Photo: Stana, ***Križevec***

Analiza in prikaz velikih omrežij s programom Pajek

Andrej Mrvar (FDV)
Vladimir Batagelj (FMF)

Biostatistični center, 4. oktober, 2005

Velika omrežja

Običajna analiza podatkov (statistika): proučujemo lastnosti enot (dohodek, starost,...).

Analiza omrežij: proučujemo relacije med enotami – poleg enot obravnavamo še relacije (povezave).

Velika omrežja – omrežja z nekaj sto tisoč enotami (točkami) in povezavami.

V večini primerov so velika omrežja redka (povezav ni veliko več kot točk).

Problemi pri analizi velikih omrežij:

- časovna zahtevnost izvajanja postopkov
- vizualizacija

Velikih omrežij ponavadi ne analiziramo ali prikazujemo v celoti, ampak prej s primernimi analizami poiščemo zanimiva podomrežja.

Primeri velikih omrežij

- socialna omrežja
  - rodovniki (genealogije);
  - omrežja razširjanja inovacij ali bolezni (HIV);
  - omrežje telefonskih klicev znotraj izbrane množice številk (raziskave kriminala, primer b.net);
  - članstva v nadzornih svetih, omrežja lastništev delnic;
  - trgovanje med organizacijami, državami;
  - omrežja citiranj, omrežja soavtorstev;
  - računalniška omrežja (lokalna omrežja, Internet, povezave med predstavitvenimi stranmi);
- organske molekule v kemiji (DNA, primer dna.net);
- *protein interaction networks*;
- omrežja pridobljena iz slovarjev in drugih besedil;
- transportna omrežja (letalske povezave USAir.net, vodovodna omrežja . . .).

Nekaj zanimivih rezultatov

Omrežja z zelo velikim številom točk imajo velikokrat kratke najkrajše poti med točkami.

Primer:

- Povprečna dolžina najkrajše poti omrežja WWW, z več kot 800 milijoni točk, je okrog 19.

  Albert, R., Jeong, H., and Barabasi, A.-L. (1999):
  Diameter of the World-Wide Web. *Nature*, **401**, 130-131.
  http://www.nd.edu/~networks/Papers/401130A0.pdf

- Ocenjuje se, da je v socialnih omrežjih (koga poznaš) s preko 6 milijardami posameznikov, povprečna dolžina najkrajše poti med dvema posameznikoma okrog 6.

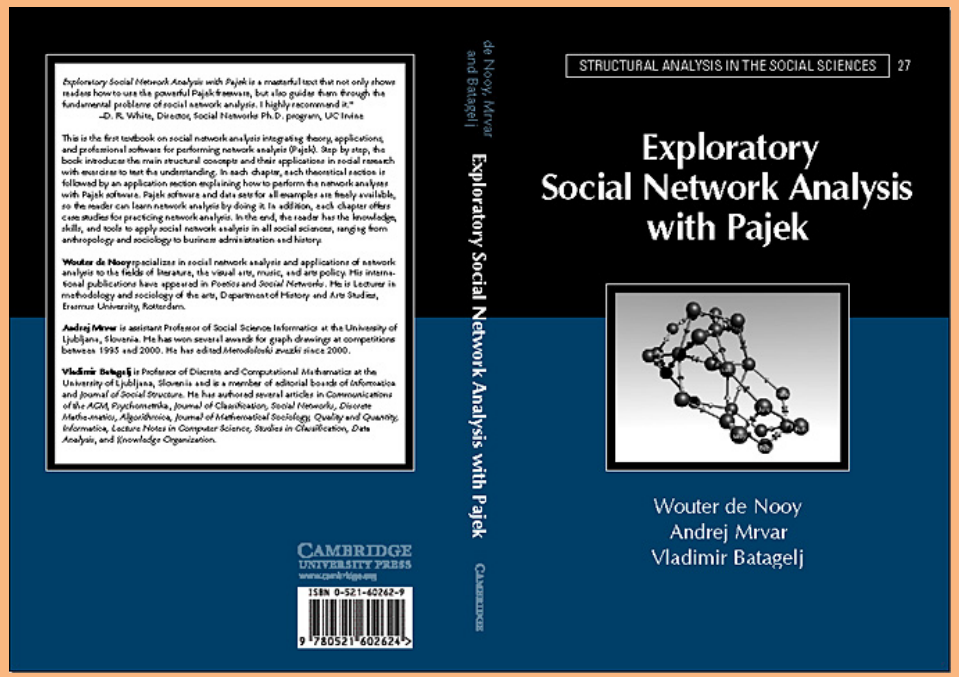
  Milgram, S. (1967): The small-world problem.
  Psychol. Today, **2**, 60-67.

Pajek

Pajek je programski paket za Windows 95/98/NT/2000/XP, ki omogoča analizo *velikih omrežij*.

Program je prosto dostopen na naslovu:

http://vlado.fmf.uni-lj.si/pub/networks/pajek/

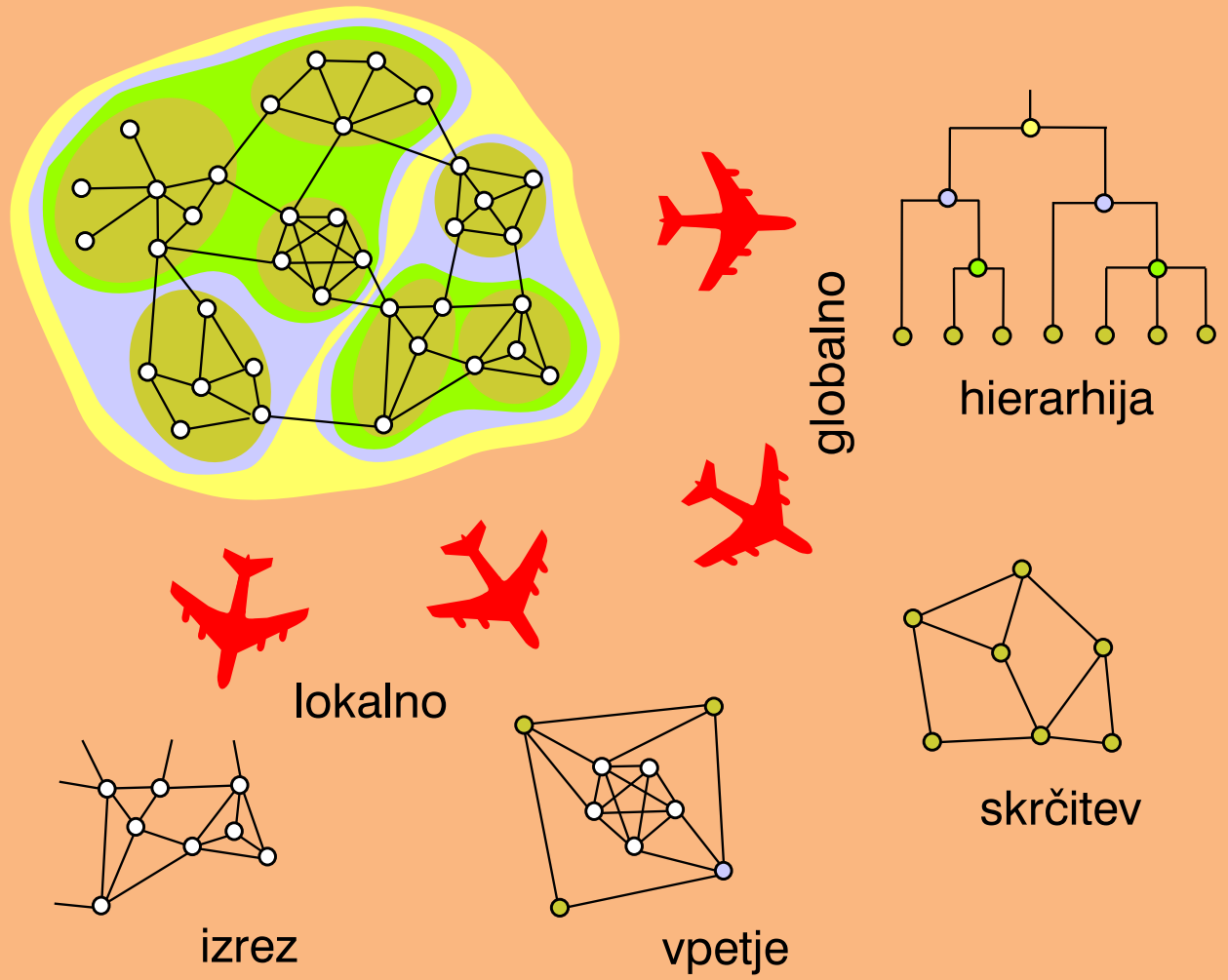


de Nooy, Mrvar, Batagelj: *Exploratory Social Network Analysis with Pajek*

Cilji pri zasnovi programa

Glavni cilji pri zasnovi programa **Pajek** so:

- podpreti *abstrakcijo* z (rekurzivno) razčlenitvijo velikega omrežja na več manjših omrežij, ki jih lahko nadalje analiziramo z uporabo običajnih metod;
- ponuditi uporabniku močna orodja za *vizualizacijo* omrežij;
- vgraditi večje število *učinkovitih* algoritmov za analizo obsežnih omrežij.


globalno
hierarhija
lokalno
skrčitev
izrez
vpetje

Podatkovne strukture

Izvedbe algoritmov v programu **`Pajek`** so oprte na šest podatkovnih struktur:

- *omrežje* – točke in povezave;
- *razbitje* – nominalne ali ordinalne lastnosti točk;
- *vektor* – številske lastnosti točk;
- *permutacija* – preureditev točk;
- *skupina* – podmnožica točk (npr. en razred iz razbitja);
- *hierarhija* – hierarhična razvrstitev točk omrežja.

