to efficiently compute these indices.

Main subnetwork (arc cut at level 0.007) of the SOM (selforganizing maps) citation network (4470 vertices, 12731 arcs).

See paper.

Walks



length $|s|$ of the walk s is the number of lines it contains.

$s = (j, h, l, g, e, f, h, l, e, c, b, a)$

$|s| = 11$

A walk is *closed* iff its initial and terminal vertex coincide.

If we don't consider the direction of the lines in the walk we get a *semiwalk* or *chain*.

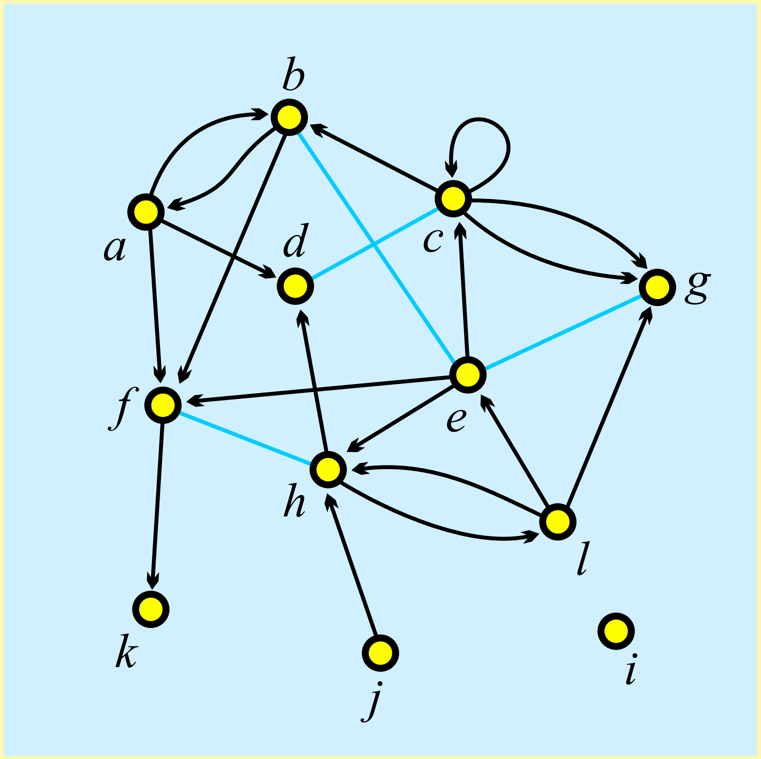
trail – walk with all lines different

path – walk with all vertices different

cycle – closed walk with all internal vertices different

A graph is *acyclic* if it doesn't contain any cycle.

Shortest paths



A shortest path from u to v is also called a *geodesic* from u to v. Its length is denoted by $d(u, v)$.

If there is no walk from u to v then $d(u, v) = \infty$.

$d(j, a) = |(j, h, d, c, b, a)| = 5$

$d(a, j) = \infty$

$\hat{d}(u, v) = \max(d(u, v), d(v, u))$

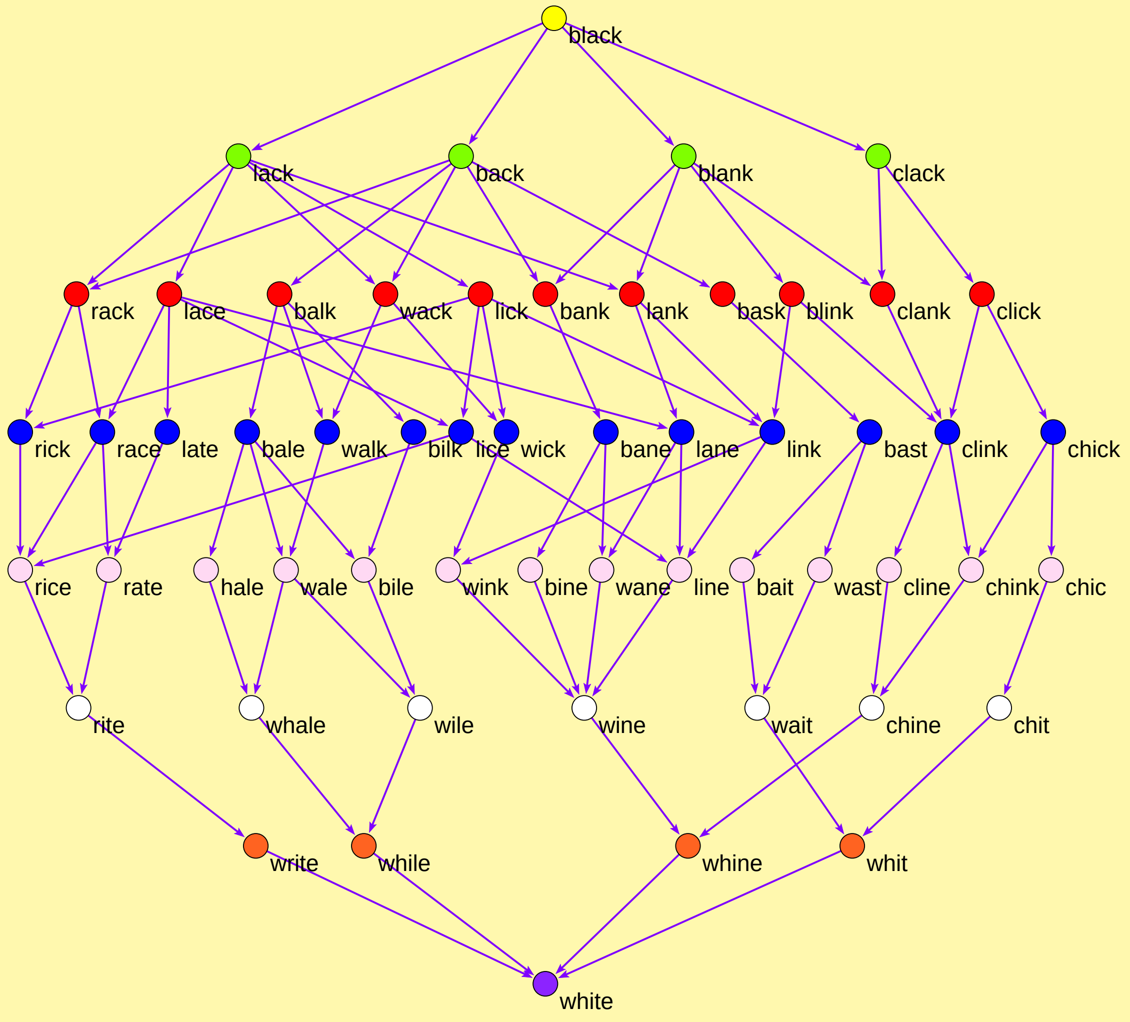
is a *distance*:

$\hat{d}(v, v) = 0$, $\hat{d}(u, v) = \hat{d}(v, u)$,

$\hat{d}(u, v) \le \hat{d}(u, t) + \hat{d}(t, v)$.

The *diameter* of a graph equals to the distance between the most distant pair of vertices: $D = \max_{u,v \in \mathcal{V}} d(u, v)$.

Shortest paths



DICT28.

Equivalence relations and Partitions

A relation R on $\mathcal{V}$ is an *equivalence* relation iff it is reflexive $\forall v \in \mathcal{V} : vRv$, symmetric $\forall u, v \in \mathcal{V} : uRv \Rightarrow vRu$, and transitive $\forall u, v, z \in \mathcal{V} : uRz \wedge zRv \Rightarrow uRv$.

Each equivalence relation determines a partition into *equivalence classes* $[v] = \{u : vRu\}$.

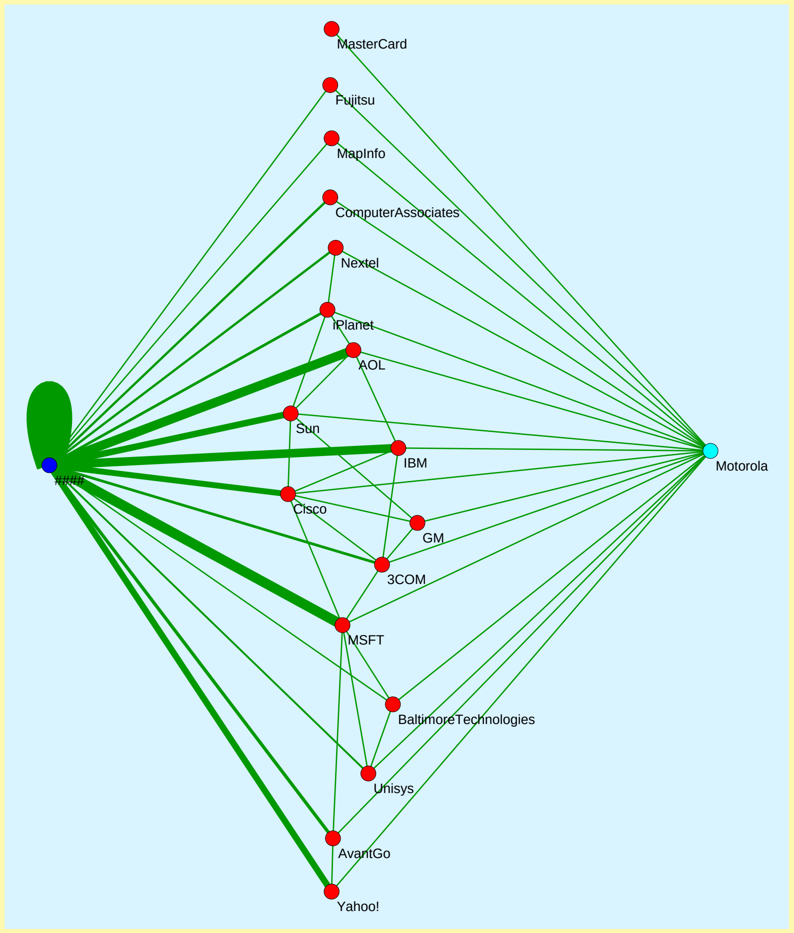
Each partition $\mathbf{C}$ determines an equivalence relation $uRv \Leftrightarrow \exists C \in \mathbf{C} : u \in C \wedge v \in C$.

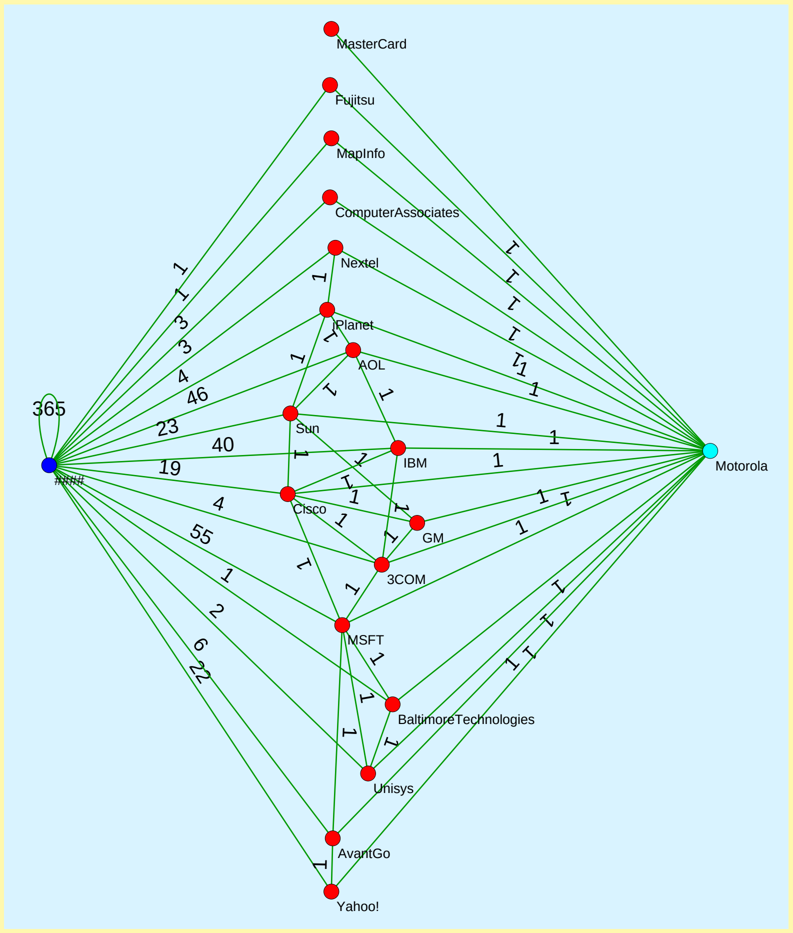
k-neighbors of v is the set of vertices on 'distance' k from v, $N^k(v) = \{u \in v : d(v, u) = k\}$.

