Network Analysis
Structure of Networks 4
Cohesion

Vladimir Batagelj

University of Ljubljana

ECPR Summer School, July 30 – August 16, 2008

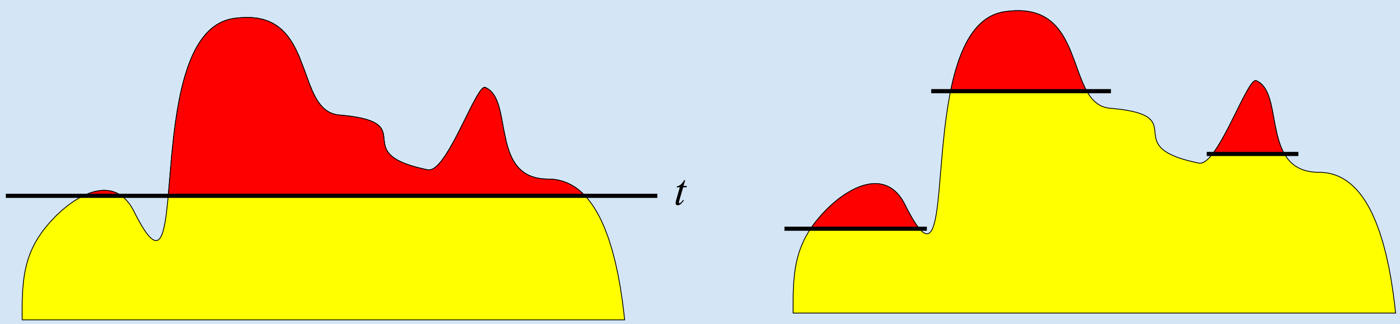
Faculty of Social Sciences, University of Ljubljana

version: August 11, 2008 / 04 : 00

Outline

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.

See details.

. . . 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 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.

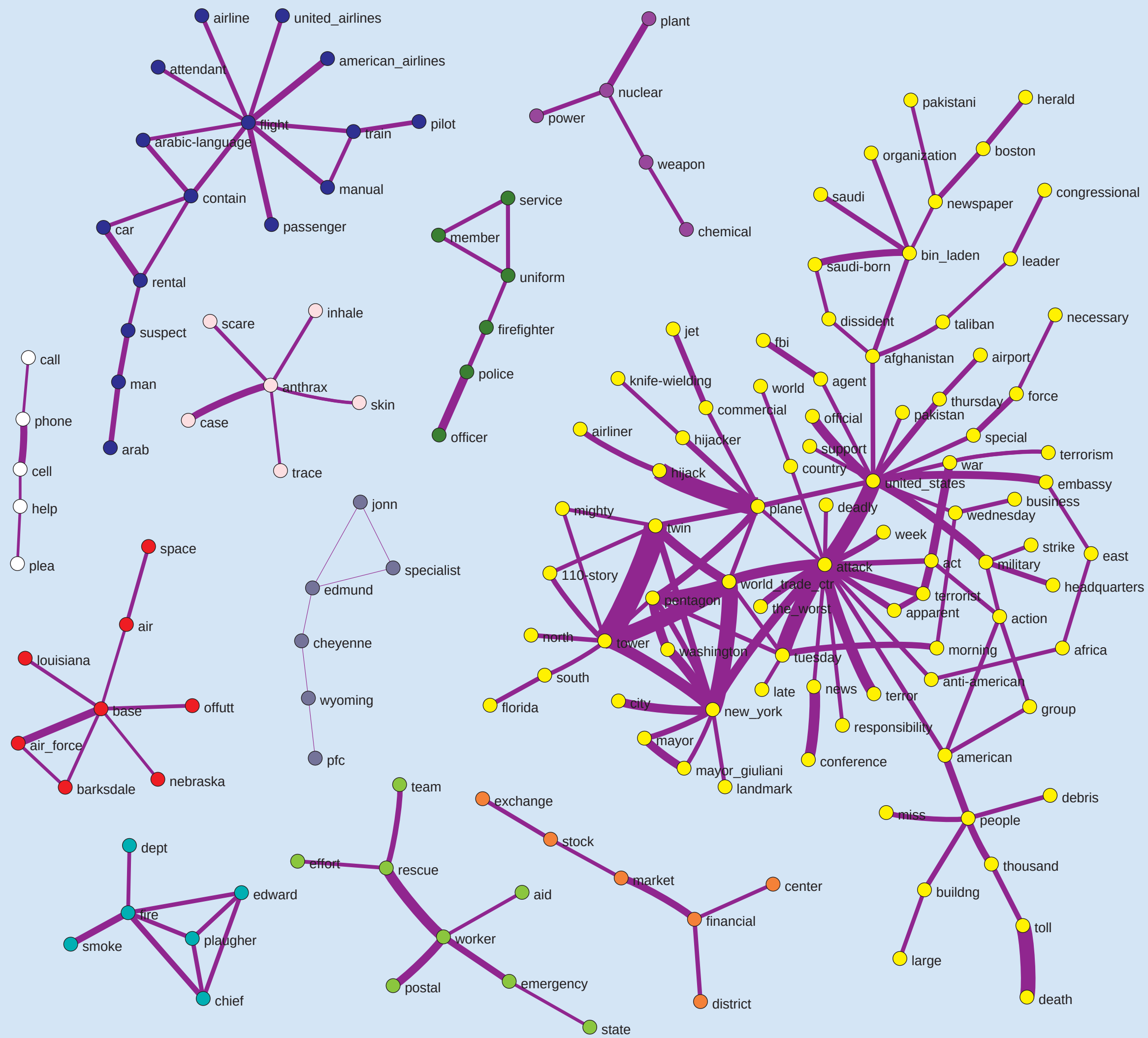
Algorithm for determining regular line islands

