Analysis of Large Networks with Pajek

Vladimir Batagelj

University of Ljubljana

ESSIR 2011 – 8th European Summer School on Information Retrieval

29 Aug - 02 Sep 2011, Koblenz, Germany

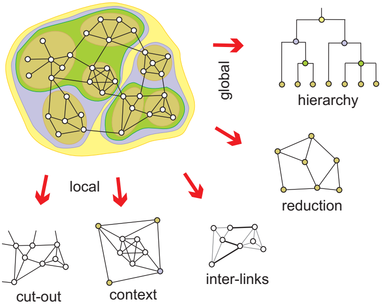
Outline

Pajek and large networks

Pajek is a program for analysis and visualization of *large* networks.
large ≡ the network can be stored in the computer memory.
Network = Graph + Data

Pajek is mostly a two men (A. Mrvar and V. Batagelj) project. We started to develop Pajek in 1996. It was assembled from experiences and code from my projects on graph algorithms in eighties and first half of nineties, and Andrej's master thesis on graph visualization. It is programmed in Delphi Pascal for Windows 32. A 64-bit Windows Delphi version is ready for release.
In November 2010 we also started to develop a new basic network analysis library (64-bit, C++).
Large networks are *sparse* (Dunbar number). For large structures already quadratic algorithms are too slow.

Pajek's backround



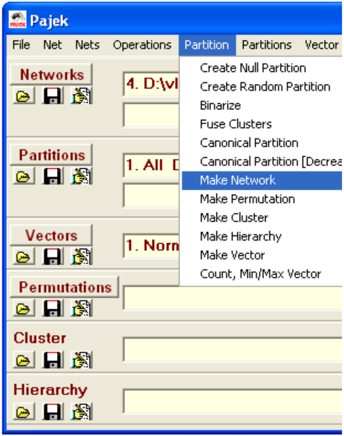
The main goals in the design of Pajek are:

- to support abstraction by (recursive) *decomposition* of a large network into several smaller networks that can be treated further using more sophisticated methods;
- to provide the user with some powerful *visualization* tools;
- to implement a selection of efficient *subquadratic* algorithms for analysis of large networks.

New algorithms

- vertex and line cuts,
- vertex and line islands,
- (generalized) cores,
- triadic spectrum; 3-rings and 4-rings weights,
- fragment (motif) searching,
- hierarchical clustering with relational constraints,
- Doreian & Hummon weights in acyclic networks,
- multiplication of networks,
- fast Pathfinder algorithm, . . .

Pajek is a network 'calculator'



In Pajek analysis and visualization are performed using 6 data types:

- *network* (graph),
- *partition* (nominal or ordinal properties of vertices),
- *vector* (numerical properties of vertices),
- *cluster* (subset of vertices),
- *permutation* (reordering of vertices, ordinal properties), and
- *hierarchy* (general tree structure on vertices).

Pajek supports also *multi-relational*, *temporal* and *two-mode* networks. Low level granularity of operations – a sequence of operations is usually needed to do a task (macros); but it is also more flexible.

Pajek's network description language

Multi-relational temporal network – KEDS/WEIS

```
% Recoded by WEISmonths, Sun Nov 28 21:57:00 2004
% from http://www.ku.edu/~keds/data.dir/balk.html
*vertices 325
1 "AFG" [1-*]
2 "AFR" [1-*]
3 "ALB" [1-*]
4 "ALBMED" [1-*]
5 "ALG" [1-*]
 ...
318 "YUGGOV" [1-*]
319 "YUGMAC" [1-*]
320 "YUGMED" [1-*]
321 "YUGMTN" [1-*]
322 "YUGSER" [1-*]
323 "ZAI" [1-*]
324 "ZAM" [1-*]
325 "ZIM" [1-*]
*arcs :0 "*** ABANDONED"
*arcs :10 "YIELD"
*arcs :11 "SURRENDER"
*arcs :12 "RETREAT"
 ...
*arcs :223 "MIL ENGAGEMENT"
*arcs :224 "RIOT"
*arcs :225 "ASSASSINATE TORTURE"
*arcs
224: 314 153 1 [4]              890402  YUG     KSV     224  (RIOT)  RIOT-TORN
212: 314 83 1 [4]               890404  YUG     ETHALB  212  (ARREST PERSON) ALB ETHNIC JAILED
224: 3 83 1 [4]                 890407  ALB     ETHALB  224  (RIOT)  RIOTS
123: 83 153 1 [4]               890408  ETHALB  KSV     123  (INVESTIGATE)   PROBING
 ...
42: 105 63 1 [175]              030731  GER     CYP     042  (ENDORSE)       GAVE SUPPORT
212: 295 35 1 [175]             030731  UNWCT   BOSSER  212  (ARREST PERSON) SENTENCED TO PRIS
43: 306 87 1 [175]              030731  VAT     EUR     043  (RALLY) RALLIED
13: 295 35 1 [175]              030731  UNWCT   BOSSER  013  (RETRACT)       CLEARED
121: 295 22 1 [175]             030731  UNWCT   BAL     121  (CRITICIZE)     CHARGES
122: 246 295 1 [175]            030731  SER     UNWCT   122  (DENIGRATE)     TESTIFIED
121: 35 295 1 [175]             030731  BOSSER  UNWCT   121  (CRITICIZE)     ACCUSED
```