The set of all k-neighbors, $k = 0, 1, \ldots$ of v is a partition of $\mathcal{V}$.

k-neighborhood of v, $N^{(k)}(v) = \{u \in v : d(v, u) \leq k\}$.

Motorola’s neighborhood





The thickness of edges is a square root of its value.

Important vertices in network

It seems that the most important distinction between different vertex *indices* is based on the view/decision whether the network is considered directed or undirected. This gives us two main types of indices:

- directed case: measures of *importance*; with two subgroups: measures of *influence*, based on out-going arcs; and measures of *support*, based on incoming arcs;
- undirected case: measures of *centrality*, based on all lines.

For undirected networks all three types of measures coincide.

If we change the direction of all arcs (replace the relation with its inverse relation) the measure of influence becomes a measure of support, and vice versa.

. . . Important vertices in network

The real meaning of measure of importance depends on the relation described by a network. For example the most 'important' person for the relation '__ doesn't like to work with __' is in fact the least popular person.

Removal of an important vertex from a network produces a substantial change in structure/functioning of the network.

Normalization

Let $p : \mathcal{V} \to \mathbb{R}$ be an index in network $\mathcal{N} = (\mathcal{V}, \mathcal{L})$. If we want to compare indices p over different networks we have to make them comparable. Usually we try to achieve this by *normalization* of p.

Let $\mathcal{N} \in \mathbf{N}(\mathcal{V})$, where $\mathbf{N}(\mathcal{V})$ is a selected set of networks over the same set of vertices V,

$$p_{max} = \max_{\mathcal{N}\in\mathbf{N}(\mathcal{V})} \max_{v\in\mathcal{V}} p_{\mathcal{N}}(v) \qquad \text{and} \qquad p_{min} = \min_{\mathcal{N}\in\mathbf{N}(\mathcal{V})} \min_{v\in\mathcal{V}} p_{\mathcal{N}}(v)$$

then we define the normalized index as

$$p'(v) = \frac{p(v) - p_{min}}{p_{max} - p_{min}} \in [0, 1]$$

Degrees

The simplest index are the degrees of vertices. Since for simple networks $\deg_{min} = 0$ and $\deg_{max} = n - 1$, the corresponding normalized indices are

centrality $$\deg'(v) = \frac{\deg(v)}{n-1}$$

and similary

support $$\mathrm{indeg}'(v) = \frac{\mathrm{indeg}(v)}{n}$$

influence $$\mathrm{outdeg}'(v) = \frac{\mathrm{outdeg}(v)}{n}$$

Instead of degrees in original network we can consider also the degrees with respect to the reachability relation (transitive closure).

Closeness

Most indices are based on the distance $d(u, v)$ between vertices in a network $\mathcal{N} = (\mathcal{V}, \mathcal{L})$. Two such indices are

radius $r(v) = \max_{u \in \mathcal{V}} d(v, u)$

total closeness $S(v) = \sum_{u \in \mathcal{V}} d(v, u)$

These two indices are measures of influence – to get measures of support we have to replace in definitions $d(u, v)$ with $d(v, u)$.

If the network is not strongly connected r_{max} and S_{max} equal ∞. Sabidussi (1966) introduced a related measure $1/S(v)$; or in its normalized form

closeness $$cl(v) = \frac{n-1}{\sum_{u \in \mathcal{V}} d(v, u)}$$

$D = \max_{u,v \in \mathcal{V}} d(v, u)$ is called the *diameter* of network.

Betweeness

Important are also the vertices that can control the information flow in the network. If we assume that this flow uses only the shortest paths (geodesics) we get a measure of *betweeness* (Anthonisse 1971, Freeman 1977)

$$b(v) = \frac{1}{(n-1)(n-2)} \sum_{\substack{u,t \in \mathcal{V}: g_{u,t}>0 \\ u\neq v, t\neq v, u\neq t}} \frac{g_{u,t}(v)}{g_{u,t}}$$

where $g_{u,t}$ is the number of geodesics from u to t; and $g_{u,t}(v)$ is the number of those among them that pass through vertex v.

If we know matrices $[d_{u,v}]$ and $[g_{u,v}]$ we can determine also $g_{u,v}(t)$ by:

$$g_{u,v}(t) = \begin{cases} g_{u,t} \cdot g_{t,v} & d_{u,t} + d_{t,v} = d_{u,v} \\ 0 & \text{otherwise} \end{cases}$$

For computation of geodesic matrix see Brandes.

Hubs and authorities

To each vertex v of a network $\mathcal{N} = (\mathcal{V}, \mathcal{L})$ we assign two values: quality of its content (*authority*) x_v and quality of its references (*hub*) y_v.

A good authority is selected by good hubs; and good hub points to good authorities (see Klienberg).

$$x_v = \sum_{u:(u,v)\in\mathcal{L}} y_u \qquad \text{and} \qquad y_v = \sum_{u:(v,u)\in\mathcal{L}} x_u$$

Let $\mathbf{W}$ be a matrix of network $\mathcal{N}$ and $\mathbf{x}$ and $\mathbf{y}$ authority and hub vectors. Then we can write these two relations as $\mathbf{x} = \mathbf{W}^T\mathbf{y}$ and $\mathbf{y} = \mathbf{W}\mathbf{x}$.

We start with $\mathbf{y} = [1, 1, \ldots, 1]$ and then compute new vectors $\mathbf{x}$ and $\mathbf{y}$. After each step we normalize both vectors. We repeat this until they stabilize.

We can show that this procedure converges. The limit vector $\mathbf{x}^*$ is the principal eigen vector of matrix $\mathbf{W}^T\mathbf{W}$; and $\mathbf{y}^*$ of matrix $\mathbf{W}\mathbf{W}^T$.

...Hubs and authorities

Similar procedures are used in search engines on the web to evaluate the importance of web pages.

PageRank, PageRank / Google, HITS / AltaVista, SALSA, teorija.

Examples: Krebs, Krempl.

Clustering

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be simple undirected graph. *Clustering* in vertex v is usually measured as a quotient between the number of lines in subgraph $\mathcal{G}^1(v) = \mathcal{G}(N^1(v))$ induced by the neighbors of vertex v and the number of lines in the complete graph on these vertices:

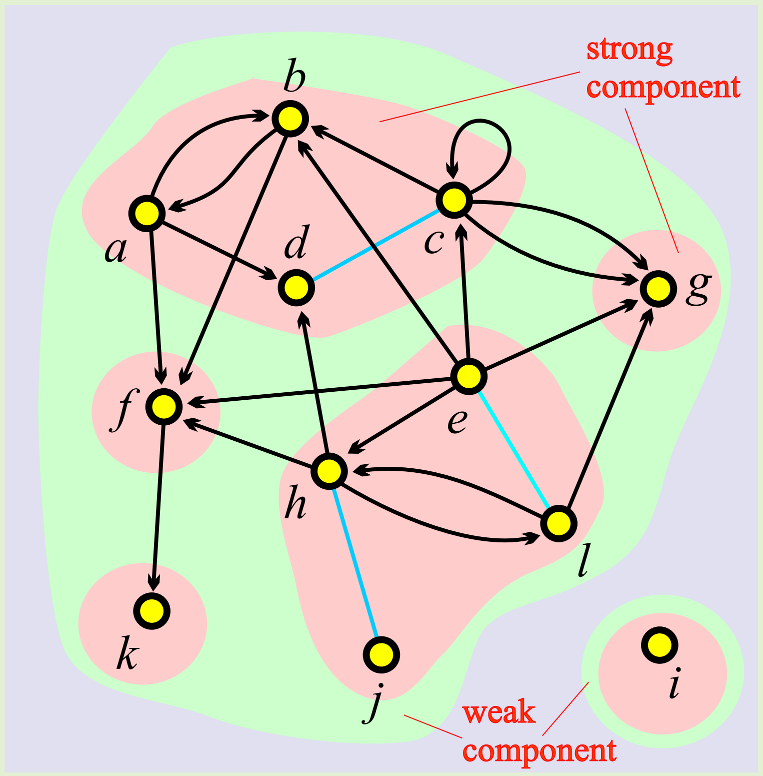
$$C(v) = \begin{cases} \dfrac{2|\mathcal{L}(\mathcal{G}^1(v))|}{\deg(v)(\deg(v)-1)} & \deg(v) > 1 \\ 0 & \text{otherwise} \end{cases}$$

We can consider also the size of vertex neighborhood by the following correction

$$C_1(v) = \frac{\deg(v)}{\Delta} C(v)$$

where Δ is the maximum degree in graph $\mathcal{G}$. This measure attains its largest value in vertices that belong to an isolated clique of size Δ.

Connectivity



Vertex u is *reachable* from vertex v iff there exists a walk with initial vertex v and terminal vertex u.

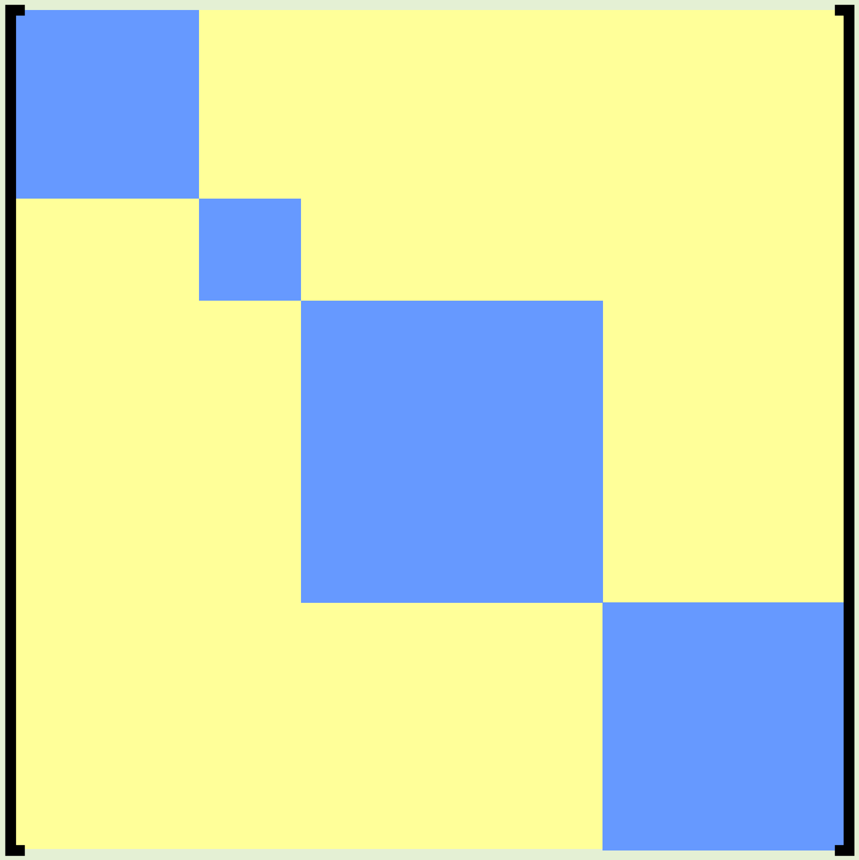
Vertex v is *weakly connected* with vertex u iff there exists a semiwalk with v and u as its end-vertices.

Vertex v is *strongly connected* with vertex u iff they are mutually reachable.

Weak and strong connectivity are equivalence relations.

Equivalence classes induce weak/strong *components*.

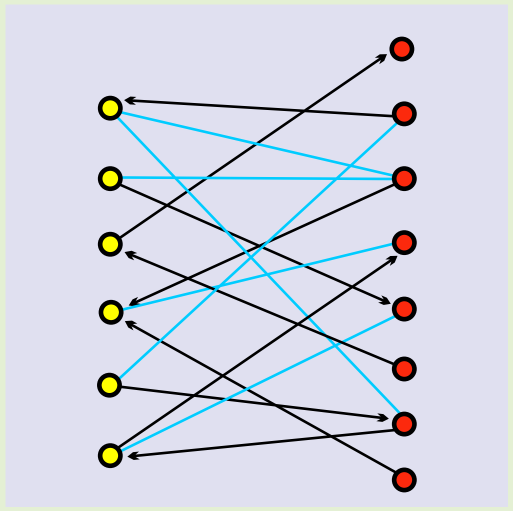
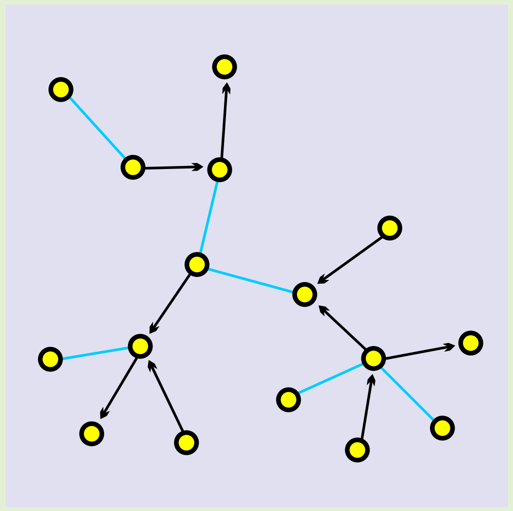
Weak components



Reordering the vertices of network such that the vertices from the same class of weak partition are put together we get a matrix representation consisting of diagonal blocks – weak components.