- 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



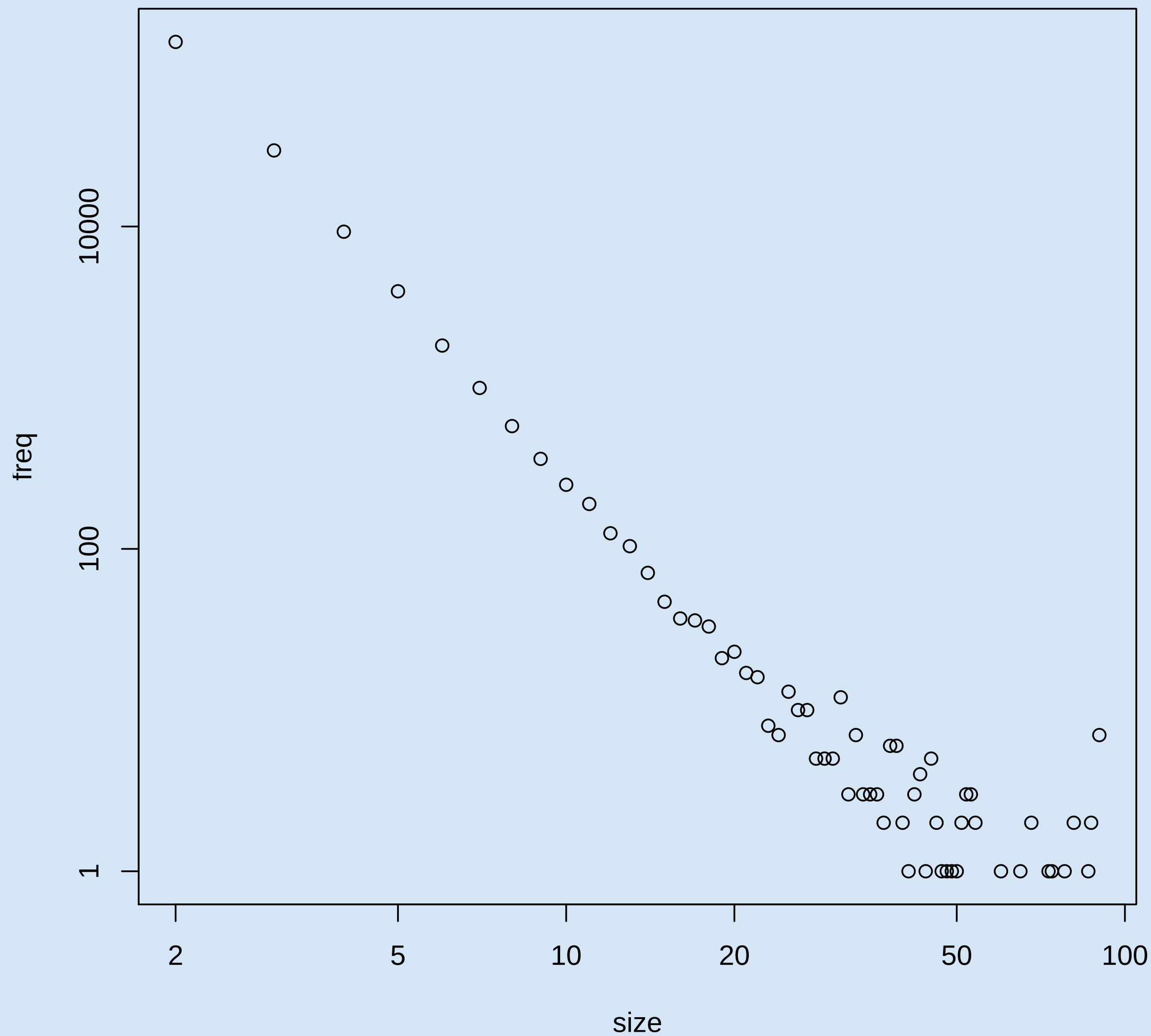
Using CRA S. Corman and K. Dooley produced the *Reuters terror news network* that is based on all stories released during 66 consecutive days by the news agency Reuters concerning the September 11 attack on the US. The vertices of a network are words (terms); there is an edge between two words iff they 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