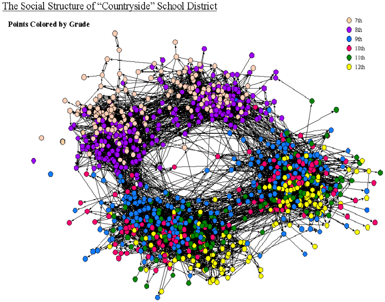
Network visualization

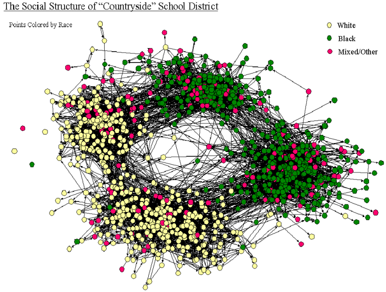
Standard network visualization methods can produce readable results for not too large and relatively sparse networks.
For denser networks the matrix representation is usually the right choice.
In network analysis it is very important to support also visualization of additional data.
It seems that interactive layouts are the future of network visualization.
In Pajek the following visualization tools are available:

- spring embedders: Kamada Kawai, Fruchterman Reingold
- eigen vectors
- acyclic
- manual improvements
- matrix representation

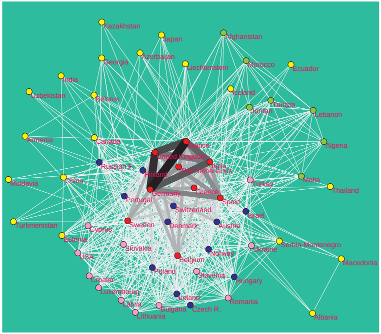
In nineties we won several first prizes at the Graph Drawing competitions.

Display of properties – school (Moody)





Analysis of Countries.net

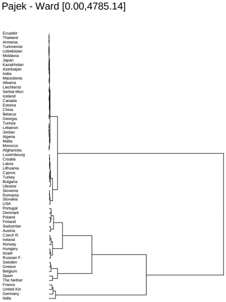


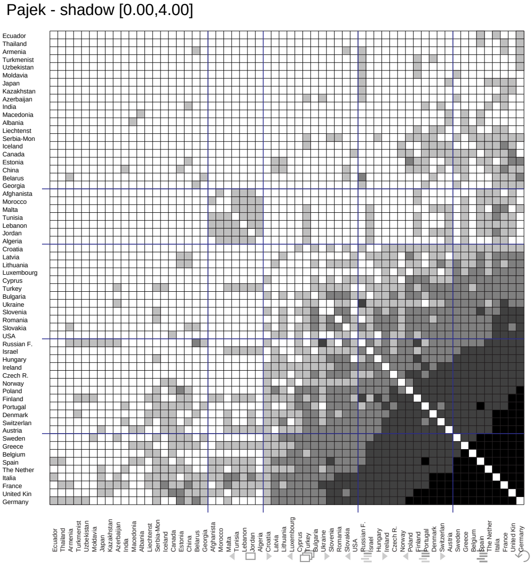
To obtain picture in which the stronger lines cover weaker lines we have to sort them `Net/Transform/Sort lines/Line values/Ascending` For dense (sub)networks we get better visualization by using matrix display. In this case we also recoded values (2,10,50). To determine clusters we used Ward's clustering procedure with dissimilarity measure d_5 (corrected Euclidean distance).

The permutation determined by hierarchy can often be improved by changing the positions of clusters. We get a typical center-periphery structure.

More: Batagelj, V.: Complex Networks, Visualization of. R.A. Meyers, ed., Encyclopedia of Complexity and Systems Science, Springer 2009: 1253-1268.

Matrix display



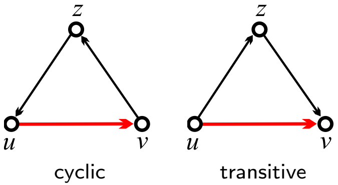


k-rings

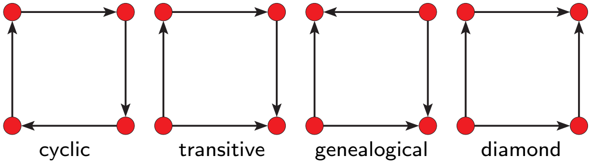
A *k-ring* is a simple closed chain of length k. Using k-rings we can define a weight of edges as
$w_k(e) = \#$ of different k-rings containing the edge $e \in \mathcal{E}$
The edges belonging to cliques have large weights. Therefore these weights can be used to identify the dense parts of a network.
The k-rings can be efficiently determined only for small values of k – 3, 4, 5. The 3-rings (triangular) weights were implemented in Pajek in May 2002 and 4-rings in August 2005.
On the k-rings we can also base the notion of short cycle connectivity which provides us with another decomposition of networks.
In two-mode network there are no 3-rings. The densest substructures are complete bipartite subgraphs $K_{p,q}$. They contain many 4-rings.