Most problems can be solved separately on each component and afterward these solutions combined into final solution.

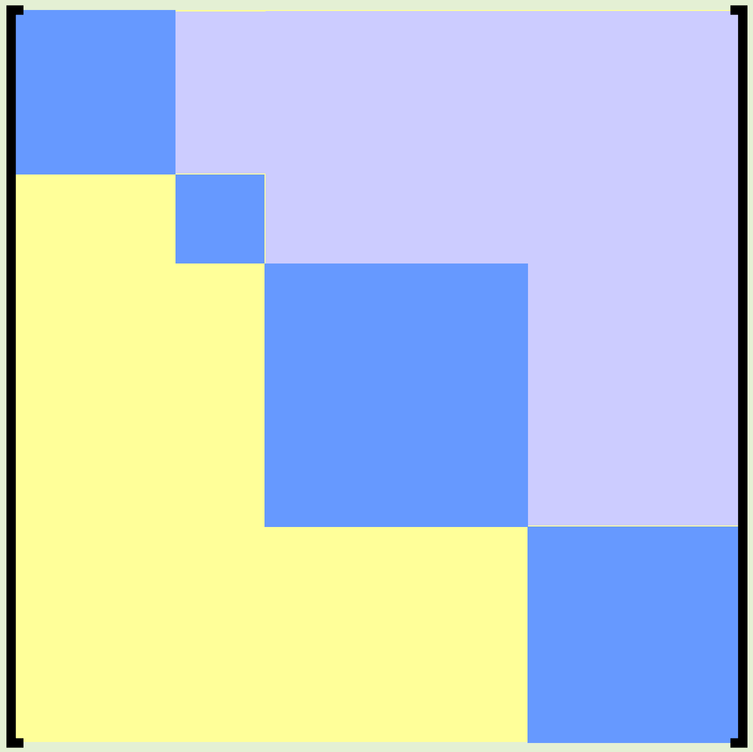
Special graphs – bipartite, tree

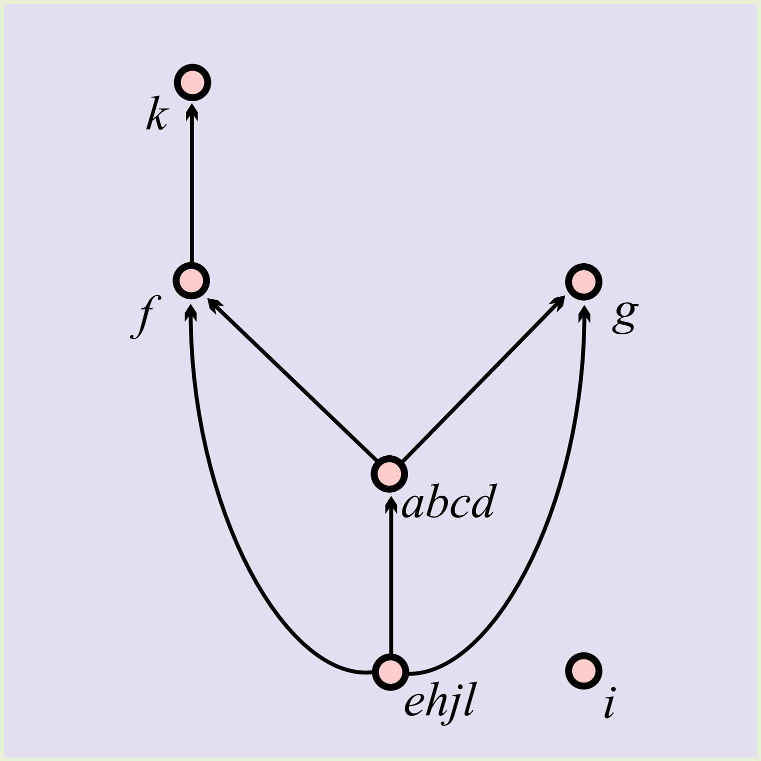



A graph $\mathcal{G} = (\mathcal{V}, \mathcal{L})$ is *bipartite* iff its set of vertices $\mathcal{V}$ can be partitioned into two sets $\mathcal{V}_1$ and $\mathcal{V}_2$ such that every line from $\mathcal{L}$ has one end-vertex in $\mathcal{V}_1$ and the other in $\mathcal{V}_2$.

A weakly connected graph $\mathcal{G}$ is a *tree* iff it doesn't contain loops and semicycles of length at least 3.

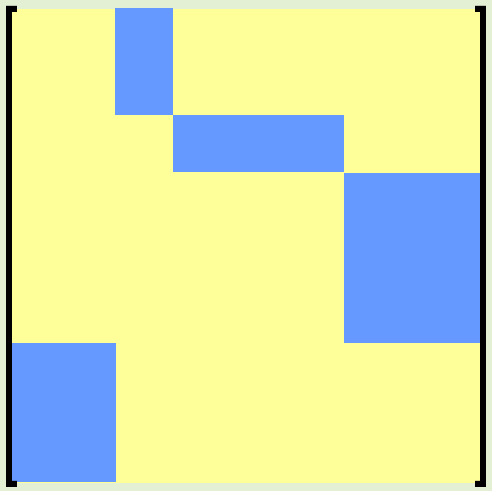
Reduction (condensation)

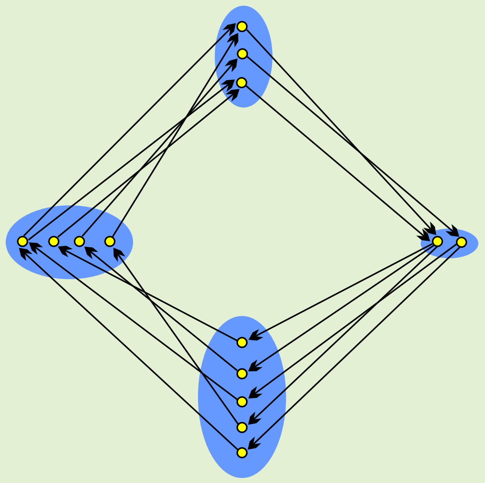




If we shrink every strong component of a given graph into a vertex, delete all loops and identify parallel arcs the obtained *reduced* graph is acyclic. For every acyclic graph an *ordering* / *level* function $i : \mathcal{V} \to \mathbb{N}$ exists s.t. $(u, v) \in \mathcal{A} \Rightarrow i(u) < i(v)$.

. . . internal structure of strong components



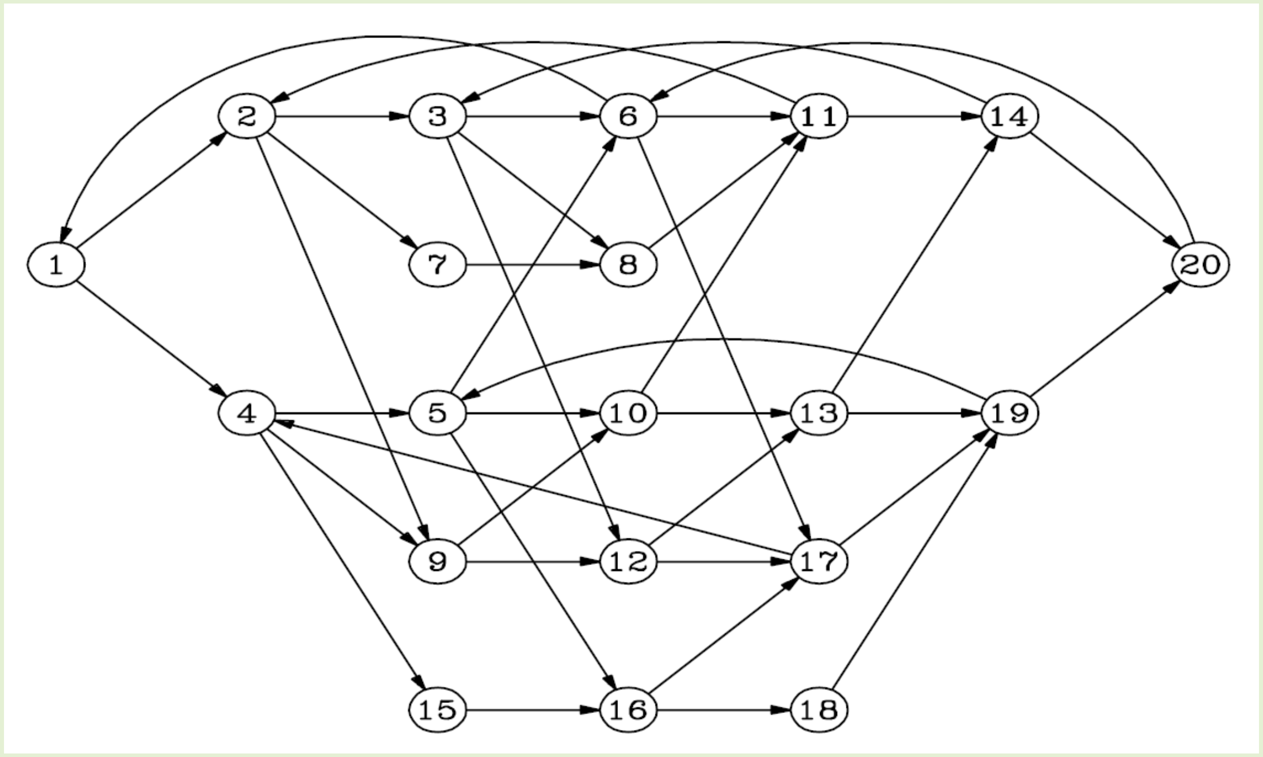


Let d be the largest common divisor of lengths of closed walks in a strong component.

The component is said to be *simple*, iff $d = 1$; otherwise it is *periodical* with a period d.

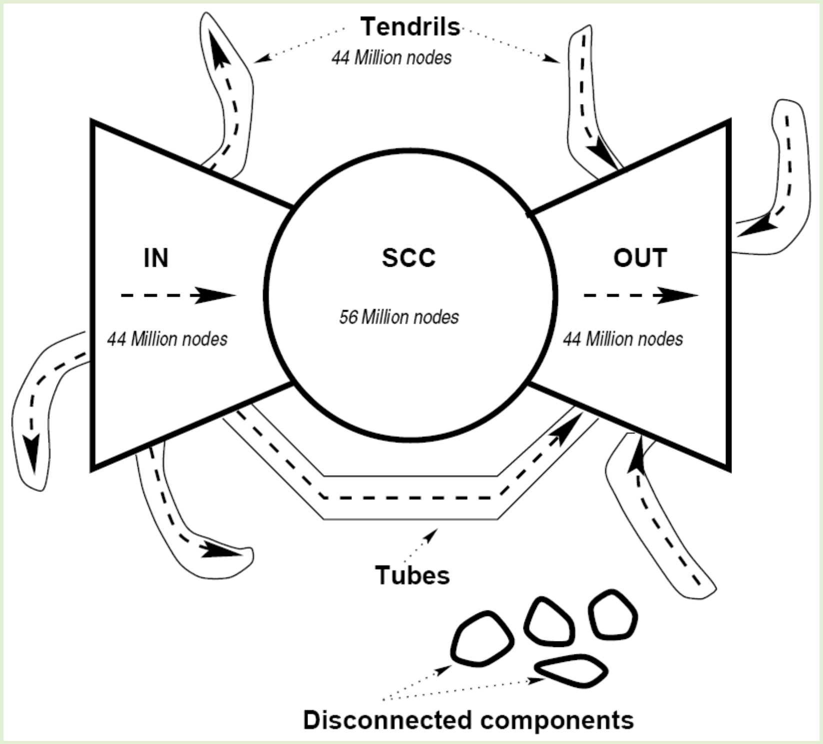
The set of vertices $\mathcal{V}$ of strongly connected directed graph $\mathcal{G} = (\mathcal{V}, R)$ can be partitioned into d clusters $\mathcal{V}_1, \mathcal{V}_2, \ldots, \mathcal{V}_d$, s.t. for every arc $(u, v) \in R$ holds $u \in \mathcal{V}_i \Rightarrow v \in \mathcal{V}_{(i \bmod d)+1}$.