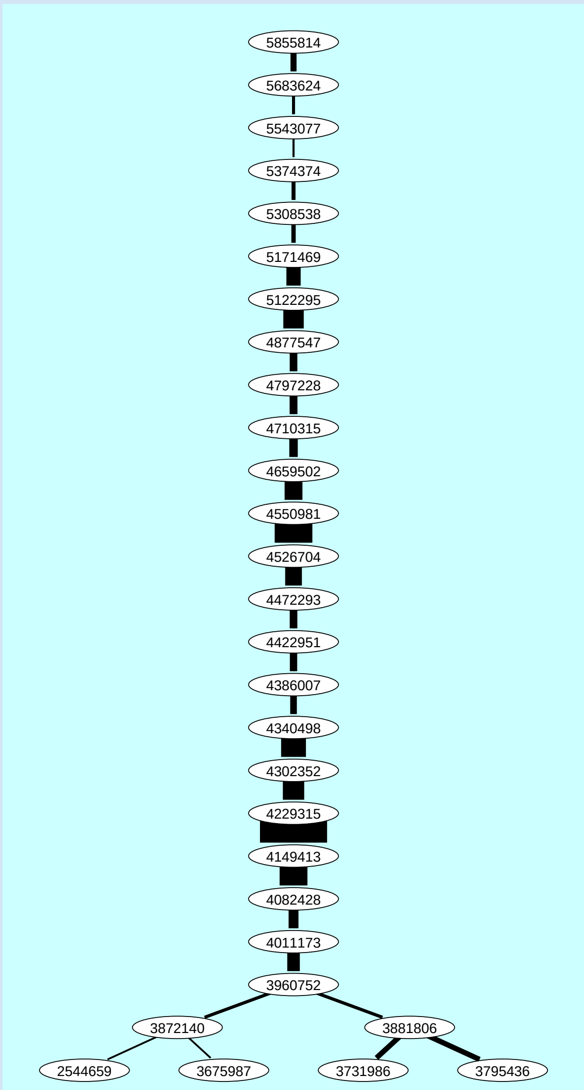
As an example, let us look at Nber network of US Patents. It has 3774768 vertices and 16522438 arcs (1 loop). We computed SPC weights in it and determined all (2,90)-islands. The reduced network has 470137 vertices, 307472 arcs and for different k: C_2 =187610, C_5 =8859,C_{30} =101, C_{50} =30 islands.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