Directed 3- and 4-rings

There are 2 types of directed 3-rings:

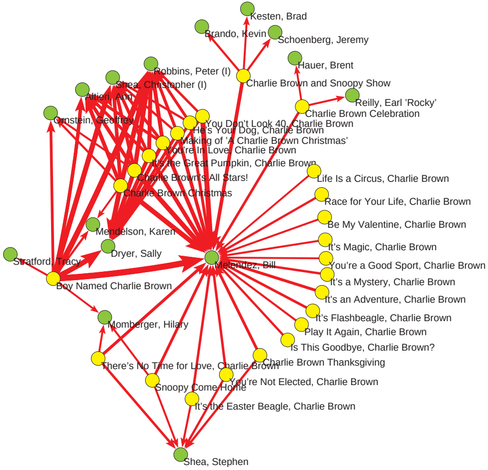


and 4 types of directed 4-rings:

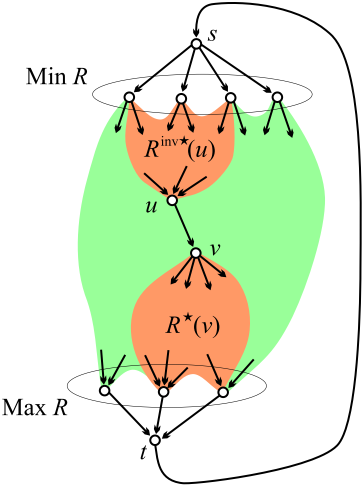


In the case of transitive rings Pajek provides a special weight counting on how many transitive rings the arc is a *shortcut*.

Charlie Brown island for w_4 in IMDB



Hummon and Doreian weights



In an acyclic network the *search path count* (SPC) weights (Hummon and Doreian, 1989) are based on counters $n(u, v)$ that count the number of different paths from s to t through the arc (u, v). To compute $n(u, v)$ we introduce two auxiliary quantities: $n^-(v)$ counts the number of different paths from s to v, and $n^+(v)$ counts the number of different paths from v to t. They can be efficiently computed. $n(u, v) = n^-(u) \cdot n^+(v)$.

Properties of SPC weights

The values of counters $n(u,v)$ form a flow in the citation network – the *Kirchoff's vertex law* holds: For every vertex u in a standardized citation network *incoming flow* = *outgoing flow*:

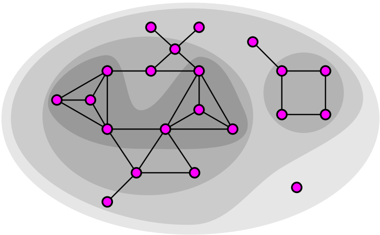
$$\sum_{v:vRu} n(v,u) = \sum_{v:uRv} n(u,v) = n^-(u) \cdot n^+(u)$$

The weight $n(t,s)$ equals to the total flow through network and provides a natural normalization of weights

$$w(u,v) = \frac{n(u,v)}{n(t,s)} \quad \Rightarrow \quad 0 \leq w(u,v) \leq 1$$

and if C is a minimal arc-cut-set $\sum_{(u,v)\in C} w(u,v) = 1$.
In large networks the values of weights can grow very large. This should be considered in the implementation of the algorithms.

Cores and generalized cores

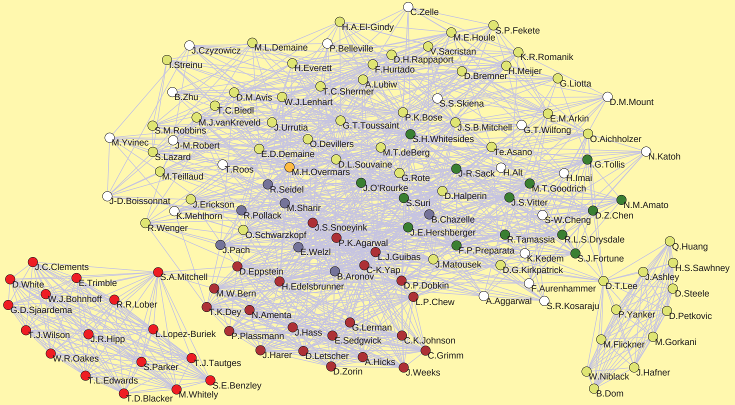


The notion of core was introduced by Seidman in 1983. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph. A subgraph $\mathcal{H} = (W, \mathcal{E}|W)$ induced by the set W is a k-*core* or a *core of order* k iff $\forall v \in W : \deg_{\mathcal{H}}(v) \geq k$, and $\mathcal{H}$ is a maximal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

Algorithm: 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.

Cores of orders 10–21 in Computational Geometry



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 $\mathbf{N}$, or a *p-function* for short, is a function $p(v, U)$, $v \in \mathcal{V}$, $U \subseteq \mathcal{V}$ with real values. Let $N_U(v) = N(v) \cap U$.
Some examples of p-functions:

$$p_S(v, U) = \sum_{u \in N_U(v)} w(v, u), \text{ where } w : \mathcal{E} \to \mathbb{R}_0^+$$

$$p_M(v, U) = \max_{u \in N_U(v)} w(v, u), \text{ where } w : \mathcal{E} \to \mathbb{R}$$

$$p_k(v, U) = \text{number of cycles of length } k \text{ through vertex } v \text{ in } (U, \mathcal{E}|U)$$

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.