…internal structure of strong components



Bonhoure's periodical graph. Pajek data

Bow-tie structure of the Web graph



Kumar &: The Web as a graph

Let $\mathcal{S}$ be the *largest strong component* in network $\mathcal{N}$; $\mathcal{W}$ the weak component containing $\mathcal{S}$; $\mathcal{I}$ the set of vertices from which $\mathcal{S}$ can be reached; $\mathcal{O}$ the set of vertices reachable from $\mathcal{S}$; $\mathcal{T}$ (tubes) set of vertices (not in $\mathcal{S}$) on paths from $\mathcal{I}$ to $\mathcal{O}$; $\mathcal{R} = \mathcal{W} \setminus (\mathcal{I} \cup \mathcal{S} \cup \mathcal{O} \cup \mathcal{T})$ (tendrils); and $\mathcal{D} = \mathcal{V} \setminus \mathcal{W}$. The partition

$$\{\mathcal{I}, \mathcal{S}, \mathcal{O}, \mathcal{T}, \mathcal{R}, \mathcal{D}\}$$

is called the *bow-tie* partition of $\mathcal{V}$.

Biconnectivity

Vertices u and v are *biconnected* iff they are connected (in both directions) by two independent (no common internal vertex) paths.

Biconnectivity determines a partition of the set of lines.

A vertex is an *articulation* vertex iff its deletion increases the number of weak components in a graph.

A line is a *bridge* iff its deletion increases the number of weak components in a graph.

k-connectivity

Vertex connectivity κ of graph $\mathcal{G}$ is equal to the smallest number of vertices that, if deleted, induce a disconnected graph or the trivial graph K_1.

Line connectivity λ of graph $\mathcal{G}$ is equal to the smallest number of lines that, if deleted, induce a disconnected graph or the trivial graph K_1.

Whitney's inequality: $\kappa(\mathcal{G}) \leq \lambda(\mathcal{G}) \leq \delta(\mathcal{G})$.

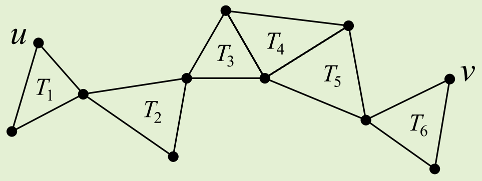
Graph $\mathcal{G}$ is *(vertex) $k-$connected*, if $\kappa(\mathcal{G}) \geq k$ and is *line $k-$connected*, if $\lambda(\mathcal{G}) \geq k$.

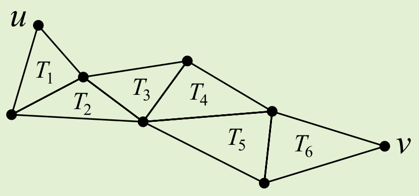
Whitney / Menger theorem: Graph $\mathcal{G}$ is vertex/line $k-$connected iff every pair of vertices can be connected with k vertex/line internally disjoint (semi)walks.

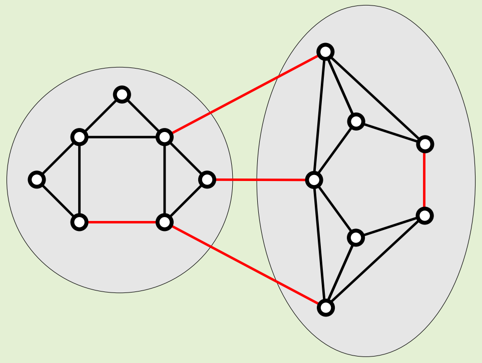
Triangular and short cycle connectivities

In an undirected graph we call a *triangle* a subgraph isomorphic to K_3.

A sequence $(T_1, T_2, \ldots, T_s)$ of triangles of $\mathcal{G}$ *(vertex) triangularly connects* vertices $u, v \in \mathcal{V}$ iff $u \in T_1$ and $v \in T_s$ or $u \in T_s$ and $v \in T_1$ and $\mathcal{V}(T_{i-1}) \cap \mathcal{V}(T_i) \neq \emptyset$, $i = 2, \ldots s$. It *edge triangularly connects* vertices $u, v \in \mathcal{V}$ iff a stronger version of the second condition holds $\mathcal{E}(T_{i-1}) \cap \mathcal{E}(T_i) \neq \emptyset$, $i = 2, \ldots s$.

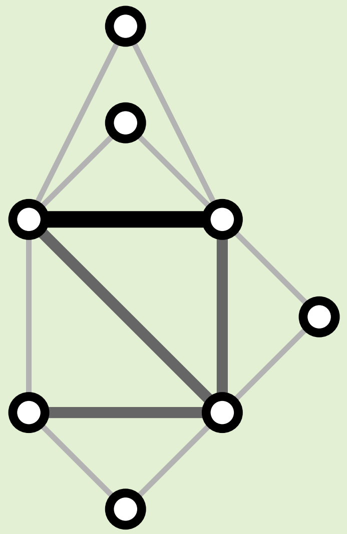






Vertex triangular connectivity is an equivalence on $\mathcal{V}$; and edge triangular connectivity is an equivalence on $\mathcal{E}$. See the paper.

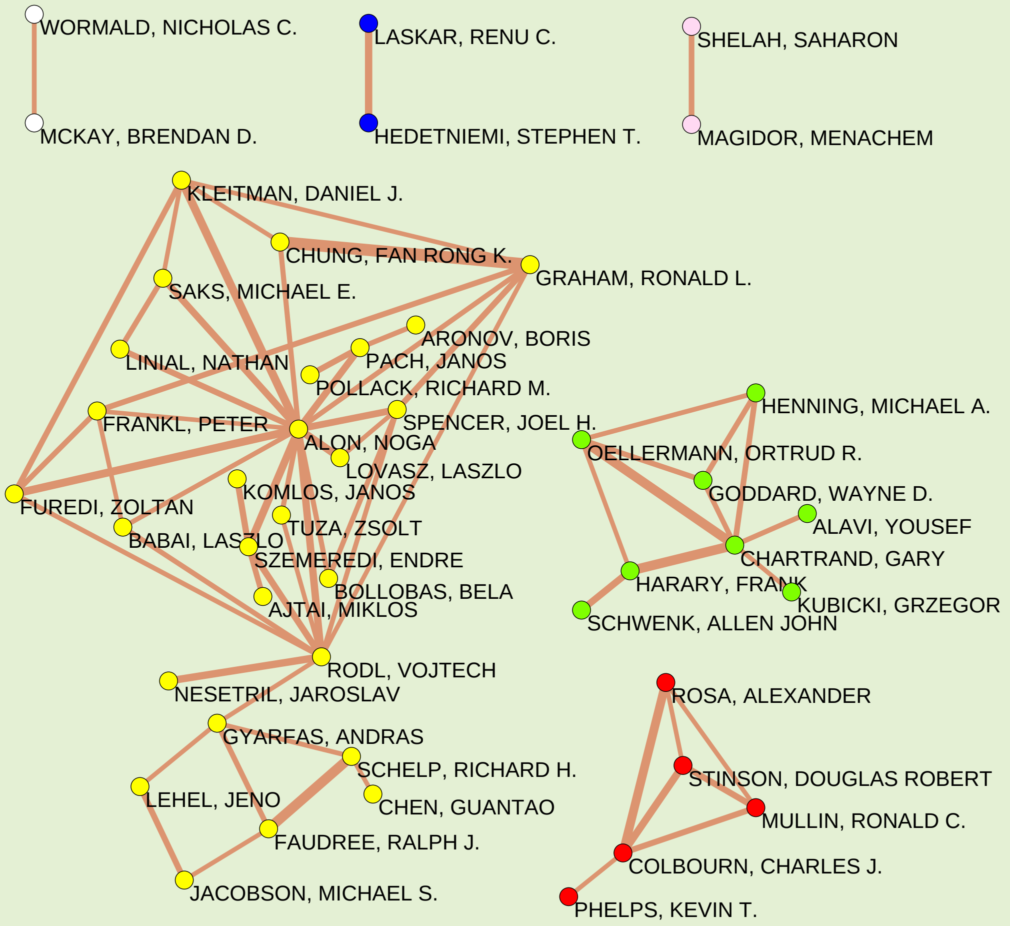
Triangular network



Let $\mathcal{G}$ be a simple undirected graph. A *triangular network* $\mathcal{N}_T(\mathcal{G}) = (\mathcal{V}, \mathcal{E}_T, w)$ determined by $\mathcal{G}$ is a subgraph $\mathcal{G}_T = (\mathcal{V}, \mathcal{E}_T)$ of $\mathcal{G}$ which set of edges $\mathcal{E}_T$ consists of all triangular edges of $\mathcal{E}(\mathcal{G})$. For $e \in \mathcal{E}_T$ the weight $w(e)$ equals to the number of different triangles in $\mathcal{G}$ to which e belongs.

Triangular networks can be used to efficiently identify dense clique-like parts of a graph. If an edge e belongs to a k-clique in $\mathcal{G}$ then $w(e) \geq k - 2$.

Edge-cut at level 16 of triangular network of Erdős collaboration graph



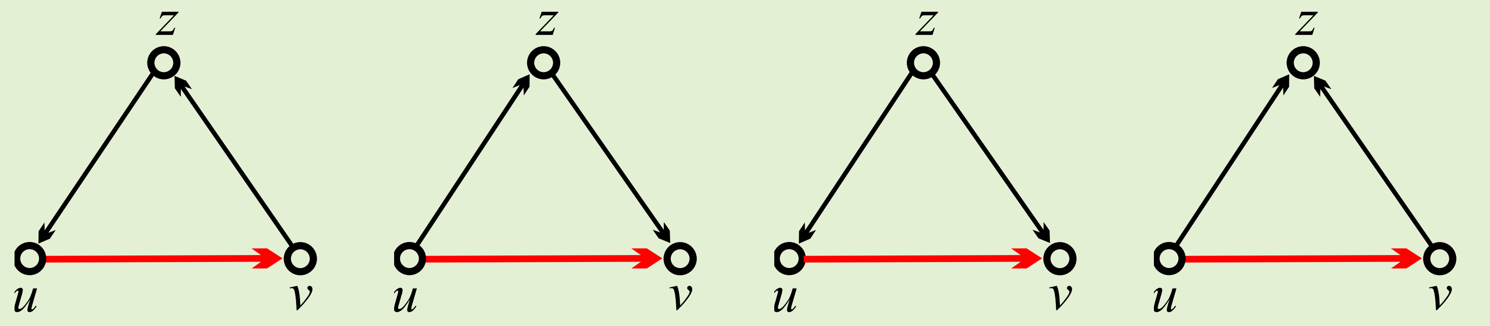
without Erdős,

$n = 6926$,

$m = 11343$

Triangular connectivity in directed graphs

If the graph $\mathcal{G}$ is mixed we replace edges with pairs of opposite arcs. In the following let $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ be a simple directed graph without loops. For a selected arc $(u, v) \in \mathcal{A}$ there are four different types of directed triangles: **cyc**lic, **tra**nsitive, **in**put and **out**put.