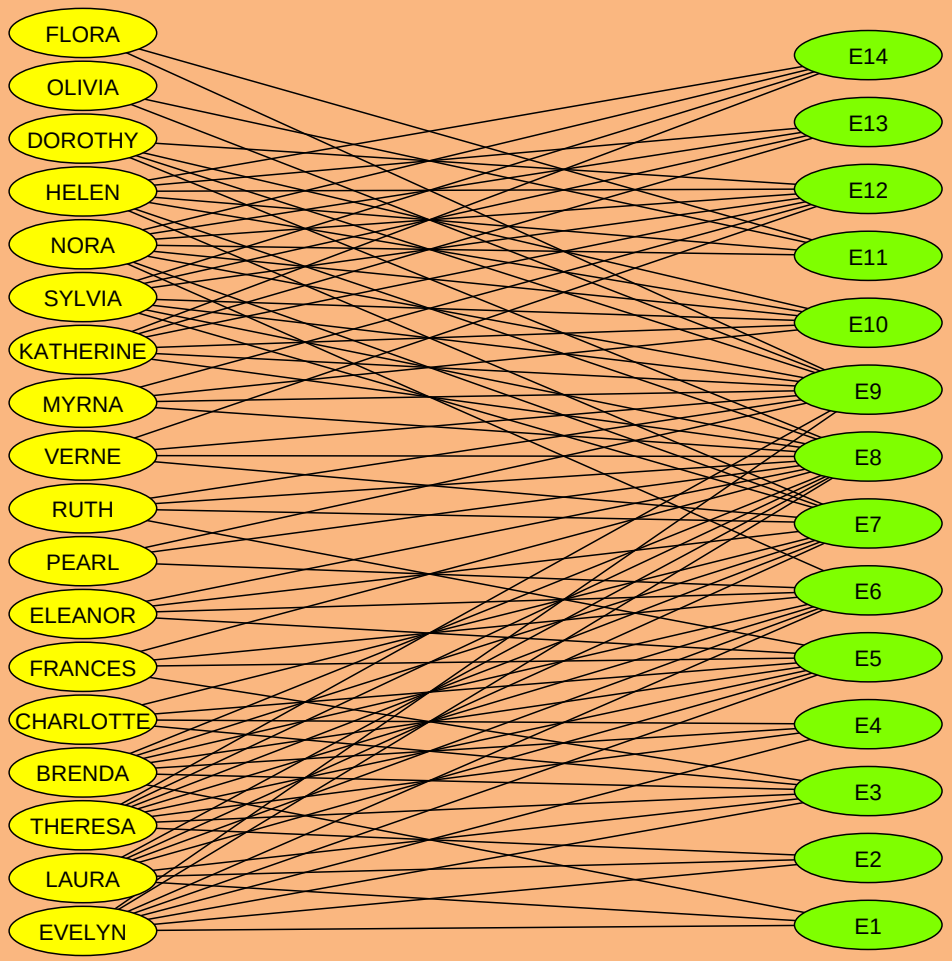
Dvovrstna omrežja

Dvovrstno omrežje sestavljata dve množici enot (npr. osebe in dogodki), relacija pa ti dve množici povezuje, npr. vključenost oseb v družabne dogodke. Primerov takih omrežij je še ogromno:

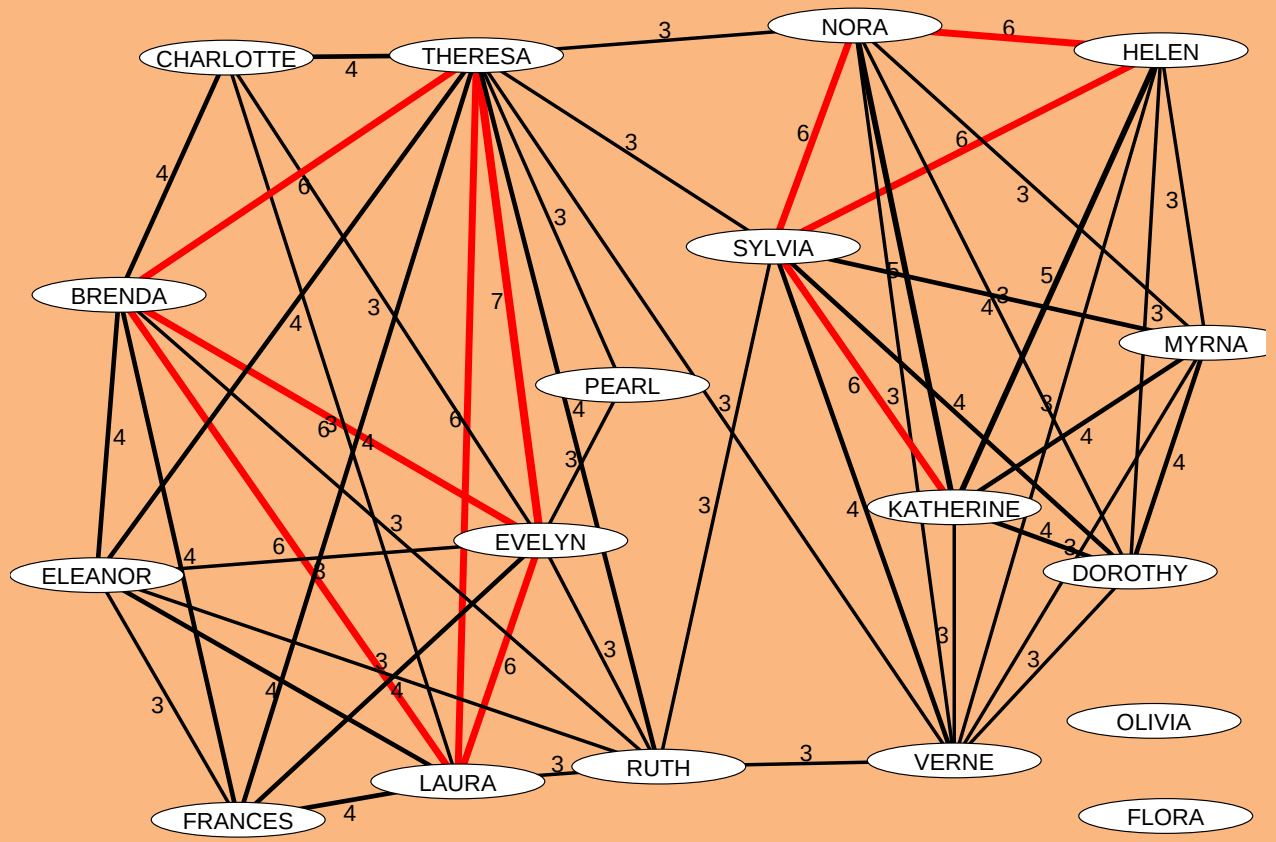
- *Omrežje citiranj*, katerega prva množica točk so avtorji, druga množica točk so članki, povezave med avtorji in članki pa povedo, kateri avtor je citiral kateri članek.
- *Omrežje nakupov*, v katerem so prva množica kupci, druga množica artikli, povezava pa pove, kateri artikel je kupec kupil.
- *Bralci in revije, ki jih osebe berejo* – primer 124 slovenskih revij (slike SVG)

Tu je analiza omrežij najbližje ’rudarjenju’ (data mining).

Primer...