Additional p-functions

relative density

$$p_\gamma(v,\mathcal{C}) = \frac{\deg(v,\mathcal{C})}{\max_{u\in N(v)} \deg(u)}, \text{ if } \deg(v) > 0;\ 0, \text{ otherwise}$$

diversity

$$p_\delta(v,\mathcal{C}) = \max_{u\in N^+(v,\mathcal{C})} \deg(u) - \min_{u\in N^+(v,\mathcal{C})} \deg(u)$$

average weight

$$p_a(v,\mathcal{C}) = \frac{1}{|N(v,\mathcal{C})|} \sum_{u\in N(v,\mathcal{C})} w(v,u), \text{ if } N(v,\mathcal{C}) \neq \emptyset;\ 0,$$

otherwise

Generalized cores algorithm

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

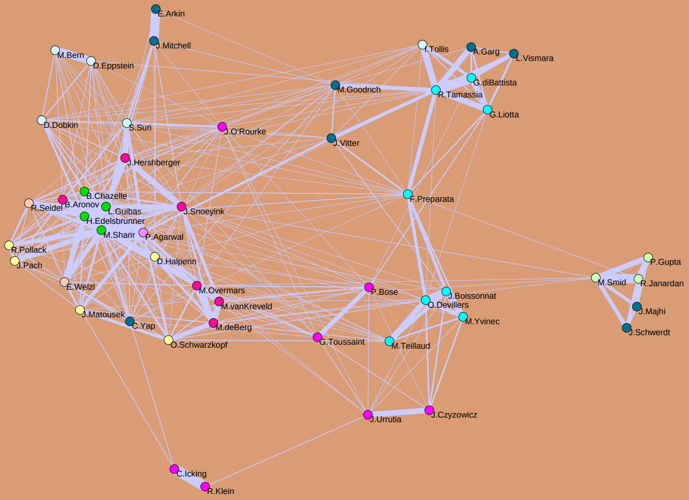
$$C_1 \subset C_2 \Rightarrow \forall v \in \mathcal{V} : (p(v, C_1) \leq p(v, C_2))$$

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.
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



Cuts

The standard approach to find interesting groups inside a network is based on properties/weights – they can be *measured* or *computed* from network structure.

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

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

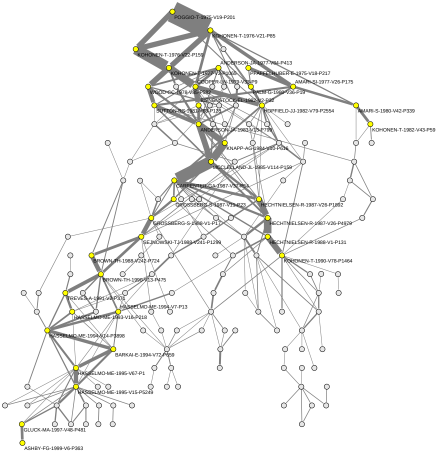
and $\mathcal{L}(\mathcal{V}')$ is the set of lines from $\mathcal{L}$ that have both endpoints in $\mathcal{V}'$.

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 $\mathcal{L}'$.

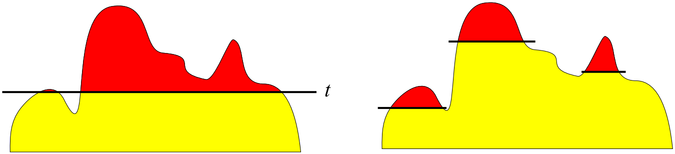
SOM citation network



Main subnetwork for Hummon and Doreian SPC weights (arc cut at level 0.007) of the SOM (self-organizing maps) citation network (4470 vertices, 12731 arcs).
See paper.

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.



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

Islands

Islands are very general and efficient approach to determine the *important* subnetworks in a given network.

We have to express the goals of our analysis with a related property of the vertices or weight of the lines. Using this property we determine the islands of an appropriate size (in the interval k to K).

In large networks we can get many islands which we have to inspect individually and interpret their content.

An important property of the islands is that they identify locally important subnetworks on different levels. Therefore they detect also emerging groups.

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

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

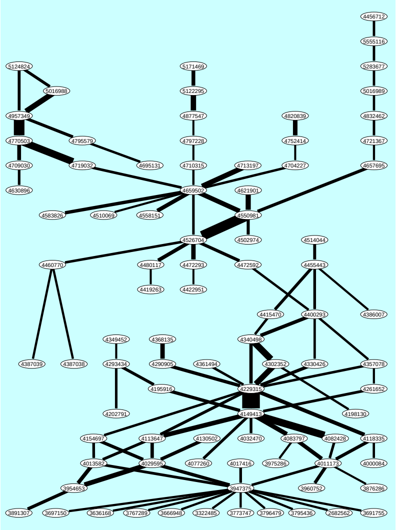
A set of vertices $C \subseteq \mathcal{V}$ is a *regular line island* in network $\mathcal{N} = (\mathcal{V}, \mathcal{L}, w)$, $w : \mathcal{L} \to \mathbb{R}$ iff it induces a connected subgraph and the lines inside the island are 'stronger linked' among them than with the neighboring vertices – in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

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

An island is *simple* iff it has a single peak.
See details.

Islands – US patents

The Nber network of US Patents 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.



US 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

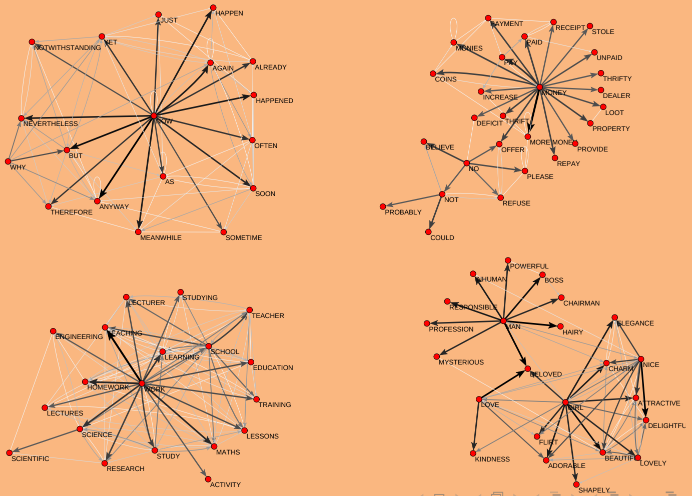
patent	date	author(s) and title
4083797	Apr 11, 1978	Oh. Nematic liquid crystal compositions
4113647	Sep 12, 1978	Coates, et al. Liquid crystalline materials
4118335	Oct 3, 1978	Krause, et al. Liquid crystalline materials of reduced viscosity
4130502	Dec 19, 1978	Eidenschink, et al. Liquid crystalline 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-substituted 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