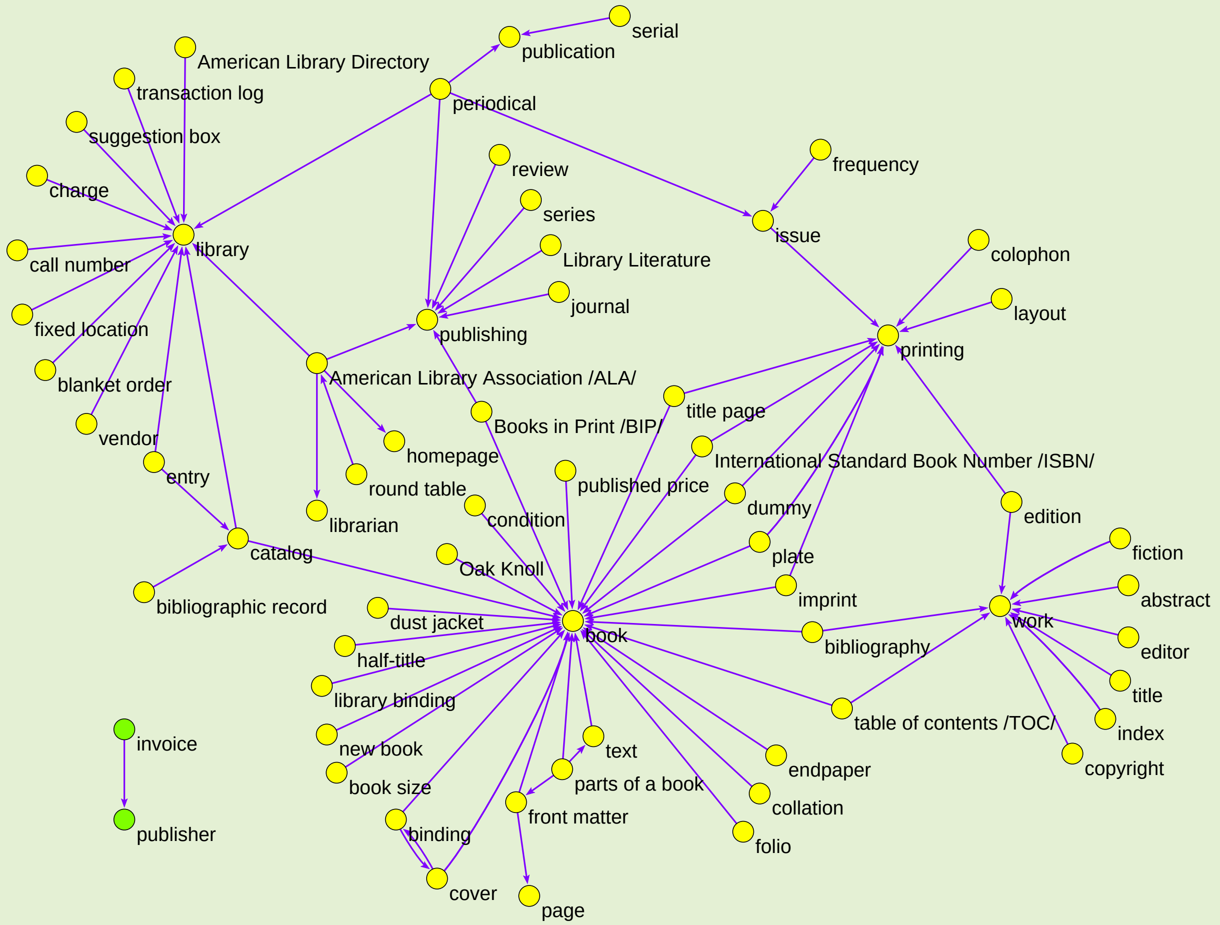


cyc tra in out

For each type we get the corresponding triangular network $\mathcal{N}_{cyc}$, $\mathcal{N}_{tra}$, $\mathcal{N}_{in}$ and $\mathcal{N}_{out}$.

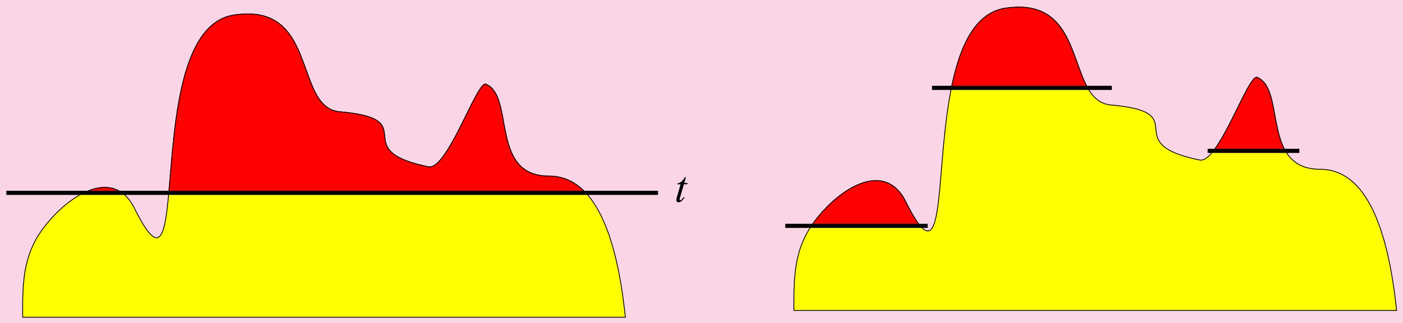
The notion of triangular connectivity can be extended to the notion of *short (semi) cycle connectivity*.

Arc-cut at level 11 of transitive triangular network of ODLIS dictionary



Islands

If we represent a given or computed value of vertices / lines as a height of vertices / lines and we immerse the network into a water up to selected level we get *islands*. Varying the level we get different islands. Islands are very general and efficient approach to determine the 'important' subnetworks in a given network.



We developed very efficient algorithms to determine the islands hierarchy and to list all the islands of selected sizes.

See details.

…Islands

A set of vertices $C \subseteq \mathcal{V}$ is a *regular vertex island* in network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, p)$, $p : \mathcal{V} \to \mathbb{R}$ iff it induces a connected subgraph and the vertices from the island are 'higher' than the neighboring vertices

$$\max_{u \in N(C)} p(u) < \min_{v \in c} p(v)$$

A set of vertices $C \subseteq \mathcal{V}$ is a *regular line island* in network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \to \mathbb{R}$ iff it induces a connected subgraph and the lines inside the island are 'stronger related' among them than with the neighboring vertices – in $\mathcal{N}$ there exists a spanning tree $\mathcal{T}$ over C such that

$$\max_{(u,v) \in \mathcal{L}, u \notin C, v \in C} w(u,v) < \min_{(u,v) \in \mathcal{T}} w(u,v)$$

Some properties of vertex islands

- The sets of vertices of connected components of vertex-cut at selected level t are regular vertex islands.
- The set $\mathcal{H}_p(\mathcal{N})$ of all regular vertex islands of network $\mathcal{N}$ is a complete hierarchy:
  - two islands are disjoint or one of them is a subset of the other
  - each vertex belongs to at least one island
- Vertex islands are invariant for the strictly increasing transformations of the property p.
- Two linked vertices cannot belong to two disjoint regular vertex islands.

Algorithm for determining regular vertex islands

- We sink the network into the water, then we lower the water level step by step.
- Each time a new vertex v appears from the water, we check with which of the already visible islands is connected.
- We join these islands and the vertex v obtaining a new (larger) island. These islands are *subislands* of the new island.
  Vertex v is a *port* of the new island (the vertex with the smallest value).
- This can be done in $\mathcal{O}(\max(n \log n, m))$ time.

Simple vertex islands

- The set of vertices $\mathcal{C} \subseteq \mathcal{V}$ is a *local vertex peak*, if it is a regular vertex island and all of its vertices have the same value.
- Vertex island with a single local vertex peak is called a *simple vertex island*.
- The types of vertex islands:
  - FLAT – all vertices have the same value
  - SINGLE – island has a single local vertex peak
  - MULTI – island has more than one local vertex peaks
- Only the islands of type FLAT or SINGLE are simple islands.

Some properties of line islands

- The sets of vertices of connected components of line-cut at selected level t are regular line islands.
- The set $\mathcal{H}_w(\mathcal{N})$ of all nondegenerated regular line islands of network $\mathcal{N}$ is hierarchy (not necessarily complete):
  - two islands are disjoint or one of them is a subset of the other
- Line islands are invariant for the strictly increasing transformations of the weight w.
- Two linked vertices may belong to two disjoint regular line islands.

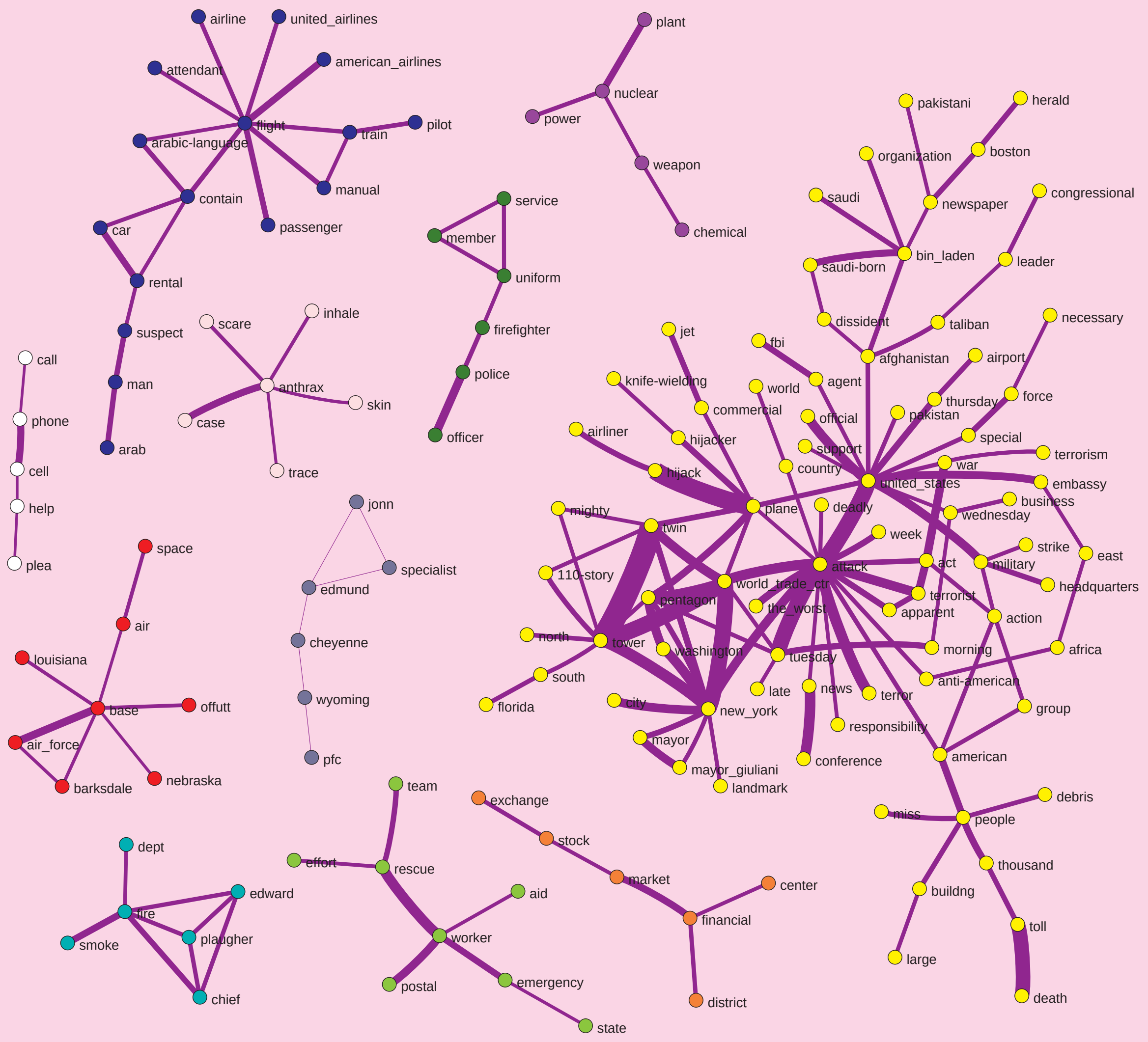
Algorithm for determining regular line islands

- We sink the network into the water, then we lower the water level step by step.
- Each time a new line e appears from the water, we check with which of the already visible islands is connected (there are exactly two such islands).
- We join these two islands obtaining a new (larger) island.
  These islands are *subislands* of the new island.
  Line e is a *port* of the new island (not necessarily the line with the smallest value).
- This can be done in $\mathcal{O}(m \log n)$ time.

Simple line islands

- The set of vertices $\mathcal{C} \subseteq \mathcal{V}$ is a *local line peak*, if it is a regular line island and there exists a spanning tree of the corresponding induced network, in which all lines have the same value as the line with the largest value.
- Line island with a single local line peak is called a *simple line island*.
- The types of line islands:
  - FLAT – there exists a spanning tree, in which all lines have the same value as the line with the largest value.
  - SINGLE – island has a single local line peak.
  - MULTI – island has more than one local line peaks.
- Only the islands of type FLAT or SINGLE are simple islands.