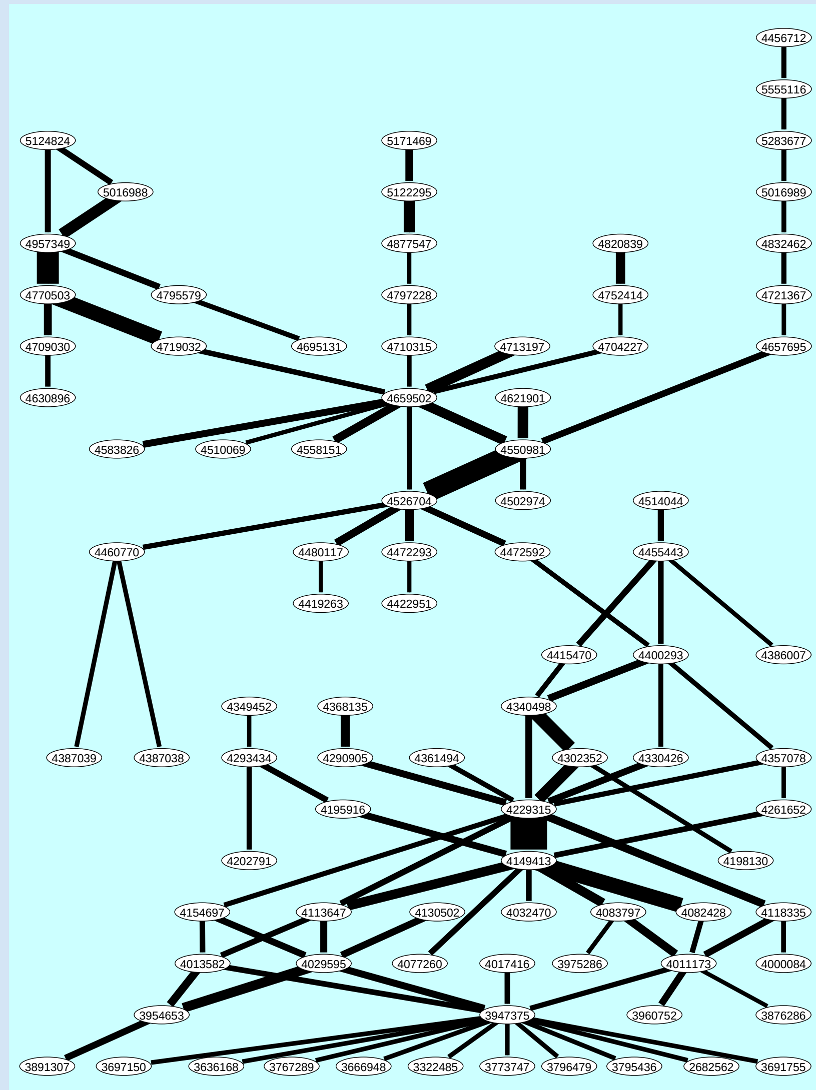
```
 [1]   0 139793  29670 9288 3966 1827 997 578 362 250
[11] 190    125    104   71   47   37  36  33  21  23
[21]  17     16      8    7   13   10  10   5   5   5
[31]  12      3      7    3    3    3   2   6   6   2
[41]   1      3      4    1    5    2   1   1   1   1
[51]   2      3      3    2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

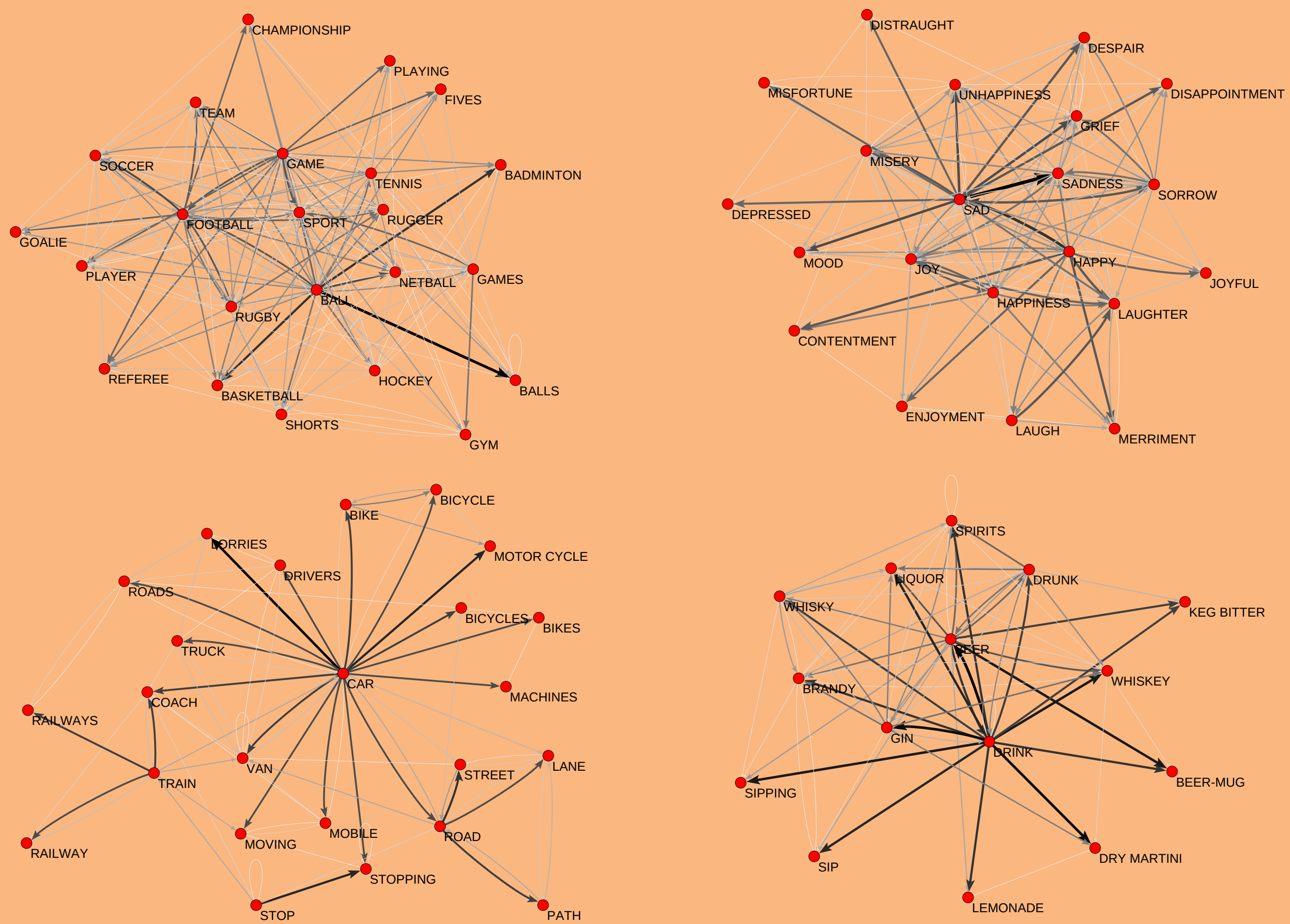
patent	date	author(s) and title
4083797	Apr 11, 1978	Oh. Nematic liquid crystal compositions
4113647	Sep 12, 1978	Coates, et al. Liquid crystalline materials
4118335	Oct 3, 1978	Krause, et al. Liquid crystalline materials of reduced viscosity