Analysis of two-mode networks

A *two-mode network* or *affiliation network* is a structure $\mathcal{N} = (\mathcal{U}, \mathcal{V}, \mathcal{A}, w)$, where $\mathcal{U}$ and $\mathcal{V}$ are disjoint sets of vertices, $\mathcal{A}$ is the set of arcs with the initial vertex in the set $\mathcal{U}$ and the terminal vertex in the set $\mathcal{V}$, and $w : \mathcal{A} \to \mathbb{R}$ is a weight. If no weight is defined we can assume a constant weight $w(u, v) = 1$ for all arcs $(u, v) \in \mathcal{A}$. The set $\mathcal{A}$ can be viewed also as a relation $A \subseteq \mathcal{U} \times \mathcal{V}$.

A two-mode network can be formally represented by rectangular matrix $\mathbf{W} = [w_{uv}]_{\mathcal{U}\times\mathcal{V}}$.

$$w_{uv} = \begin{cases} w(u,v) & (u,v) \in \mathcal{A} \\ 0 & \text{otherwise} \end{cases}$$

Analysis of two-mode networks

For direct analysis of two-mode networks we can use the *eigen-vector approach* – a two-mode variant of Kleinberg's hubs and authorities. The weight vector $(\mathbf{x}, \mathbf{y})$ on $\mathcal{U} \cup \mathcal{V}$ is determined by relations $\mathbf{y} = \mathbf{W}\mathbf{x}$ and $\mathbf{x} = \mathbf{W}^T\mathbf{y}$.

In 2005 we proposed two new direct methods: *two-mode cores* and *4-rings*.

We can also use the *clustering* and *blockmodeling* in two-mode networks.

Multiplication of networks

To a simple two-mode *network* $\mathcal{N} = (\mathcal{I}, \mathcal{J}, \mathcal{E}, w)$; where $\mathcal{I}$ and $\mathcal{J}$ are sets of *vertices*, $\mathcal{E}$ is a set of *edges* linking $\mathcal{I}$ and $\mathcal{J}$, and $w : \mathcal{E} \to \mathbb{R}$ (or some other semiring) is a *weight*; we can assign a *network matrix* $\mathbf{W} = [w_{i,j}]$ with elements: $w_{i,j} = w(i,j)$ for $(i,j) \in \mathcal{E}$ and $w_{i,j} = 0$ otherwise.

Given a pair of compatible networks $\mathcal{N}_A = (\mathcal{I}, \mathcal{K}, \mathcal{E}_A, w_A)$ and $\mathcal{N}_B = (\mathcal{K}, \mathcal{J}, \mathcal{E}_B, w_B)$ with corresponding matrices $\mathbf{A}_{\mathcal{I}\times\mathcal{K}}$ and $\mathbf{B}_{\mathcal{K}\times\mathcal{J}}$ we call a *product of networks* $\mathcal{N}_A$ and $\mathcal{N}_B$ a network $\mathcal{N}_C = (\mathcal{I}, \mathcal{J}, \mathcal{E}_C, w_C)$, where $\mathcal{E}_C = \{(i,j) : i \in \mathcal{I}, j \in \mathcal{J}, c_{i,j} \neq 0\}$ and $w_C(i,j) = c_{i,j}$ for $(i,j) \in \mathcal{E}_C$. The product matrix $\mathbf{C} = [c_{i,j}]_{\mathcal{I}\times\mathcal{J}} = \mathbf{A} * \mathbf{B}$ is defined in the standard way

$$c_{i,j} = \sum_{k\in\mathcal{K}} a_{i,k} \cdot b_{k,j}$$

In the case when $\mathcal{I} = \mathcal{K} = \mathcal{J}$ we are dealing with ordinary one-mode networks (with square matrices).

Matrix multiplication

The standard matrix multiplication has the complexity $O(|\mathcal{I}| \cdot |\mathcal{K}| \cdot |\mathcal{J}|)$ – it is too slow to be used for large networks. For sparse large networks we can multiply faster considering only nonzero elements.

In general the multiplication of large sparse networks is a 'dangerous' operation since the result can 'explode' – it is not sparse.

If at least one of the sparse networks $\mathcal{N}_A$ and $\mathcal{N}_B$ has small maximal degree on $\mathcal{K}$ then also the resulting product network $\mathcal{N}_C$ is sparse.