FLORA
OLIVIA
DOROTHY
HELEN
NORA
SYLVIA
KATHERINE
MYRNA
VERNE
RUTH
PEARL
ELEANOR
FRANCES
CHARLOTTE
BRENDA
THERESA
LAURA
EVELYN
E14
E13
E12
E11
E10
E9
E8
E7
E6
E5
E4
E3
E2
E1

...primer


NORA
HELEN
CHARLOTTE
THERESA
SYLVIA
BRENDA
MYRNA
PEARL
KATHERINE
EVELYN
ELEANOR
DOROTHY
OLIVIA
VERNE
RUTH
LAURA
FRANCES
FLORA

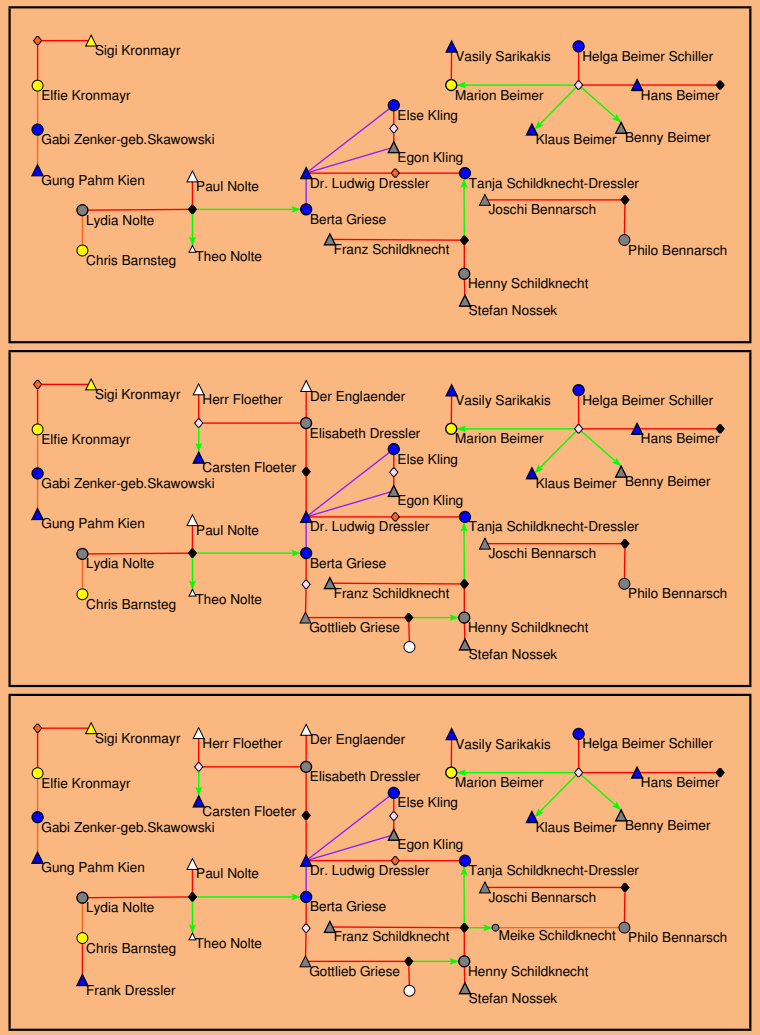
Časovna omrežja

Omrežja, ki se spreminjajo skozi čas. Primeri časovnih omrežij

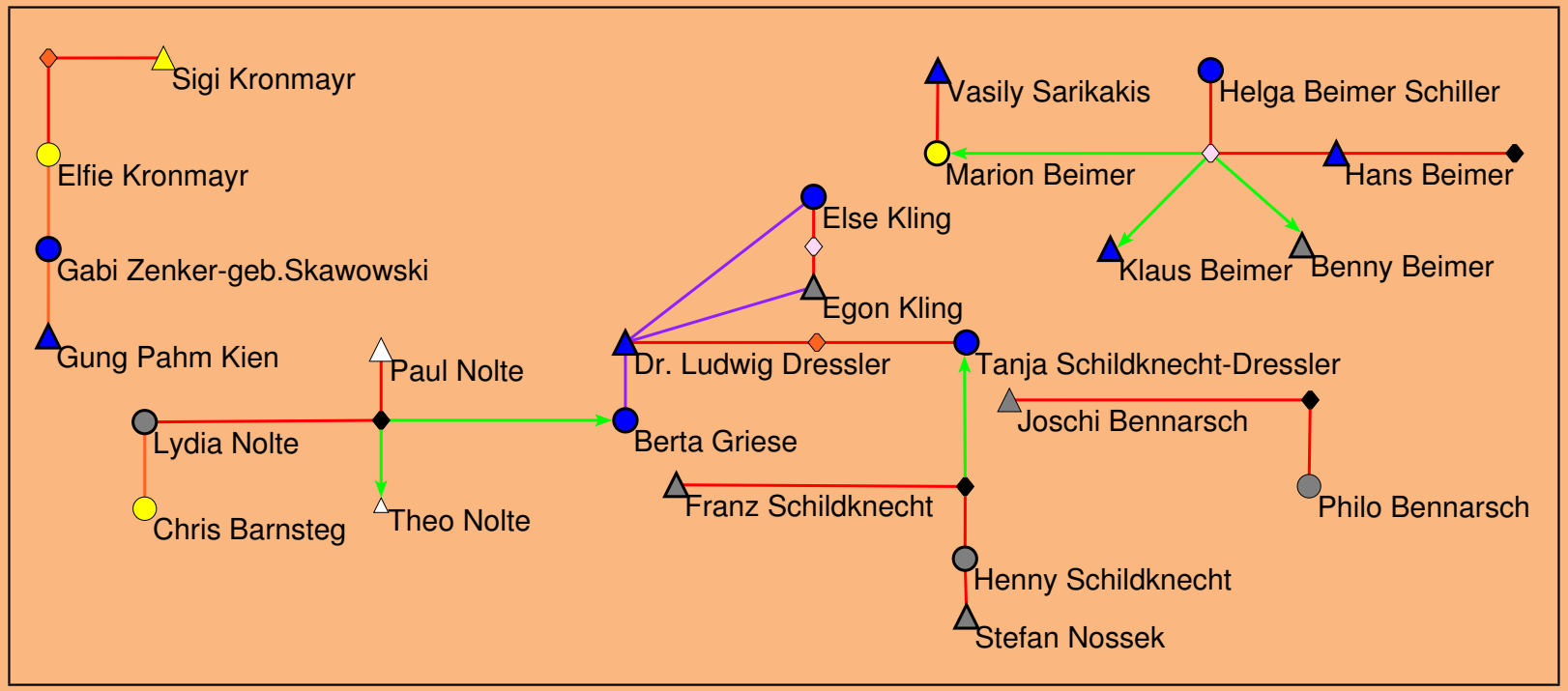
- omrežje prijateljstev v razredu skozi osemletko,
- omrežje telefonskih klicev znotraj izbrane množice številk,
- omrežje citiranj v člankih z izbranega področja,
- omrežje prehodov žoge med igralci na neki tekmi z žogo,
- omrežja okuženih z virusom HIV,
- razmerja med igralci v različnih delih televizijskih nadaljevank,
- rojevanja, umiranja in poroke v primeru rodovnikov. . .