4130502	Dec 19, 1978	Eidenschink, et al. Liquid crystalline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

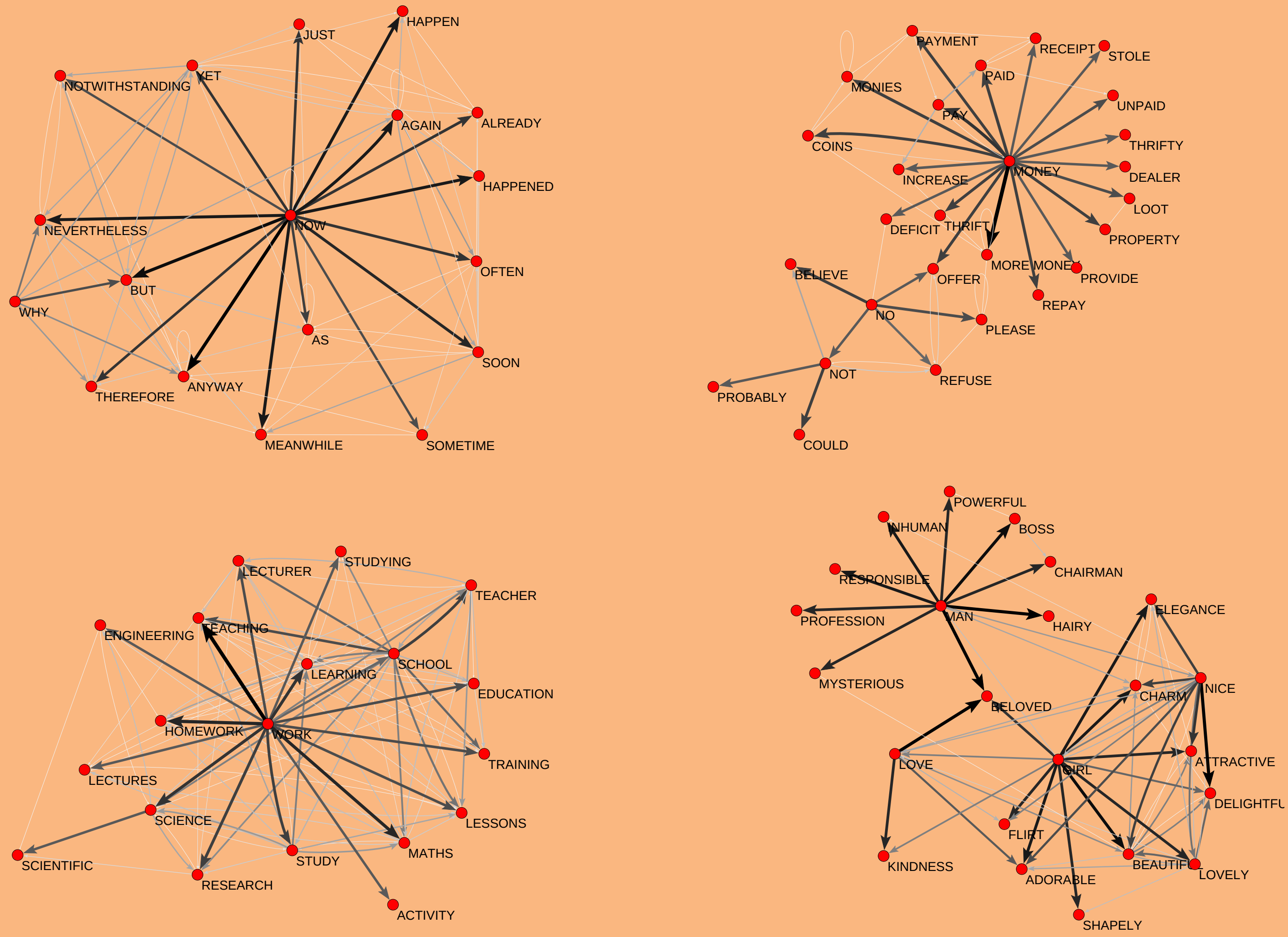
patent	date	author(s) and title
4514044	Apr 30, 1985	Gunjima, et al. 1-(Trans-4-alkylcyclohexyl)-2-(trans-4'-(p-sub stituted phenyl) cyclohexyl)ethane and liquid crystal mixture
4526704	Jul 2, 1985	Petrzilka, et al. Multiring liquid crystal esters
4550981	Nov 5, 1985	Petrzilka, et al. Liquid crystalline 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



... Islands – The Edinburgh Associative Thesaurus



Dense groups

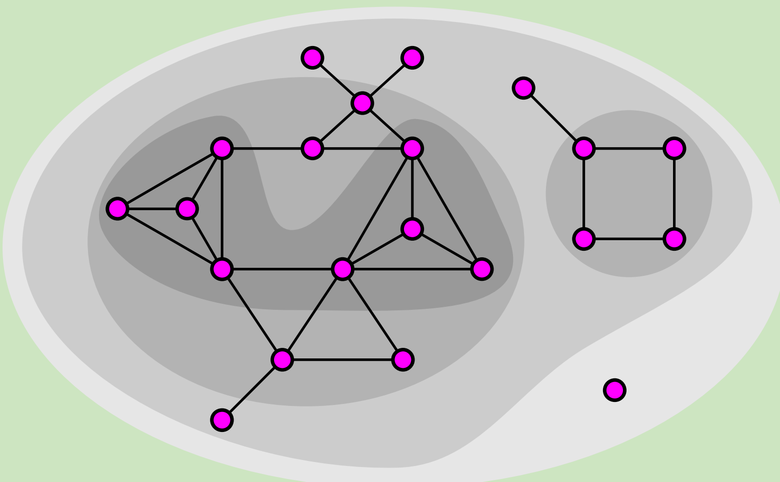