Two-mode network analysis by conversion to one-mode network

Often we transform a two-mode network $\mathcal{N} = (\mathcal{U}, \mathcal{V}, \mathcal{E}, w)$ into an ordinary (one-mode) network $\mathcal{N}_1 = (\mathcal{U}, \mathcal{E}_1, w_1)$ or/and $\mathcal{N}_2 = (\mathcal{V}, \mathcal{E}_2, w_2)$, where $\mathcal{E}_1$ and w_1 are determined by the matrix $\mathbf{W}^{(1)} = \mathbf{W}\mathbf{W}^T$, $w^{(1)}_{uv} = \sum_{z\in\mathcal{V}} w_{uz} \cdot w^T_{zv}$. Evidently $w^{(1)}_{uv} = w^{(1)}_{vu}$. There is an edge $(u : v) \in \mathcal{E}_1$ in $\mathcal{N}_1$ iff $N(u) \cap N(v) \neq \emptyset$. Its weight is $w_1(u, v) = w^{(1)}_{uv}$.
The network $\mathcal{N}_2$ is determined in a similar way by the matrix $\mathbf{W}^{(2)} = \mathbf{W}^T\mathbf{W}$.
The networks $\mathcal{N}_1$ and $\mathcal{N}_2$ are analyzed using standard methods.

Networks from data tables

RuthDELmain.csv

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	Ident	Num	File	ORGANISATION ORI(	ORG	Org	Contact Name	Street	ZIP	Project	City	Country	coun	EU	Region
2	1	1480	613.html	3D PLUS SA	3D F	3D I	LIGNIER, Olivier	641 RU	78530	IST-2001-3440	Buc	FRANCE	20	2	ÎLE DE FF
3	2	1481	613.html	3D PLUS SA	3D PLUS		LIGNIER, Olivier	641 Ru	78530	IST-2001-3440	Buc	FRANCE	20	2	ÎLE DE FF
4	3	4001	924.html	3D VISION	3D \	3D '	MARIAT, Jacques	Savoie	73375	502909	Le Bo	FRANCE	20	2	CENTRE-I
5	4	1648	160.html	3D Web Technologies	3D WEB		DENNISON, Andrew		M31 4XL	BMH4989519	Carrir	UNITED KI	60	2	NORTH W
6	5	1406	442.html	3E	3E		PALMERS, Geert	Eredier	1000	NNE5/51/1999	Bruxe	BELGIQUE	8	2	REG.BRU
7	6	1007	884.html	4M2C PATRIC SALOI	4M2C PA		N/A	CRAN/	12157	507255	Berlir	DEUTSCH	15	2	BERLIN E
8	7	7914	991.html	5T S.c.r.l.	5T S.C.R		N/A	C.so B	10126	Road2/506716	Torin(	ITALIA	26	2	NORD OV
9	8	6880	588.html	A & C 2000 S.R.L.	A &	SAI	CARLUCCI, Renz	VIALE	148	IST-2001-3454	Roma	ITALIA	26	2	LAZIO Ro
10	9	6881	588.html	A & C 2000 S.R.L.	A & C 20		CARLUCCI, Renz	Viale C	148	IST-2001-3454	Roma	ITALIA	26	2	LAZIO Ro
11	10	1647	176.html	A. BENETTI MACCHII	A. BENE		Federico BENETT	Via Pro	54033	BRST985466	Carra	ITALIA	26	2	CENTRO
12	11	6605	984.html	A. Mickiewicz Univers	A. M	Inst	PATKOWSKI, Ad	Ul. H. \	61-712	502235	Pozn	POLSKA	45	2	
13	12	6571	135.html	A.BRITO - INDUSTRI/	A.BRITO		VIEIRA DE BRIT(	5109,E	4350-11!	BRST985263	Porto	PORTUGA	46	2	CONTINEI
14	13	1813	409.html	A.L. DIGITAL LIMITEI	A.L.	A.L.	LAURIE, Ben	VOYSE	W4 4GB	IST-2000-2633	Chisv	UNITED KI	60	2	SOUTH E.
15	14	1814	409.html	A.L. Digital Limited	A.L.	DIG	LAURIE, Ben	Voysey	W4 4GB	IST-2000-2633	Chisv	UNITED KI	60	2	SOUTH E.
16	15	1885	960.html	A.P. MOLLER-MAER	A.P.	TE(	DRAGSTED, Jorr	Esplan	1098	506676	Kope	DANMARK	14	2	Københav
17	16	6731	537.html	A.S.M. S.A.	A.S.M.	S	MOYA GARCIA, .	Carrete	43206	IST-2000-3008	Reus	ESPAÑA	19	2	ESTE CA
18	17	8150	232.html	AABO AKADEMI UNI	AAB	COI	NYBACKA-WILLI	14-18B	20500	ERK5-CT-1999	Turku	SUOMI/FII	53	2	MANNER-
19	18	8152	662.html	AABO AKADEMI UNI	AAB	DEF	BJORKSTRAND,	3,Tykis	20521	EVK1-CT-2002	Turku	SUOMI/FII	53	2	
20	19	8148	959.html	AABO AKADEMI UNI	AAB	Dep	HUPA, Mikko	Domky	20500	502679	Turku	SUOMI/FII	53	2	MANNER-
21	20	8151	233.html	AABO AKADEMI UNI	AAB	DEF	NYBACKA-WILLI	Lemmi	20500	ERK6-CT-1999	Turku	SUOMI/FII	53	2	MANNER-
22	21	125	116.html	AACHEN UNIVERSIT	AAC	GIE	E. NEUSSL	Intzest	52072	BRPR980663	Aach	DEUTSCH	15	2	NORDRHE
23	22	123	104.html	AACHEN UNIVERSIT	AAC	GIE	MEISER, Lukas	Intzest	52072	BRPR980695	Aach	DEUTSCH	15	2	NORDRHE
24	23	155	364.html	AACHEN UNIVERSIT	AAC	INS	RAUHUT, Burkha	18,Eilfs	52062	G1RD-CT-200(	Aach	DEUTSCH	15	2	NORDRHE

A *data table* $\mathcal{T}$ is a set of *records* $\mathcal{T} = \{T_k : k \in \mathcal{K}\}$, where $\mathcal{K}$ is the set of *keys*. A record has the form $T_k = (k, q_1(k), q_2(k), \ldots, q_r(k))$ where $q_i(k)$ is the value of the *property* (attribute) $\mathbf{q}_i$ for the key k.