Odnosi med igralci v nadaljevanki LindenStrasse

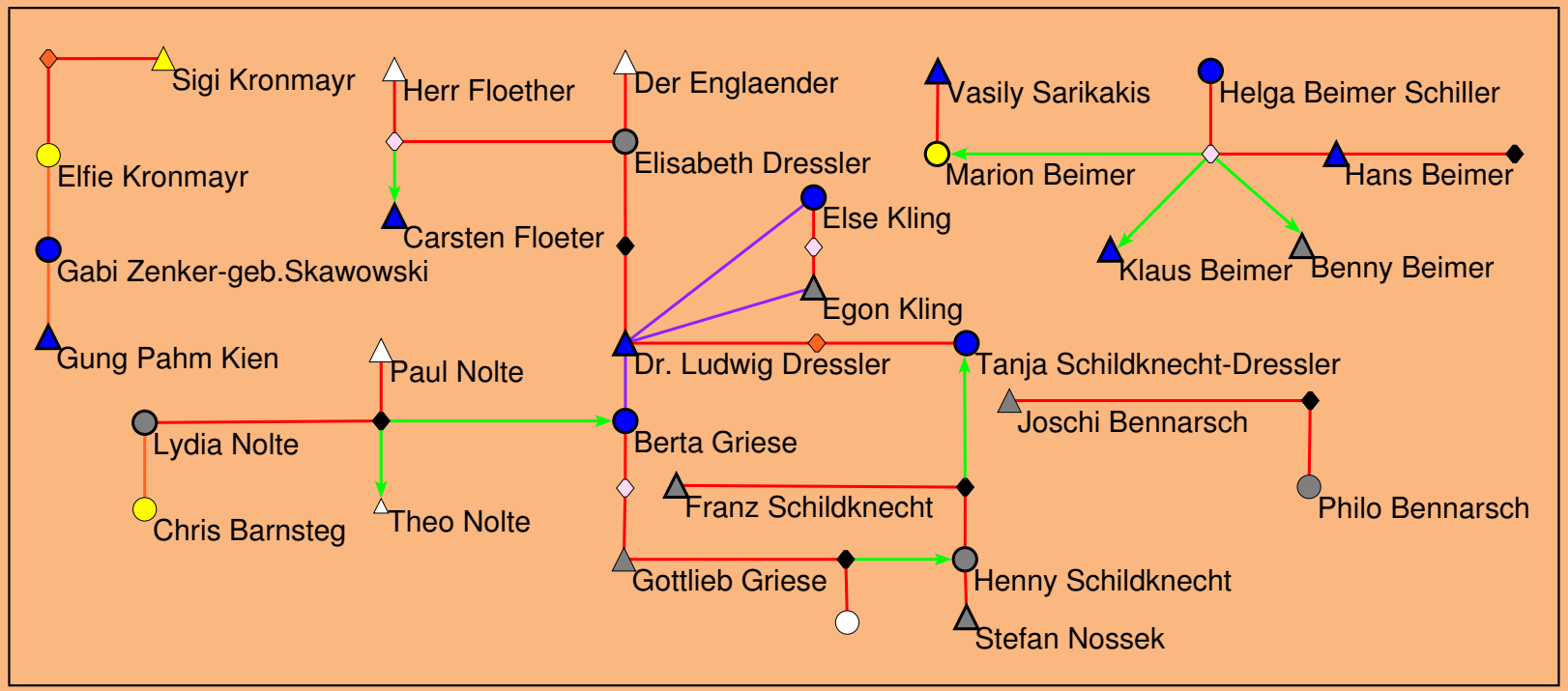
v 5., 6. in 7. delu



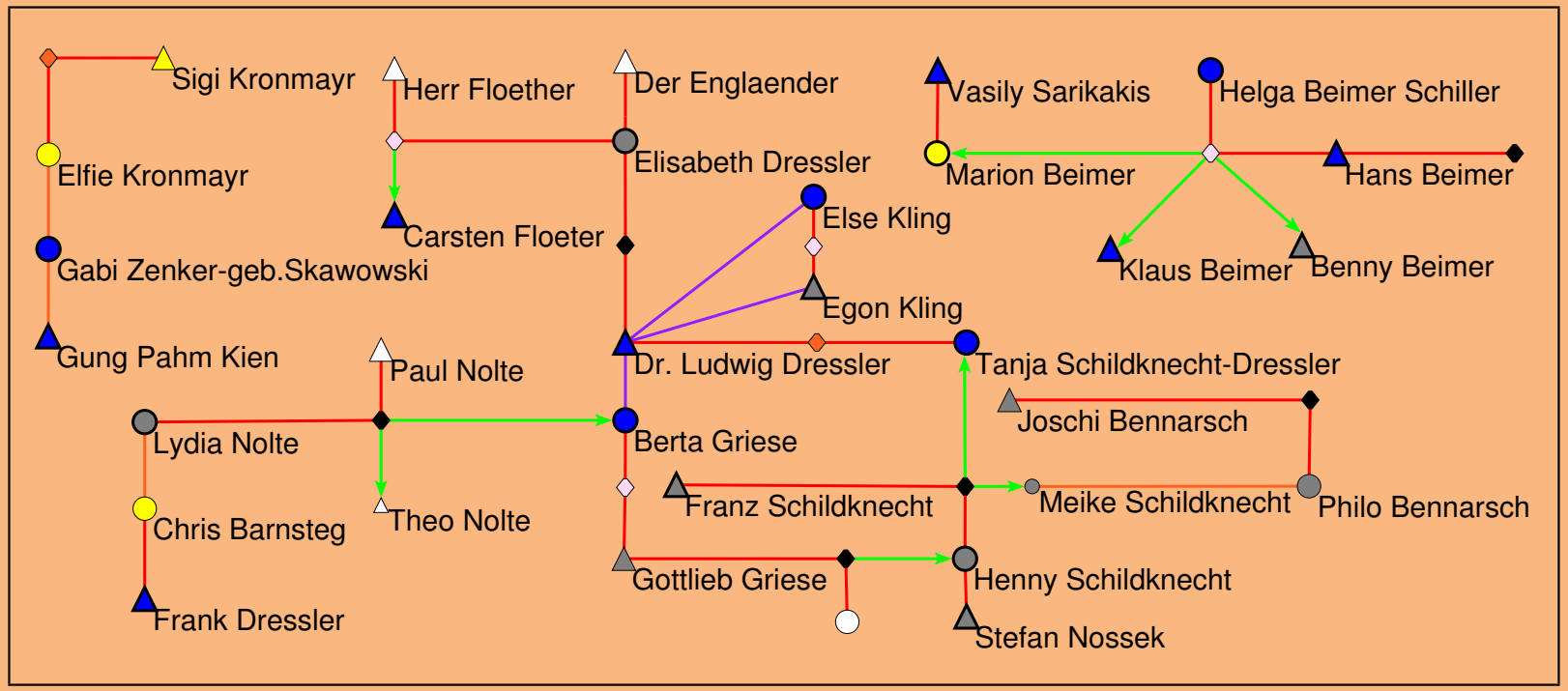
5. del


Sigi Kronmayr
Elfie Kronmayr
Gabi Zenker-geb.Skawowski
Gung Pahm Kien
Lydia Nolte
Chris Barnsteg
Paul Nolte
Theo Nolte
Dr. Ludwig Dressler
Else Kling
Egon Kling
Berta Griese
Vasily Sarikakis
Helga Beimer Schiller
Marion Beimer
Hans Beimer
Klaus Beimer
Benny Beimer
Tanja Schildknecht-Dressler
Joschi Bennarsch
Philo Bennarsch
Franz Schildknecht
Henny Schildknecht
Stefan Nossek

6. del


Sigi Kronmayr
Herr Floether
Der Englaender
Vasily Sarikakis
Helga Beimer Schiller
Elfie Kronmayr
Elisabeth Dressler
Marion Beimer
Hans Beimer
Carsten Floeter
Else Kling
Gabi Zenker-geb.Skawowski
Klaus Beimer
Benny Beimer
Egon Kling
Gung Pahm Kien
Paul Nolte
Dr. Ludwig Dressler
Tanja Schildknecht-Dressler
Joschi Bennarsch
Lydia Nolte
Berta Griese
Philo Bennarsch
Franz Schildknecht
Chris Barnsteg
Theo Nolte
Gottlieb Griese
Henny Schildknecht
Stefan Nossek

7. del


Sigi Kronmayr
Herr Floether
Der Englaender
Vasily Sarikakis
Helga Beimer Schiller
Elfie Kronmayr
Elisabeth Dressler
Marion Beimer
Hans Beimer
Carsten Floeter
Else Kling
Gabi Zenker-geb.Skawowski
Klaus Beimer
Benny Beimer
Egon Kling
Gung Pahm Kien
Paul Nolte
Dr. Ludwig Dressler
Tanja Schildknecht-Dressler
Joschi Bennarsch
Lydia Nolte
Berta Griese
Franz Schildknecht
Meike Schildknecht
Philo Bennarsch
Chris Barnsteg
Theo Nolte
Gottlieb Griese
Henny Schildknecht
Frank Dressler
Stefan Nossek

Večrelacijska omrežja

Omrežja kjer so na isti množici enot izmerjene različne relacije. Primer: Sampsonovi menihi (prijateljstvo, zaupanje, nepriljubljenost, nezaupanje...) Možne kombinacije: podatki za Sampsonove menihe so primer časovnega večrelacijskega omrežja – relacije so izmerjene v več časovnih trenutkih.

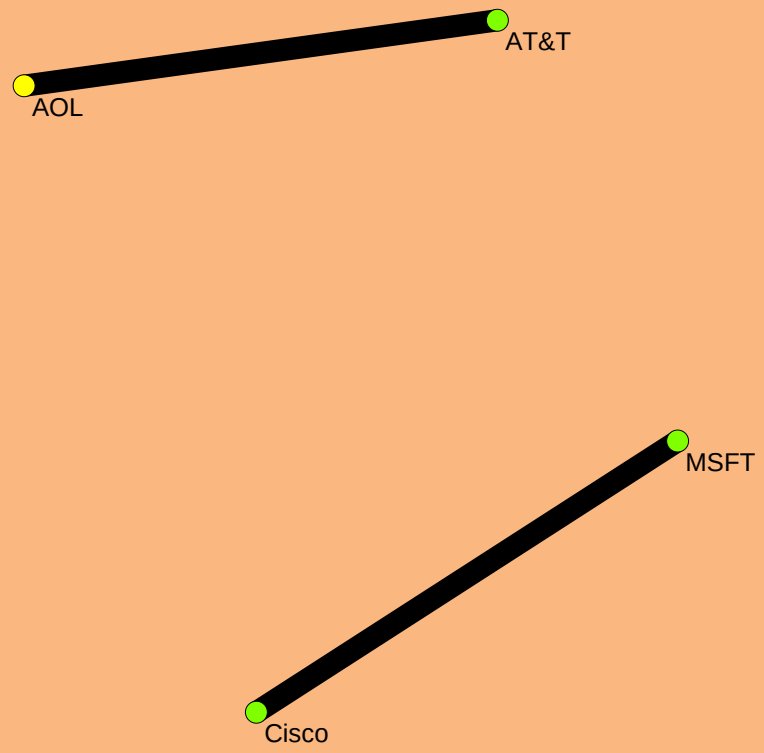
Primeri

- flor.net
- 1crn.net,
- usair.net + usair.clu,
- import.net + cont.clu,
- football.net,
- davis.net.