Several notions were proposed in attempts to formally describe dense groups in graphs.

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

s-plexes, LS sets, lambda sets, cores, ...

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

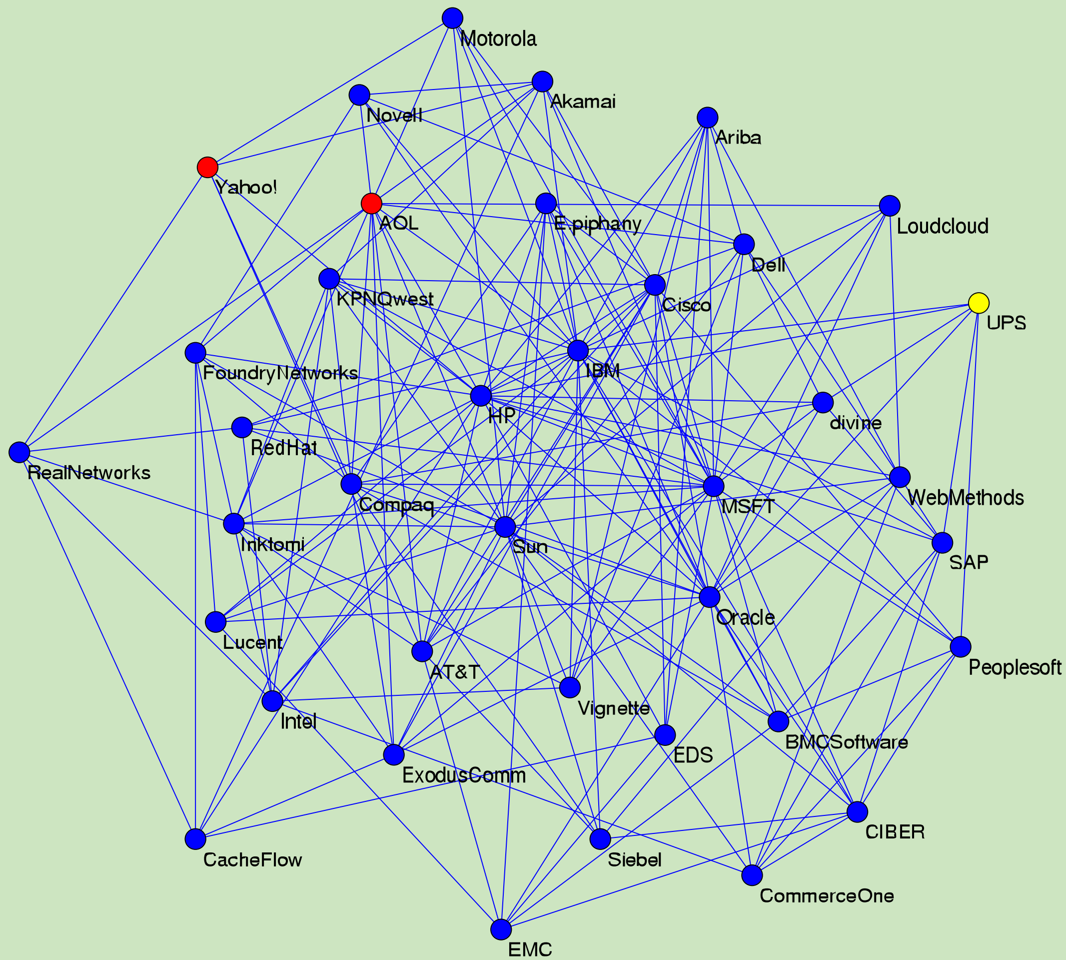
The cores, because they can be determined very efficiently, are one among few concepts that provide us with meaningful decompositions of large networks. We expect that different approaches to the analysis of large networks can be built on this basis. For example: we get the following bound on the chromatic 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 and (corrected) clustering coefficient, 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.