Islands - Reuters terror news



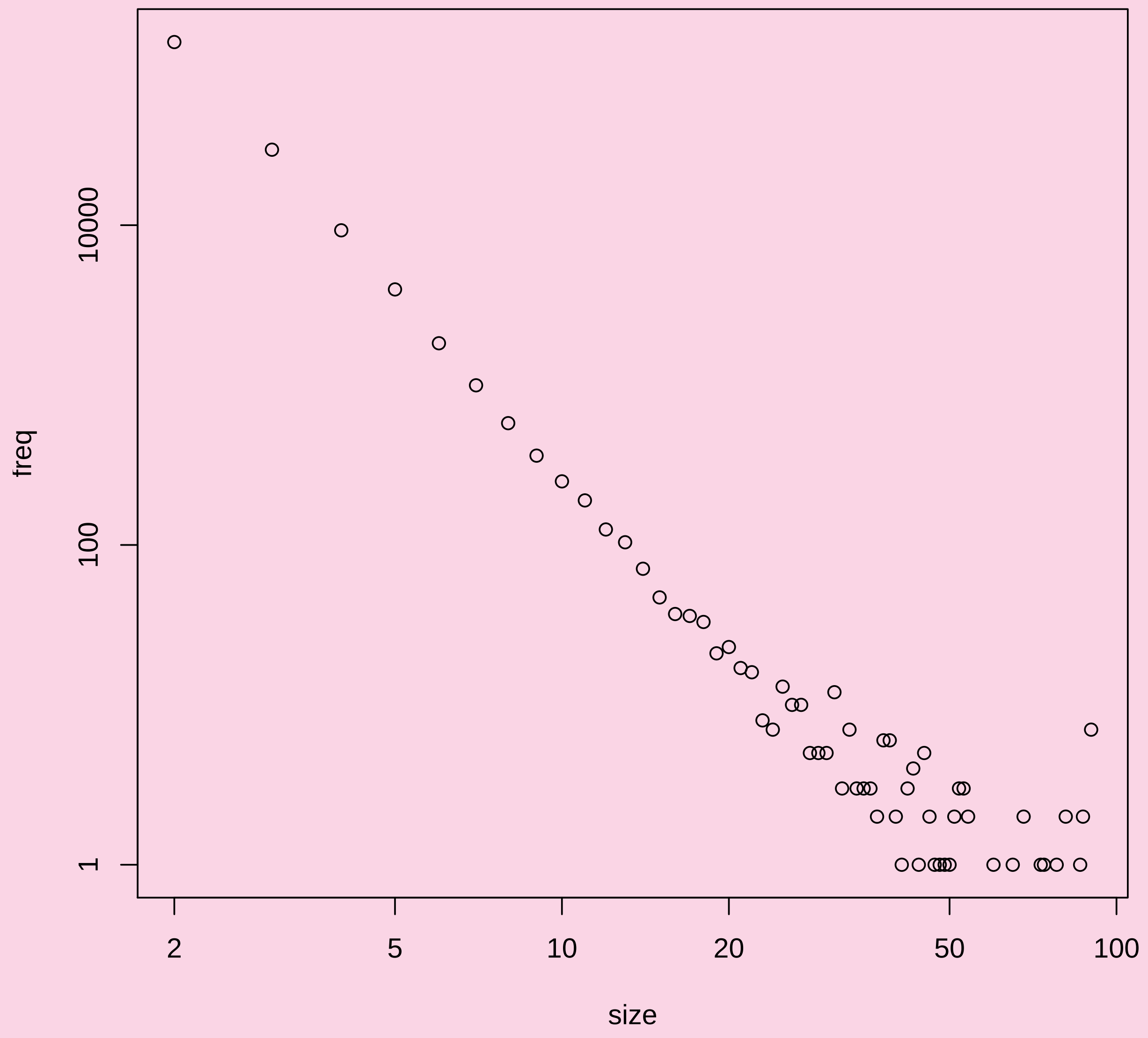
Using CRA S. Corman and K. Dooley produced the *Reuters terror news network* that is based on all stories released during 66 consecutive days by the news agency Reuters concerning the September 11 attack on the US. The vertices of a network are words (terms); there is an edge between two words iff they appear in the same text unit. The weight of an edge is its frequency. It has $n = 13332$ vertices and $m = 243447$ edges.

Islands – US patents

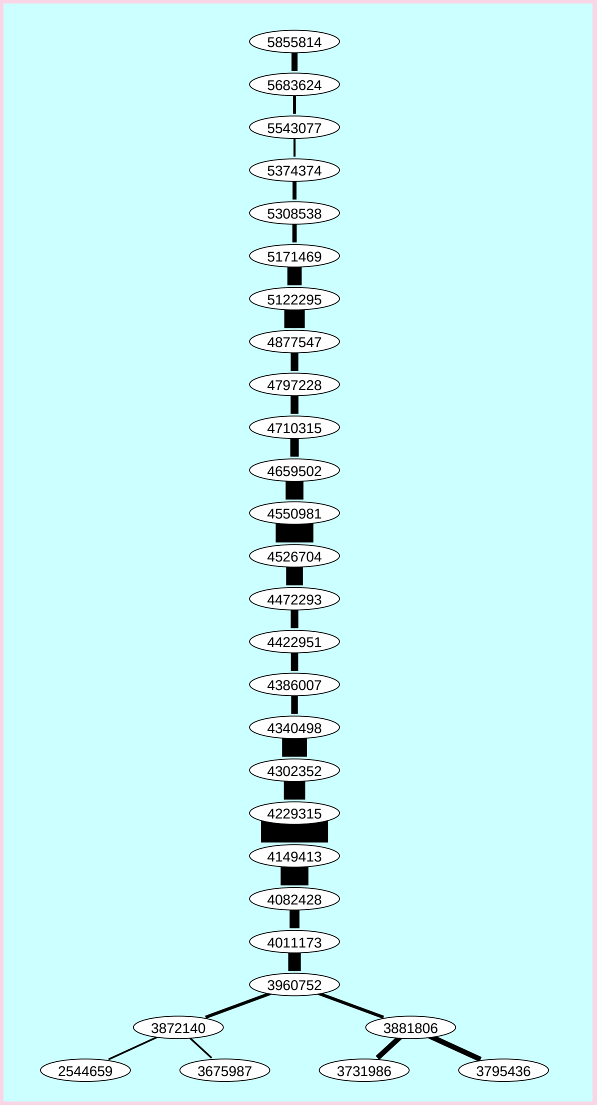
As an example, let us look at Nber network of US Patents. It has 3774768 vertices and 16522438 arcs (1 loop). We computed SPC weights in it and determined all (2,90)-islands. The reduced network has 470137 vertices, 307472 arcs and for different k: C_2 =187610, C_5 =8859,C_{30} =101, C_{50} =30 islands. Rolex

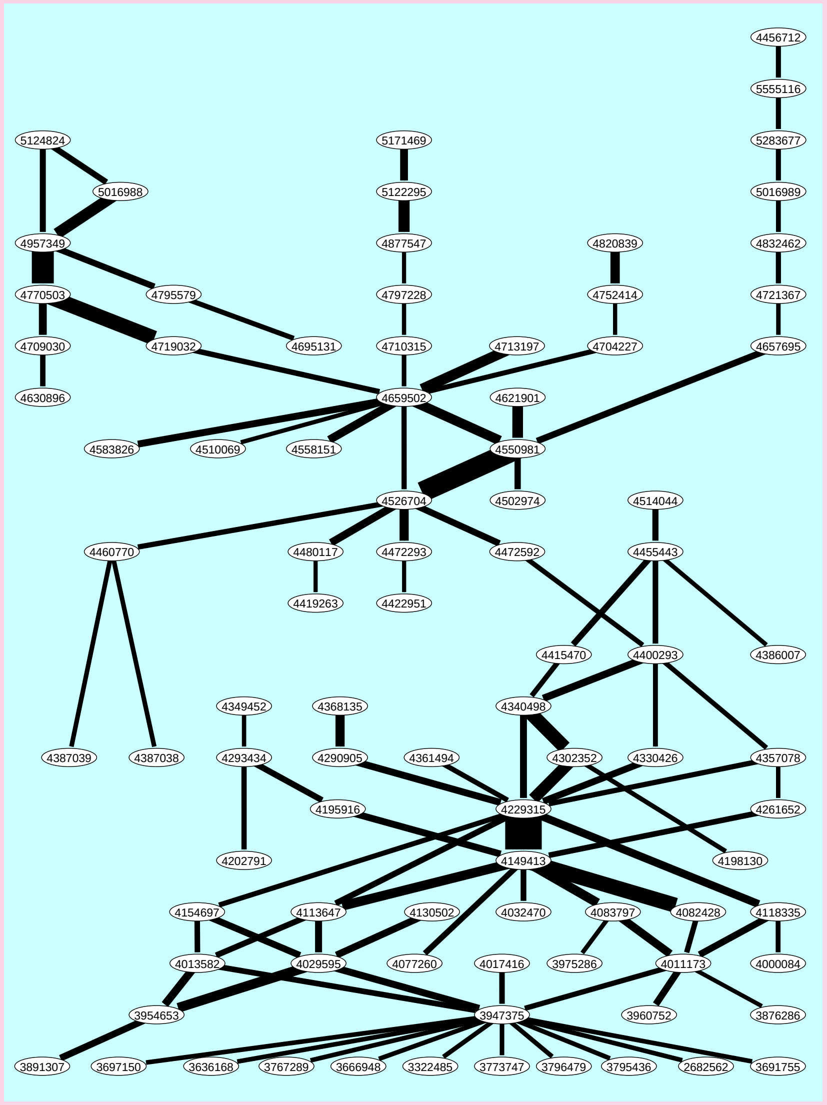
```
 [1]    0 139793  29670 9288 3966 1827 997 578 362 250
[11]  190    125    104   71   47   37  36  33  21  23
[21]   17     16      8    7   13   10  10   5   5   5
[31]   12      3      7    3    3    3   2   6   6   2
[41]    1      3      4    1    5    2   1   1   1   1
[51]    2      3      3    2    0    0   0   0   0   1
[61]    0      0      0    0    1    0   0   2   0   0
[71]    0      0      1    1    0    0   0   1   0   0
[81]    2      0      0    0    0    1   2   0   0   7
```

Island size distribution



Main path and main island of Patents





Liquid crystal display

Table 1: Patents on the liquid-crystal display

patent	date	author(s) and title
2544659	Mar 13, 1951	Dreyer. Dichroic light-polarizing sheet and the like and the formation and use thereof
2682562	Jun 29, 1954	Wender, et al. Reduction of aromatic carbinols
3322485	May 30, 1967	Williams. Electro-optical elements utilazing an organic nematic compound
3636168	Jan 18, 1972	Josephson. Preparation of polynuclear aromatic compounds
3666948	May 30, 1972	Mechlowitz, et al. Liquid crystal termal imaging system having an undisturbed image on a disturbed background
3675987	Jul 11, 1972	Rafuse. Liquid crystal compositions and devices
3691755	Sep 19, 1972	Girard. Clock with digital display
3697150	Oct 10, 1972	Wysochi. Electro-optic systems in which an electrophoretic-like or dipolar material is dispersed throughout a liquid crystal to reduce the turn-off time
3731986	May 8, 1973	Fergason. Display devices utilizing liquid crystal light modulation
3767289	Oct 23, 1973	Aviram, et al. Class of stable trans-stilbene compounds, some displaying nematic mesophases at or near room temperature and others in a range up to 100°C
3773747	Nov 20, 1973	Steinstrasser. Substituted azoxy benzene compounds
3795436	Mar 5, 1974	Boller, et al. Nematogenic material which exhibit the Kerr effect at isotropic temperatures
3796479	Mar 12, 1974	Helfrich, et al. Electro-optical light-modulation cell utilizing a nematogenic material which exhibits the Kerr effect at isotropic temperatures
3872140	Mar 18, 1975	Klanderman, et al. Liquid crystalline compositions and method
3876286	Apr 8, 1975	Deutscher, et al. Use of nematic liquid crystalline substances
3881806	May 6, 1975	Suzuki. Electro-optical display device
3891307	Jun 24, 1975	Tsukamoto, et al. Phase control of the voltages applied to opposite electrodes for a cholesteric to nematic phase transition display
3947375	Mar 30, 1976	Gray, et al. Liquid crystal materials and devices
3954653	May 4, 1976	Yamazaki. Liquid crystal composition having high dielectric anisotropy and display device incorporating same
3960752	Jun 1, 1976	Klanderman, et al. Liquid crystal compositions
3975286	Aug 17, 1976	Oh. Low voltage actuated field effect liquid crystals compositions and method of synthesis
4000084	Dec 28, 1976	Hsieh, et al. Liquid crystal mixtures for electro-optical display devices
4011173	Mar 8, 1977	Steinstrasser. Modified nematic mixtures with positive dielectric anisotropy
4013582	Mar 22, 1977	Gavrilovic. Liquid crystal compounds and electro-optic devices incorporating them
4017416	Apr 12, 1977	Inukai, et al. P-cyanophenyl 4-alkyl-4'-biphenylcarboxylate, method for preparing same and liquid crystal compositions using same
4029595	Jun 14, 1977	Ross, et al. Novel liquid crystal compounds and electro-optic devices incorporating them
4032470	Jun 28, 1977	Bloom, et al. Electro-optic device
4077260	Mar 7, 1978	Gray, et al. Optically active cyano-biphenyl compounds and liquid crystal materials containing them
4082428	Apr 4, 1978	Hsu. Liquid crystal composition and method