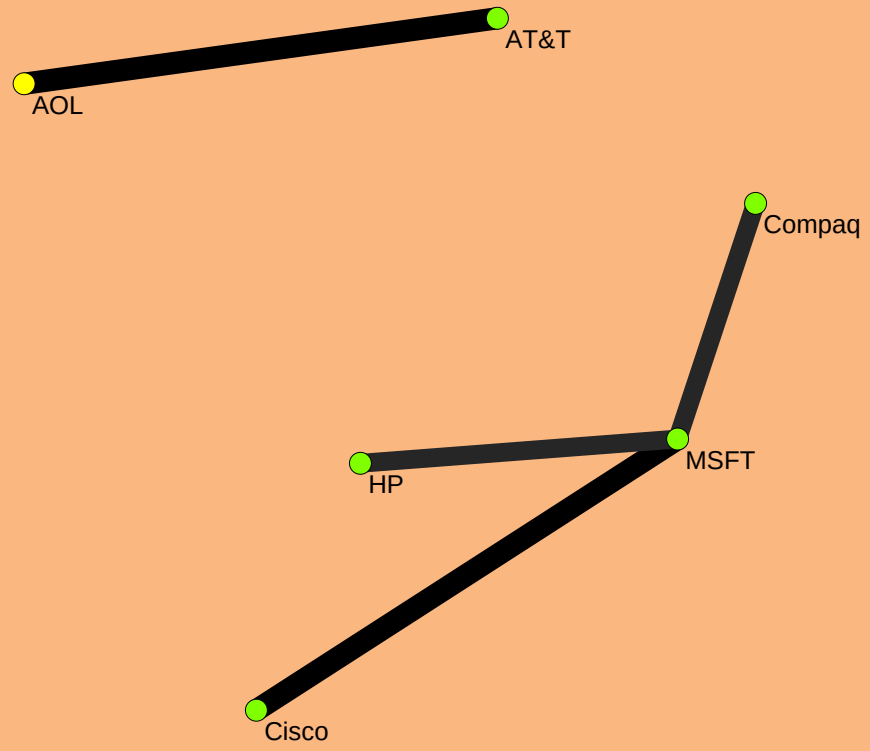
Primer: Internet Industry Partnerships

- Vsaka točka v omrežju predstavlja neko podjetje, ki se ukvarja z internetom. Podjetji sta povezani, če sta objavili skupna vlaganja, strateški plan ali drugo vrsto partnerstva.
- Omrežje predstavlja samo podmnožico celotne industrije na področju interneta v obdobju od 1998 do 2001. Vsebuje 219 točk (podjetij) in 631 povezav.
- Podjetja so razdeljena v tri skupine glede na njihovo vrsto: rumena = vsebina, rdeča = trgovina, zelena = infrastruktura.
- http://www.orgnet.com/netindustry.html

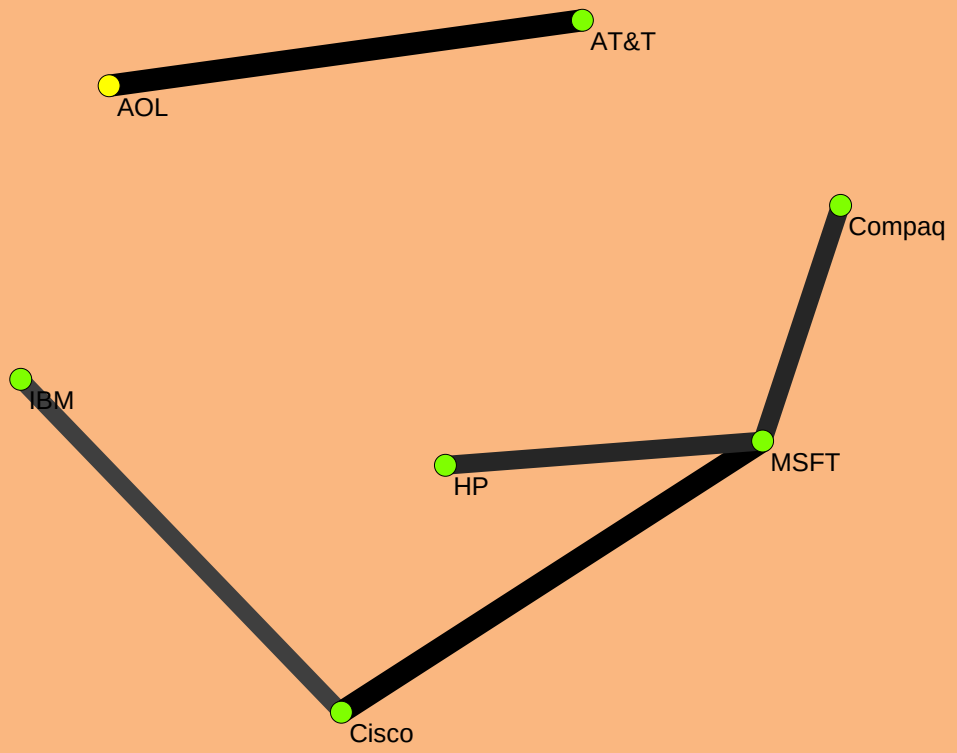
Povezave, ki pripadajo vsaj 12 trikotnikom



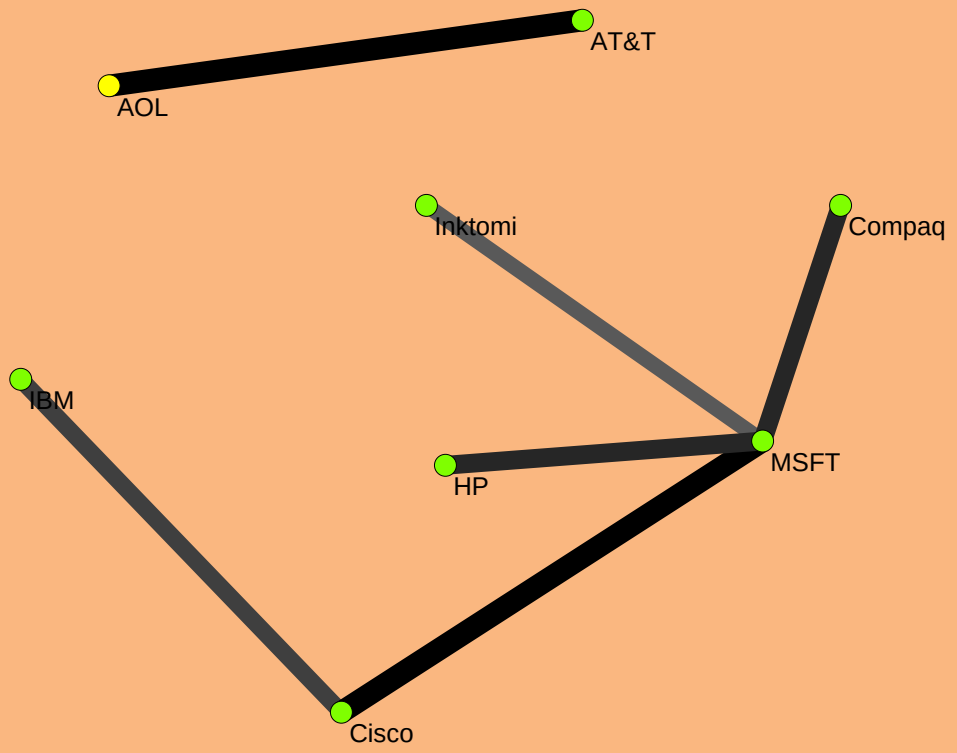
Povezave, ki pripadajo vsaj 11 trikotnikom