Networks from data tables

Suppose that the property $\mathbf{q}$ has the range $2^{\mathcal{Q}}$. For example:
Authors[WasFau] = { S. Wasserman, K. Faust },
PubYear[WasFau] = { 1994 }, ...
If $\mathcal{Q}$ is finite (it can always be transformed in such set by partitioning the set $\mathcal{Q}$ and recoding the values) we can assign to the property $\mathbf{q}$ a two-mode network $\mathcal{K} \times \mathbf{q} = (\mathcal{K}, \mathcal{Q}, \mathcal{E}, w)$ where $(k, v) \in \mathcal{E}$ iff $v \in q(k)$, and $w(k, v) = 1$.
Also, for properties $\mathbf{q}_i$ and $\mathbf{q}_j$ we can define a two-mode network
$\mathbf{q}_i \times \mathbf{q}_j = (\mathcal{Q}_i, \mathcal{Q}_j, \mathcal{E}, w)$ where $(u, v) \in \mathcal{E}$ iff
$\exists k \in \mathcal{K} : (q_i(k) = u \wedge q_j(k) = v)$, and
$w(u, v) = \text{card}(\{k \in \mathcal{K} : (q_i(k) = u \wedge q_j(k) = v)\})$.
It holds $[\mathbf{q}_i \times \mathbf{q}_j]^T = \mathbf{q}_j \times \mathbf{q}_i$ and
$\mathbf{q}_i \times \mathbf{q}_j = [\mathcal{K} \times \mathbf{q}_i]^T * [\mathcal{K} \times \mathbf{q}_j] = [\mathbf{q}_i \times \mathcal{K}] * [\mathcal{K} \times \mathbf{q}_j]$.
We can join a pair of properties $\mathbf{q}_i$ and $\mathbf{q}_j$ also with respect to the third property $\mathbf{q}_s$: we get a two-mode network $[\mathbf{q}_i \times \mathbf{q}_j]/\mathbf{q}_s = [\mathbf{q}_i \times \mathbf{q}_s] * [\mathbf{q}_s \times \mathbf{q}_j]$.

EU projects on simulation

For the meeting **The Age of Simulation** at Ars Electronica in Linz, January 2006 a dataset of EU projects on simulation was collected by FAS research, Vienna and stored in the form of Excel table (`SimPro.csv`). The rows are the projects participants (idents) and colomns correspond to different their properties. Three two-mode networks were produced from this table using Jürgen Pfeffer's `Text2Pajek` program:

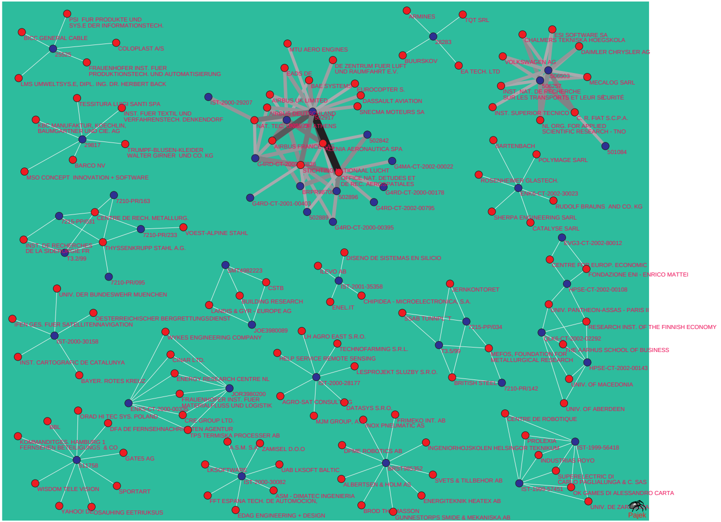
- `project.net` – $\mathbf{P}$ = [idents × projects]
- `country.net` – $\mathbf{C}$ = [idents × countries]
- `institution.net` – $\mathbf{U}$ = [idents × institutions]

|idents| = 8869, |projects| = 933, |institutions| = 3438, |countries| = 60.

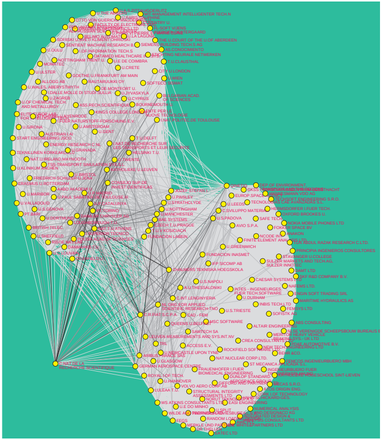
Since all three networks have the common set (idents) we can derive from them using *network multiplication* several interesting networks:

- `ProjInst.net` – $\mathbf{W}$ = [projects × institutions] = $\mathbf{P}^T * \mathbf{U}$
- `Countries.net` – $\mathbf{S}$ = [countries × countries] = $\mathbf{C}^T * \mathbf{C}$
- `Institutions.net` – $\mathbf{Q}$ = [institutions × institutions] = $\mathbf{W}^T * \mathbf{W}$
- ...