Table 2: Patents on the liquid-crystal display

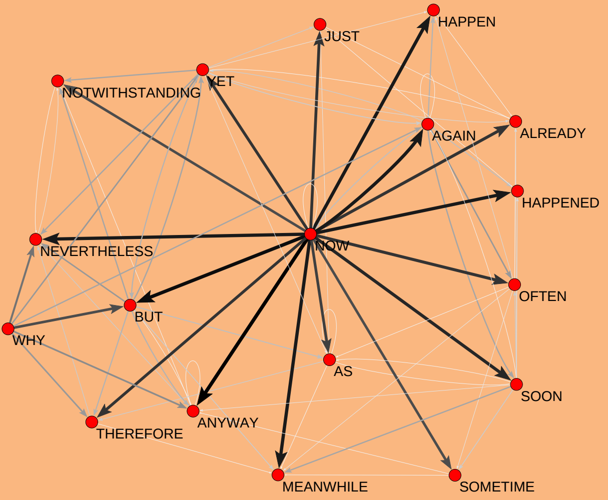
patent	date	author(s) and title
4083797	Apr 11, 1978	Oh. Nematic liquid crystal compositions
4113647	Sep 12, 1978	Coates, et al. Liquid crystalline materials
4118335	Oct 3, 1978	Krause, et al. Liquid crystalline materials of reduced viscosity
4130502	Dec 19, 1978	Eidenschink, et al. Liquid crystalline cyclohexane derivatives
4149413	Apr 17, 1979	Gray, et al. Optically active liquid crystal mixtures and liquid crystal devices containing them
4154697	May 15, 1979	Eidenschink, et al. Liquid crystalline hexahydroterphenyl derivatives
4195916	Apr 1, 1980	Coates, et al. Liquid crystal compounds
4198130	Apr 15, 1980	Boller, et al. Liquid crystal mixtures
4202791	May 13, 1980	Sato, et al. Nematic liquid crystalline materials
4229315	Oct 21, 1980	Krause, et al. Liquid crystalline cyclohexane derivatives
4261652	Apr 14, 1981	Gray, et al. Liquid crystal compounds and materials and devices containing them
4290905	Sep 22, 1981	Kanbe. Ester compound
4293434	Oct 6, 1981	Deutscher, et al. Liquid crystal compounds
4302352	Nov 24, 1981	Eidenschink, et al. Fluorophenylcyclohexanes, the preparation thereof and their use as components of liquid crystal dielectrics
4330426	May 18, 1982	Eidenschink, et al. Cyclohexylbiphenyls, their preparation and use in dielectrics and electrooptical display elements
4340498	Jul 20, 1982	Sugimori. Halogenated ester derivatives
4349452	Sep 14, 1982	Osman, et al. Cyclohexylcyclohexanoates
4357078	Nov 2, 1982	Carr, et al. Liquid crystal compounds containing an alicyclic ring and exhibiting a low dielectric anisotropy and liquid crystal materials and devices incorporating such compounds
4361494	Nov 30, 1982	Osman, et al. Anisotropic cyclohexyl cyclohexylmethyl ethers
4368135	Jan 11, 1983	Osman. Anisotropic compounds with negative or positive DC-anisotropy and low optical anisotropy
4386007	May 31, 1983	Krause, et al. Liquid crystalline naphthalene derivatives
4387038	Jun 7, 1983	Fukui, et al. 4-(Trans-4'-alkylcyclohexyl) benzoic acid 4'"-cyano-4"-biphenylyl esters
4387039	Jun 7, 1983	Sugimori, et al. Trans-4-(trans-4'-alkylcyclohexyl)-cyclohexane carboxylic acid 4'"-cyanobiphenyl ester
4400293	Aug 23, 1983	Romer, et al. Liquid crystalline cyclohexylphenyl derivatives
4415470	Nov 15, 1983	Eidenschink, et al. Liquid crystalline fluorine-containing cyclohexylbiphenyls and dielectrics and electro-optical display elements based thereon
4419263	Dec 6, 1983	Praefcke, et al. Liquid crystalline cyclohexylcarbonitrile derivatives
4422951	Dec 27, 1983	Sugimori, et al. Liquid crystal benzene derivatives
4455443	Jun 19, 1984	Takatsu, et al. Nematic halogen Compound
4456712	Jun 26, 1984	Christie, et al. Bismaleimide triazine composition
4460770	Jul 17, 1984	Petrzilka, et al. Liquid crystal mixture
4472293	Sep 18, 1984	Sugimori, et al. High temperature liquid crystal substances of four rings and liquid crystal compositions containing the same
4472592	Sep 18, 1984	Takatsu, et al. Nematic liquid crystalline compounds
4480117	Oct 30, 1984	Takatsu, et al. Nematic liquid crystalline compounds
4502974	Mar 5, 1985	Sugimori, et al. High temperature liquid-crystalline ester compounds
4510069	Apr 9, 1985	Eidenschink, et al. Cyclohexane derivatives

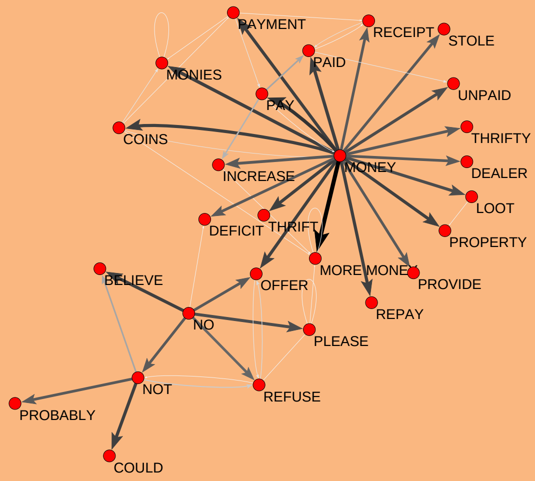
Table 3: Patents on the liquid-crystal display

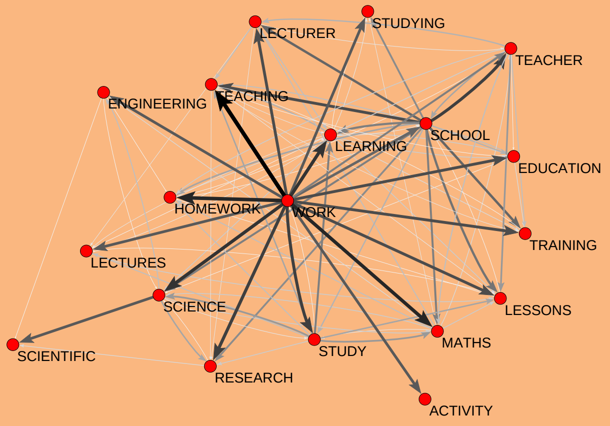
patent	date	author(s) and title
4514044	Apr 30, 1985	Gunjima, et al. 1-(Trans-4-alkylcyclohexyl)-2-(trans-4'-(p-sub stituted phenyl) cyclohexyl)ethane and liquid crystal mixture
4526704	Jul 2, 1985	Petrzilka, et al. Multiring liquid crystal esters
4550981	Nov 5, 1985	Petrzilka, et al. Liquid crystalline esters and mixtures
4558151	Dec 10, 1985	Takatsu, et al. Nematic liquid crystalline compounds
4583826	Apr 22, 1986	Petrzilka, et al. Phenylethanes
4621901	Nov 11, 1986	Petrzilka, et al. Novel liquid crystal mixtures
4630896	Dec 23, 1986	Petrzilka, et al. Benzonitriles
4657695	Apr 14, 1987	Saito, et al. Substituted pyridazines
4659502	Apr 21, 1987	Fearon, et al. Ethane derivatives
4695131	Sep 22, 1987	Balkwill, et al. Disubstituted ethanes and their use in liquid crystal materials and devices
4704227	Nov 3, 1987	Krause, et al. Liquid crystal compounds
4709030	Nov 24, 1987	Petrzilka, et al. Novel liquid crystal mixtures
4710315	Dec 1, 1987	Schad, et al. Anisotropic compounds and liquid crystal mixtures therewith
4713197	Dec 15, 1987	Eidenschink, et al. Nitrogen-containing heterocyclic compounds
4719032	Jan 12, 1988	Wachtler, et al. Cyclohexane derivatives
4721367	Jan 26, 1988	Yoshinaga, et al. Liquid crystal device
4752414	Jun 21, 1988	Eidenschink, et al. Nitrogen-containing heterocyclic compounds
4770503	Sep 13, 1988	Buchecker, et al. Liquid crystalline compounds
4795579	Jan 3, 1989	Vauchier, et al. 2,2'-difluoro-4-alkoxy-4'-hydroxydiphenyls and their derivatives, their production process and their use in liquid crystal display devices
4797228	Jan 10, 1989	Goto, et al. Cyclohexane derivative and liquid crystal composition containing same
4820839	Apr 11, 1989	Krause, et al. Nitrogen-containing heterocyclic esters
4832462	May 23, 1989	Clark, et al. Liquid crystal devices
4877547	Oct 31, 1989	Weber, et al. Liquid crystal display element
4957349	Sep 18, 1990	Clerc, et al. Active matrix screen for the color display of television pictures, control system and process for producing said screen
5016988	May 21, 1991	Iimura. Liquid crystal display device with a birefringent compensator
5016989	May 21, 1991	Okada. Liquid crystal element with improved contrast and brightness
5122295	Jun 16, 1992	Weber, et al. Matrix liquid crystal display
5124824	Jun 23, 1992	Kozaki, et al. Liquid crystal display device comprising a retardation compensation layer having a maximum principal refractive index in the thickness direction
5171469	Dec 15, 1992	Hittich, et al. Liquid-crystal matrix display
5283677	Feb 1, 1994	Sagawa, et al. Liquid crystal display with ground regions between terminal groups
5308538	May 3, 1994	Weber, et al. Supertwist liquid-crystal display
5374374	Dec 20, 1994	Weber, et al. Supertwist liquid-crystal display
5543077	Aug 6, 1996	Rieger, et al. Nematic liquid-crystal composition
5555116	Sep 10, 1996	Ishikawa, et al. Liquid crystal display having adjacent electrode terminals set equal in length
5683624	Nov 4, 1997	Sekiguchi, et al. Liquid crystal composition
5855814	Jan 5, 1999	Matsui, et al. Liquid crystal compositions and liquid crystal display elements

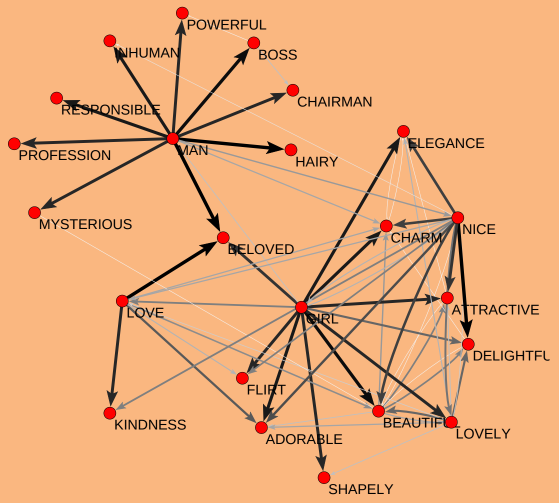
Islands – The Edinburgh Associative Thesaurus

$n = 23219$, $m = 325624$, transitivity weight









Dense groups

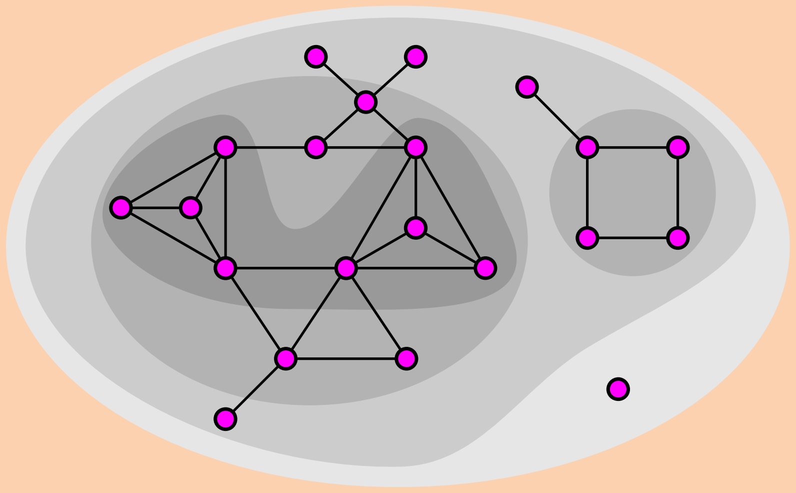
Several notions were proposed in attempts to formally describe dense groups in graphs.