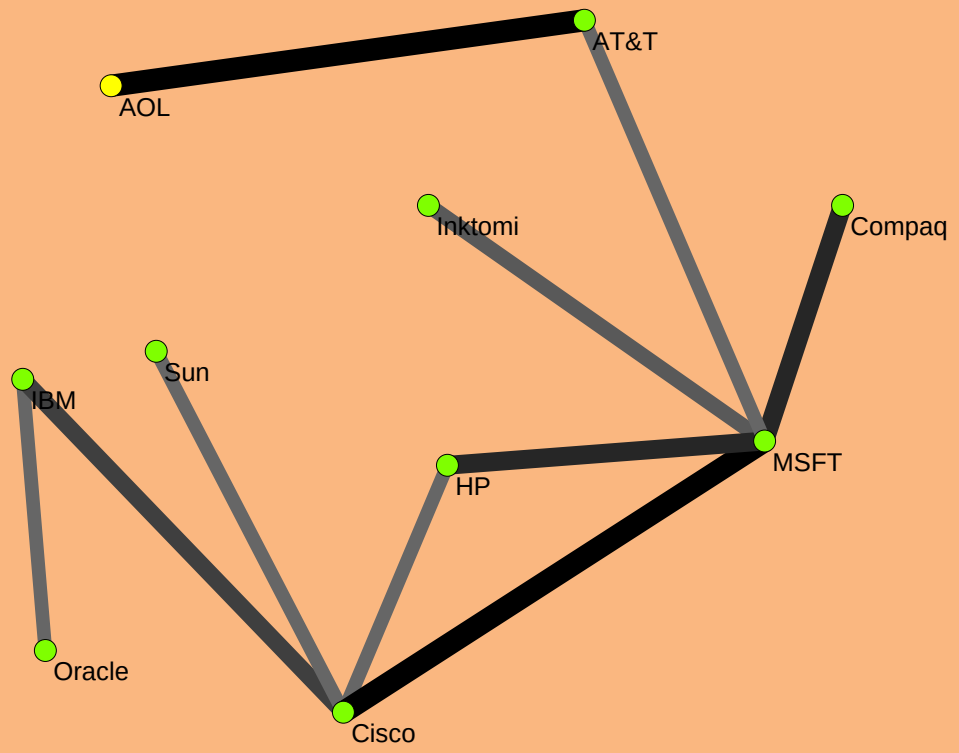
Povezave, ki pripadajo vsaj 10 trikotnikom



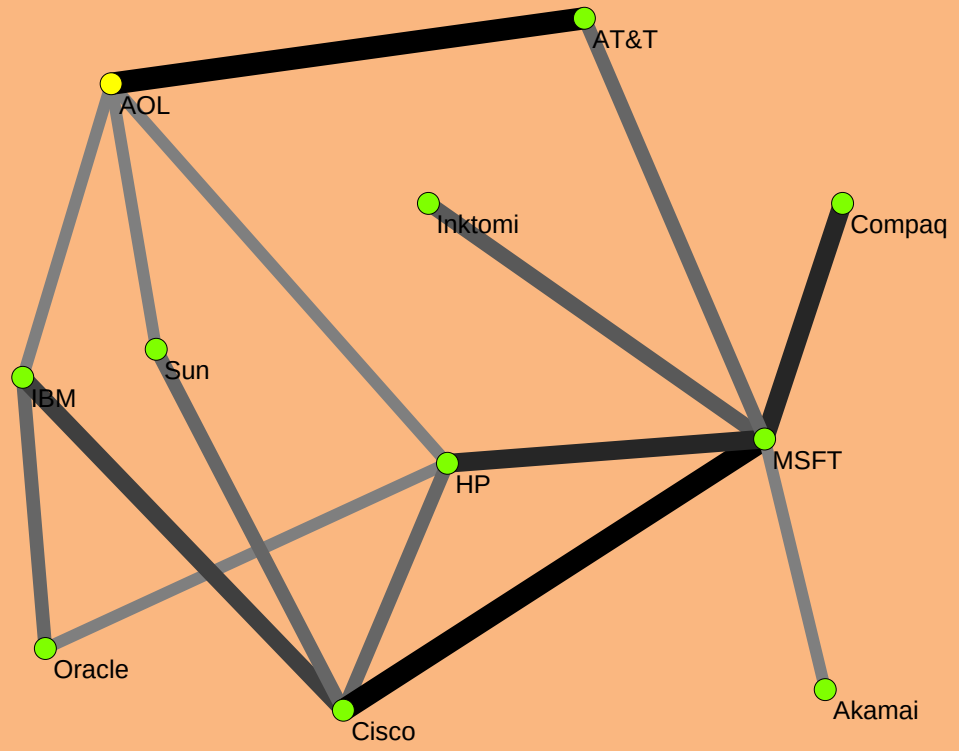
Povezave, ki pripadajo vsaj 9 trikotnikom



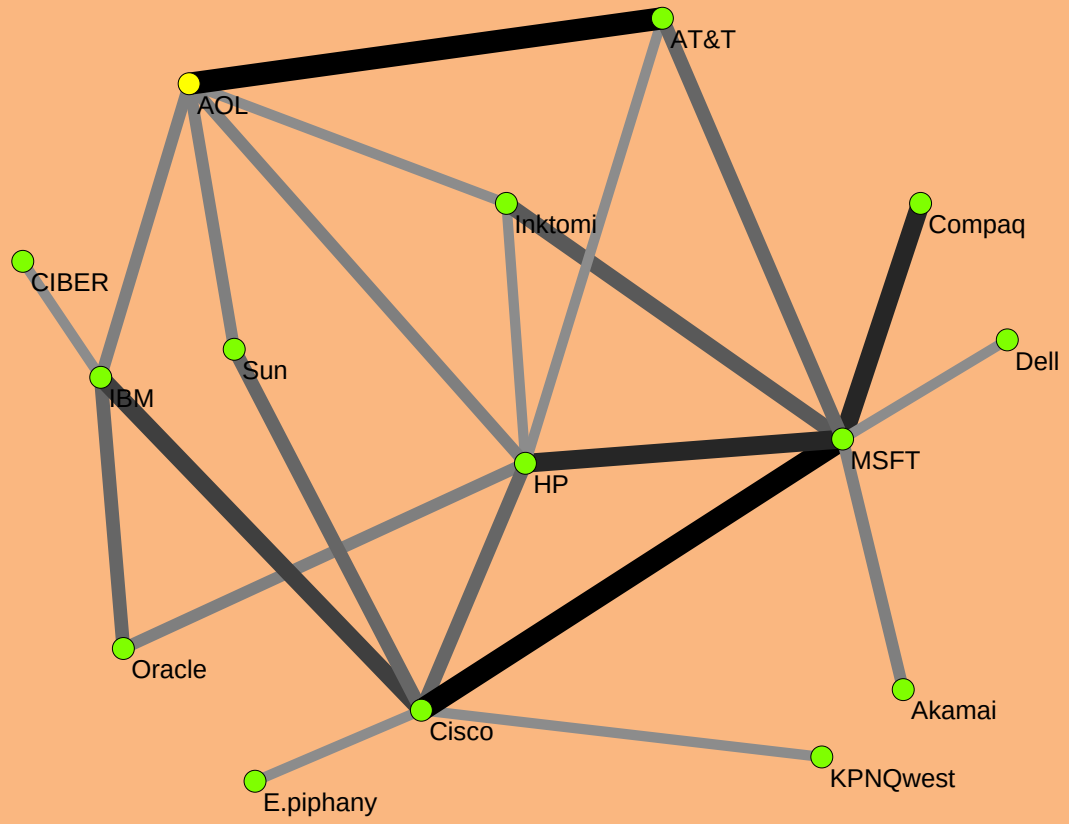
Povezave, ki pripadajo vsaj 8 trikotnikom



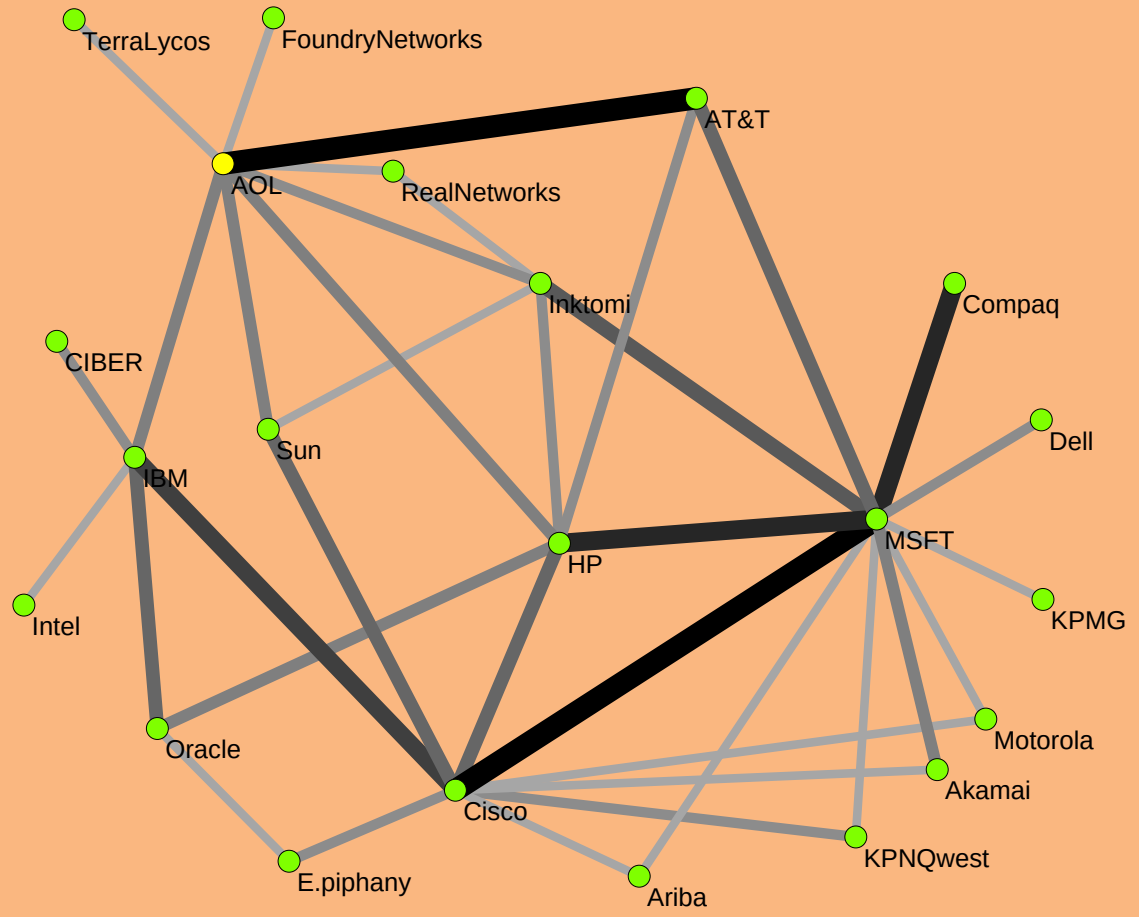
Povezave, ki pripadajo vsaj 7 trikotnikom