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

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

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; 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 $p(v, U) = p(v, N_U(v))$.

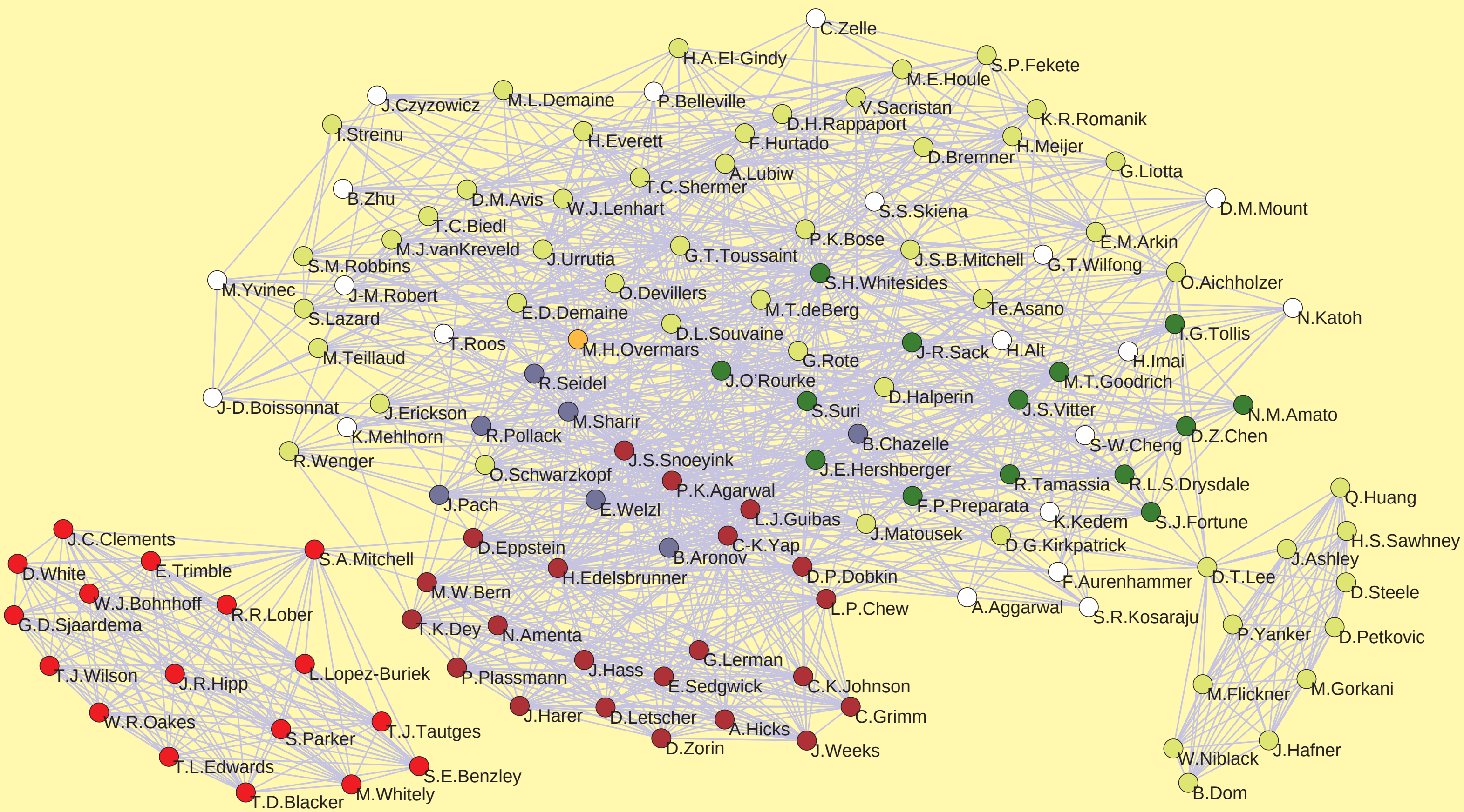
The degrees, p_S and p_M are local; but p_k is **not** local for $k \geq 4$. For a local p-function an $O(m \max(\Delta, \log n))$ algorithm for determining the p-core levels exists, assuming that $p(v, N_C(v))$ can be computed in $O(\deg_C(v))$.

For details see the paper.

Cores and generalized cores / Pajek commands

```
File/Network/Read  [Geom.net]
Net/Partitions/Core/All
Info/Partition
Operations/Extract from Network/Partition [13-*]
Draw/Draw-Partition
Layout/Energy/Kamada-Kawai
Options/Values of lines/Similarities
Layout/Energy/Kamada-Kawai
Operations/Extract from Network/Partition [21]
Draw
Layout/Energy/Kamada-Kawai
Options/Values of lines/Forget
Layout/Energy/Kamada-Kawai
[select Geom.net]
Net/Vector/PCore/Sum/All
Info/Vector
Vector/Make Partition/by Intervals/Selected Thresholds [45]
Info/Partition
Operations/Extract from Network/Partition [2]
Draw
Options/Values of lines/Similarities
Layout/Energy/Fruchterman-Reingold
```

Cores of orders 10–21 in Computational Geometry



p_S-core at level 46 of Geombib network