Clique of order k is a maximal complete subgraph (isomorphic to K_k), $k \geq 3$.

s-plexes, LS sets, lambda sets, cores, $\dots$

For all of them, except for cores, it turned out that they are difficult to detemine.

Cores and generalized cores



The notion of core was introduced by Seidman in 1983. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph. A subgraph $\mathcal{H} = (W, \mathcal{E}|W)$ induced by the set W is a k-*core* or a *core of order* k iff $\forall v \in W : \deg_{\mathcal{H}}(v) \geq k$, and $\mathcal{H}$ is a maximal subgraph with this property. The core of maximum order is also called the *main* core.

The *core number* of vertex v is the highest order of a core that contains this vertex. The degree $\deg(v)$ can be: in-degree, out-degree, in-degree + out-degree, etc., determining different types of cores.

Properties of cores

From the figure, representing 0, 1, 2 and 3 core, we can see the following properties of cores:

- The cores are nested: $i < j \implies \mathcal{H}_j \subseteq \mathcal{H}_i$
- Cores are not necessarily connected subgraphs.

An efficient algorithm for determining the cores hierarchy is based on the following property:

> If from a given graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ we recursively delete all vertices, and edges incident with them, of degree less than k, the remaining graph is the k-core.

…Properties of cores

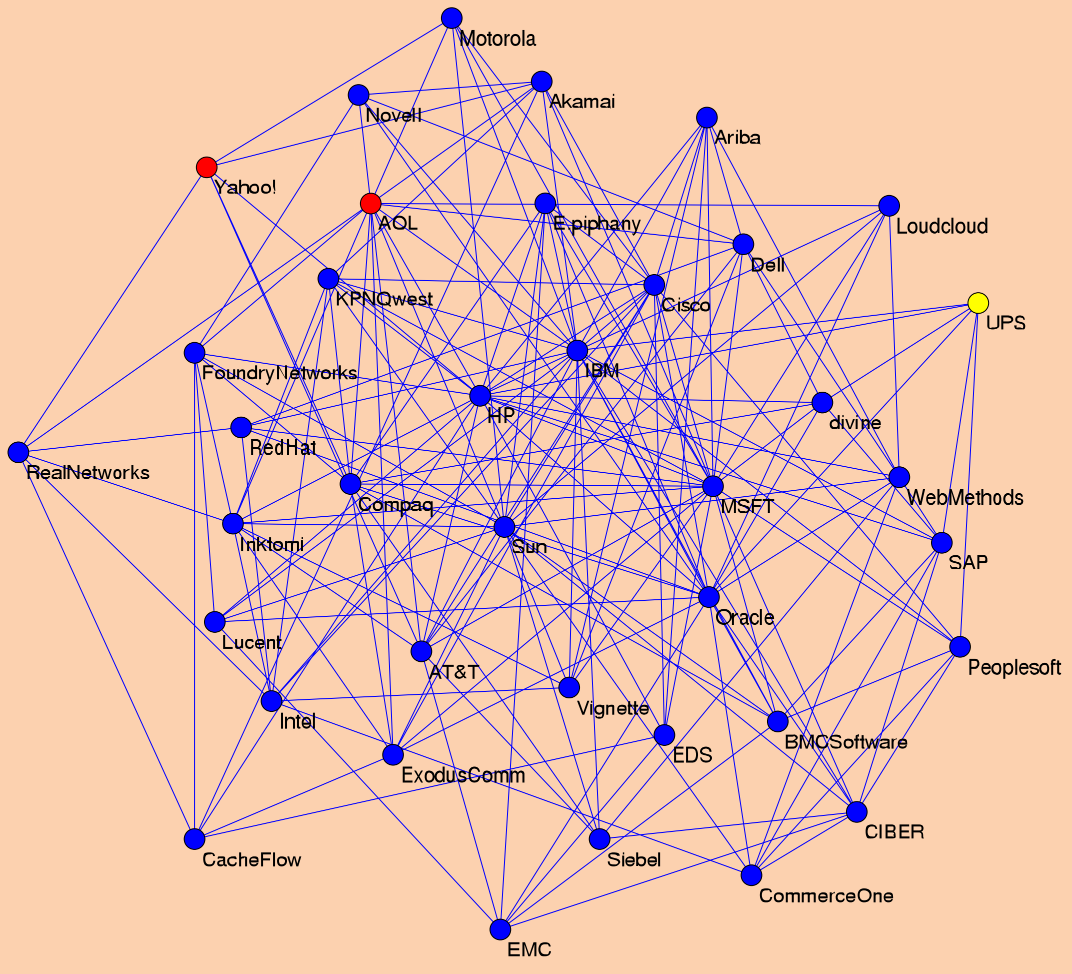
The cores, because they can be determined very efficiently, are one among few concepts that provide us with meaningful decompositions of large networks. We expect that different approaches to the analysis of large networks can be built on this basis. For example: we get the following bound on the chromatic number of a given graph $\mathcal{G}$

$$\chi(\mathcal{G}) \leq 1 + \mathrm{core}(\mathcal{G})$$

Cores can also be used to localize the search for interesting subnetworks in large networks since: if it exists, a k-component is contained in a k-core; and a k-clique is contained in a k-core.

For details see the paper.

6-core of Krebs Internet industries



Generalized cores

The notion of core can be generalized to networks. Let $\mathcal{N} = (\mathcal{V}, \mathcal{E}, w)$ be a network, where $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph and $w : \mathcal{E} \to \mathbb{R}$ is a function assigning values to edges. A *vertex property function* on **N**, or a p-*function* for short, is a function $p(v, U)$, $v \in \mathcal{V}$, $U \subseteq \mathcal{V}$ with real values. Let $N_U(v) = N(v) \cap U$. Besides degrees, here are some examples of p-functions:

$$
\begin{aligned}
p_S(v,U) &= \sum_{u \in N_U(v)} w(v,u), \text{ where } w : \mathcal{E} \to \mathbb{R}_0^+ \\
p_M(v,U) &= \max_{u \in N_U(v)} w(v,u), \text{ where } w : \mathcal{E} \to \mathbb{R} \\
p_k(v,U) &= \text{number of cycles of length } k \text{ through vertex } v \text{ in } (U, \mathcal{E}|U)
\end{aligned}
$$

The subgraph $\mathcal{H} = (C, \mathcal{E}|C)$ induced by the set $C \subseteq \mathcal{V}$ is a p-*core at level* $t \in \mathbb{R}$ iff $\forall v \in C : t \leq p(v, C)$ and C is a maximal such set.

Generalized cores algorithm

The function p is *monotone* iff it has the property

$$C_1 \subset C_2 \Rightarrow \forall v \in \mathcal{V} : (p(v, C_1) \leq p(v, C_2))$$

The degrees and the functions p_S, p_M and p_k are monotone. For a monotone function the p-core at level t can be determined, as in the ordinary case, by successively deleting vertices with value of p lower than t; and the cores on different levels are nested

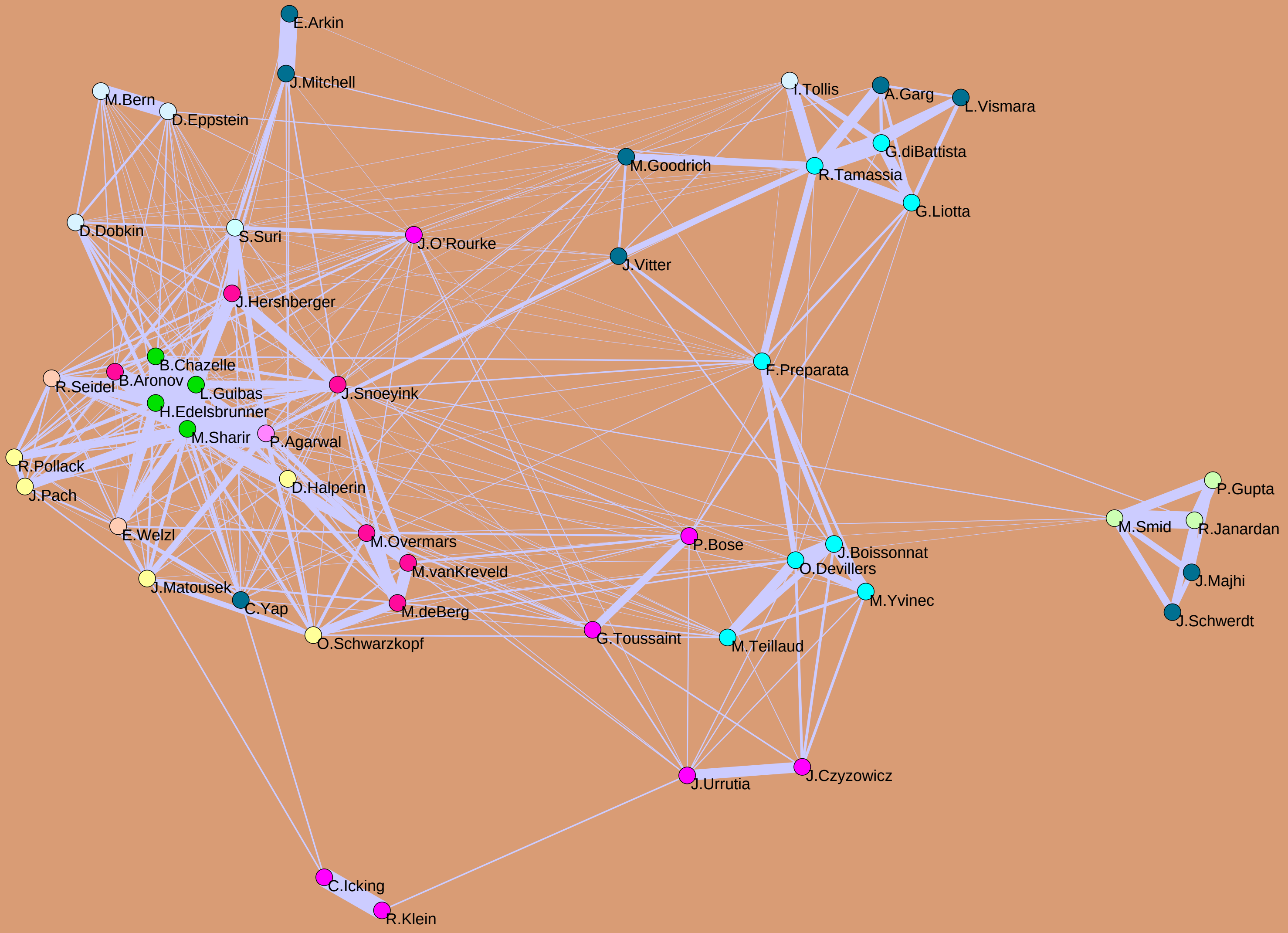
$$t_1 < t_2 \Rightarrow \mathcal{H}_{t_2} \subseteq \mathcal{H}_{t_1}$$

The p-function is *local* iff $p(v, U) = p(v, N_U(v))$.

The degrees, p_S and p_M are local; but p_k is **not** local for $k \geq 4$. For a local p-function an $O(m \max(\Delta, \log n))$ algorithm for determining the p-core levels exists, assuming that $p(v, N_C(v))$ can be computed in $O(\deg_C(v))$.

For details see the paper.

p_S-core at level 46 of Geombib network


E.Arkin
J.Mitchell
M.Bern
D.Eppstein
I.Tollis
A.Garg
L.Vismara
G.diBattista
M.Goodrich
R.Tamassia
G.Liotta
D.Dobkin
S.Suri
J.O'Rourke
J.Vitter
J.Hershberger
B.Chazelle
R.Seidel
B.Aronov
L.Guibas
J.Snoeyink
F.Preparata
H.Edelsbrunner
M.Sharir
P.Agarwal
R.Pollack
J.Pach
D.Halperin
P.Gupta
M.Smid
R.Janardan
E.Welzl
M.Overmars
P.Bose
J.Boissonnat
O.Devillers
M.vanKreveld
J.Majhi
J.Matousek
C.Yap
M.deBerg
M.Yvinec
J.Schwerdt
O.Schwarzkopf
G.Toussaint
M.Teillaud
J.Urrutia
J.Czyzowicz
C.Icking
R.Klein