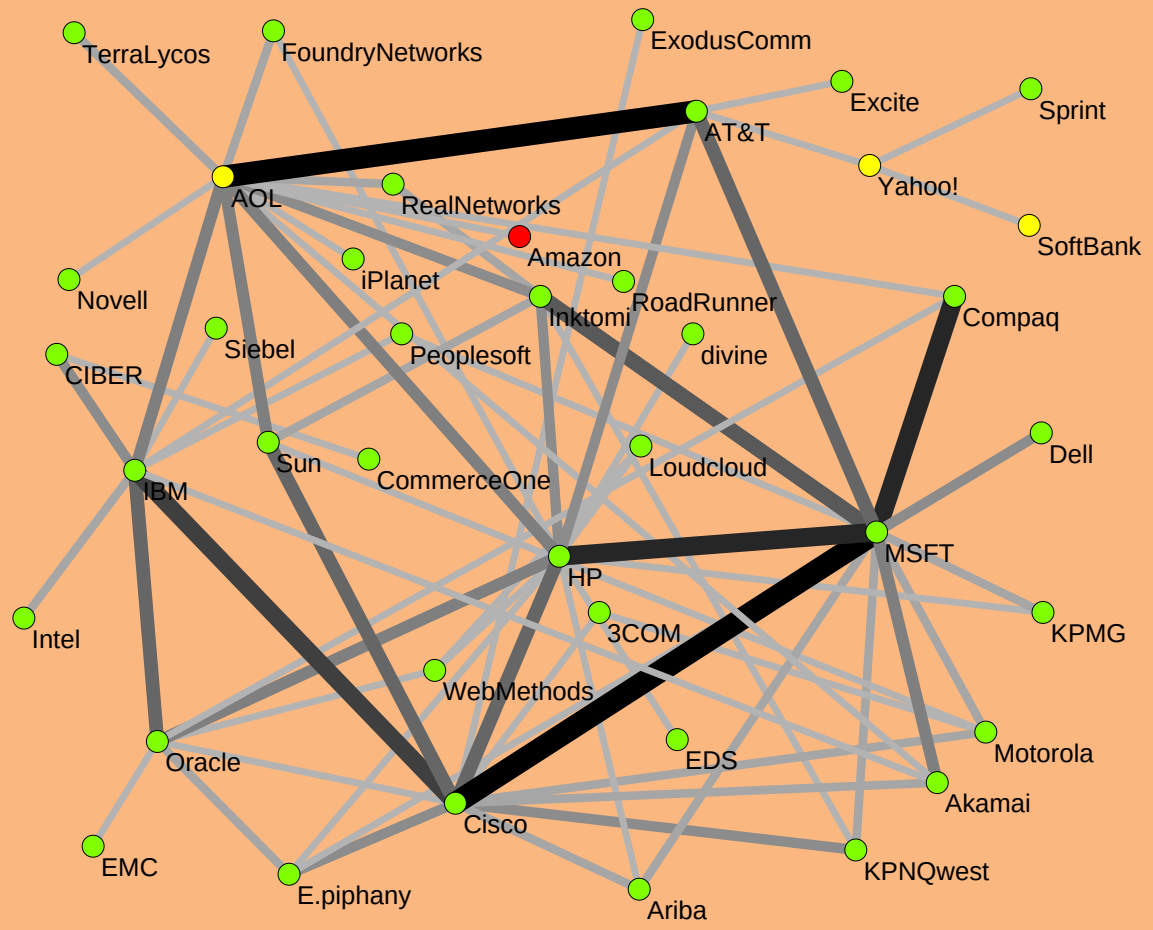
Povezave, ki pripadajo vsaj 6 trikotnikom



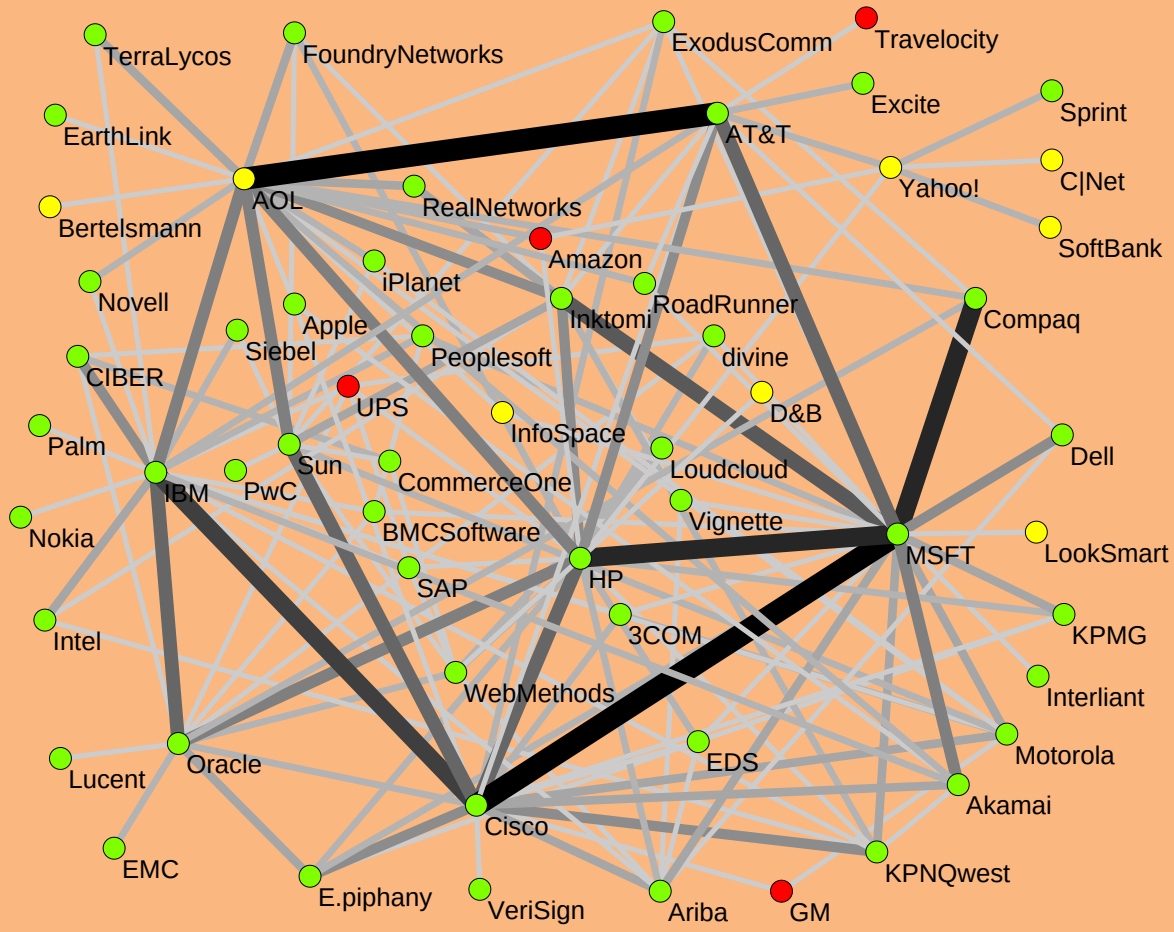
Povezave, ki pripadajo vsaj 5 trikotnikom