Analysis of ProjInst.net



Analysis of Institutions.net



To identify the most important institutions we first computed p_S-cores vector and use it to determine the corresponding vertex islands. We got essentially one large island. Again the corresponding subnetwork is very dense.

WoS2Pajek

For converting WoS file into networks in Pajek's format a program WoS2Pajek was developed (in Python). It produces the following files:

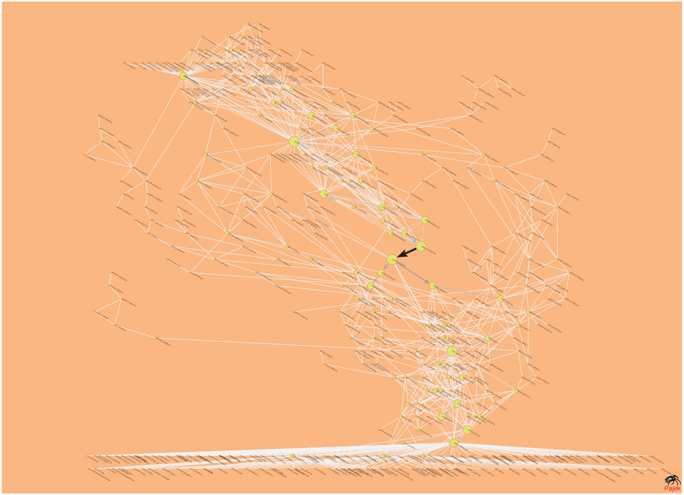
- citation network: works × works;
- authorship (two-mode) network: works × authors, for works without complete description only the first author is known;
- keywords (two-mode) network: works × keywords, only for works with complete description;
- journals (two-mode) network: works × journals, field J9;
- partition of works by the publication year;
- partition of works – complete description (1) / ISI name only (0);
- vector number of pages, PG or EP − BP +1.

WoS2Pajek/ derived networks

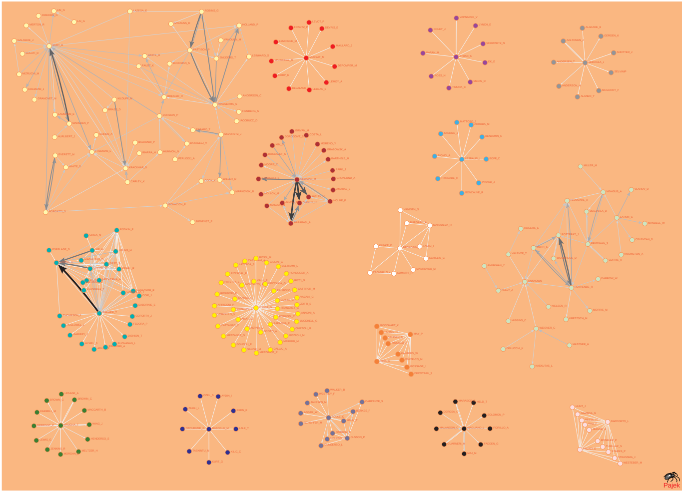
Let us denote the citation network with **Ci** , u **Ci** $v \equiv u \text{ cites } v$, and the authorship network with **WA**. Then
Co $=$ **WA**^{T} $*$ **WA** is the *collaboration network*;
or better
Co$'$ $=$ **WA**^{T} $*$ **diag**$(\frac{1}{\deg v})$ $*$ **WA**
Ca $=$ **WA**^{T} $*$ **Ci** $*$ **WA** is a network of citations between authors.
biCo $=$ **Ci** $*$ **Ci**^{T} is the *bibliographic coupling* network
coCi $=$ **Ci**^{T} $*$ **Ci** is the *co-citation* network **coCi**

Since the network can be quite large we first eliminate the only-cited works. The weight $w(a, p)$ in the *author citation* network
ACi $=$ **WA**^{T} $*$ **Ci** counts the number of times author a cited work p.
Let $b(\mathbf{A})$ denotes the binarized version of **A**. The *author co-citation* network can be obtained as **ACo** $= b($**ACi**$) * b($**ACi**$)^T$

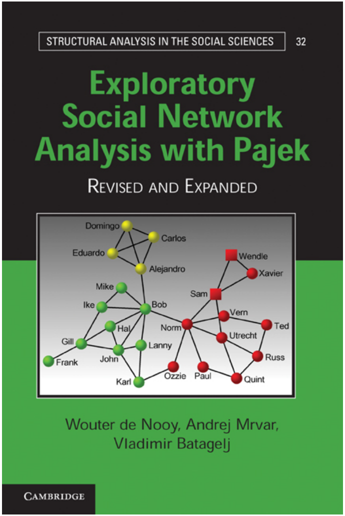
Main island in citation network for SPC in Social networks



Islands in authors' citations network in Social networks



ESNA Pajek



Pajek – program for analysis and visualization of large networks is freely available, for noncommercial use, at its web site.

http://pajek.imfm.si/

An introduction to social network analysis with Pajek is available in the book ESNA (de Nooy, Mrvar, Batagelj 2005; second, extended edition 2011).

ESNA in Japanese was published by Tokyo Denki University Press in 2010.