Povezave, ki pripadajo vsaj 4 trikotnikom



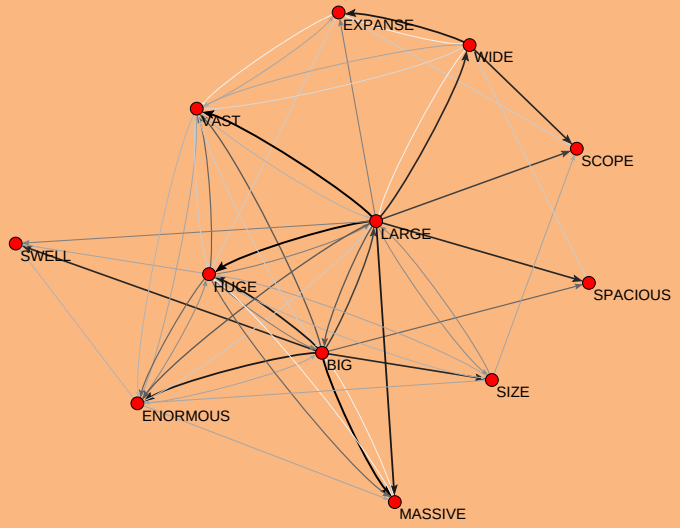
Povezave, ki pripadajo vsaj 3 trikotnikom

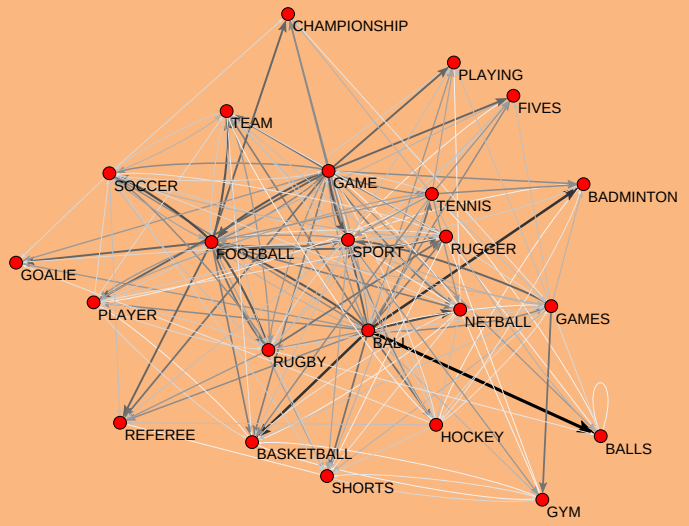


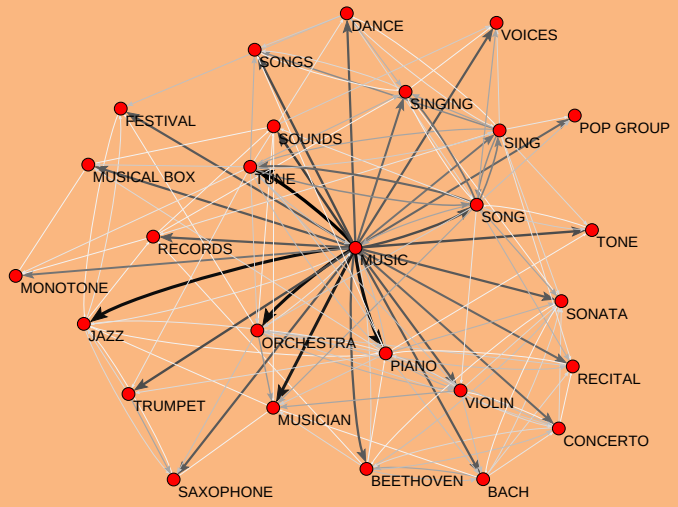
Primer: The Edinburgh Associative Thesaurus

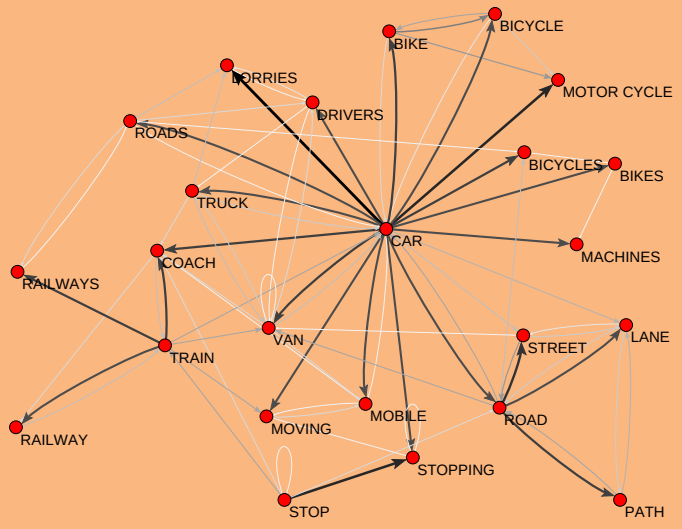
- Na britanskih univerzah so od junija 1968 do maja 1971 anketirali tamkajšnje študente. Za vsako od besed iz danega seznama so morali napisati besedo, ki jim prva pride na misel.
- Dobili so veliko omrežje, sestavljeno iz 23219 točk (besede) in 325624 usmerjenih povezav, med katerimi je tudi 564 zank.

Izbrane teme v EAT

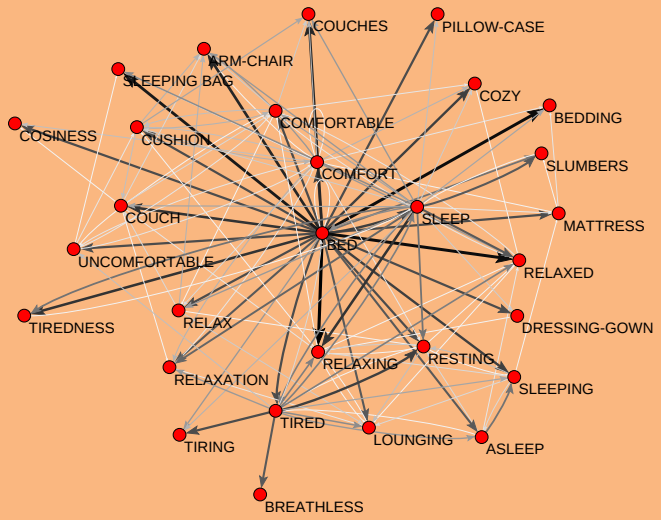


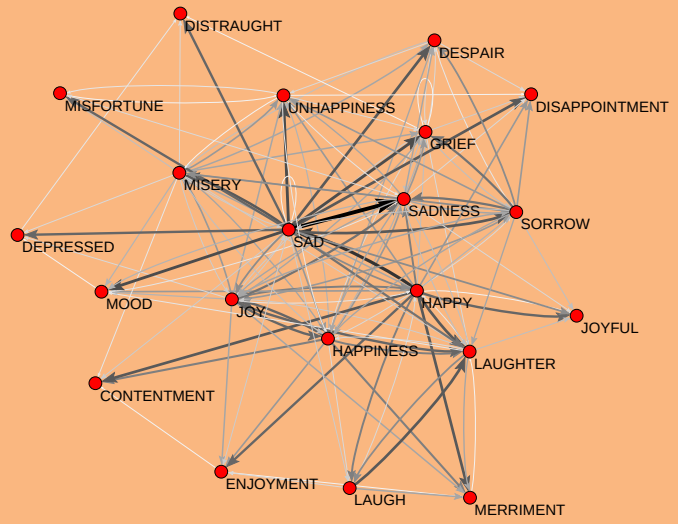


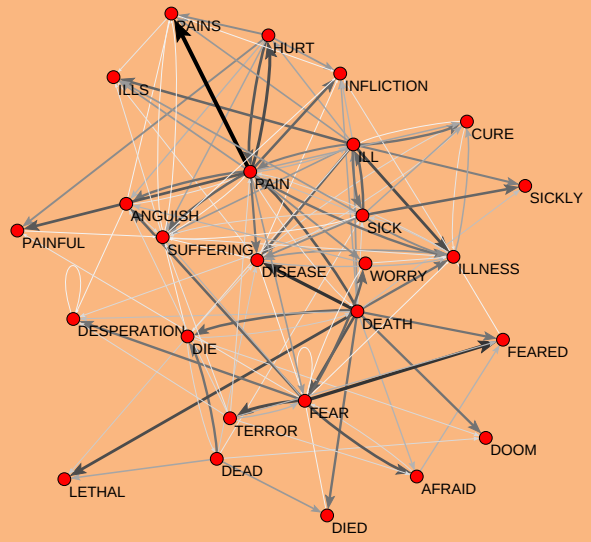


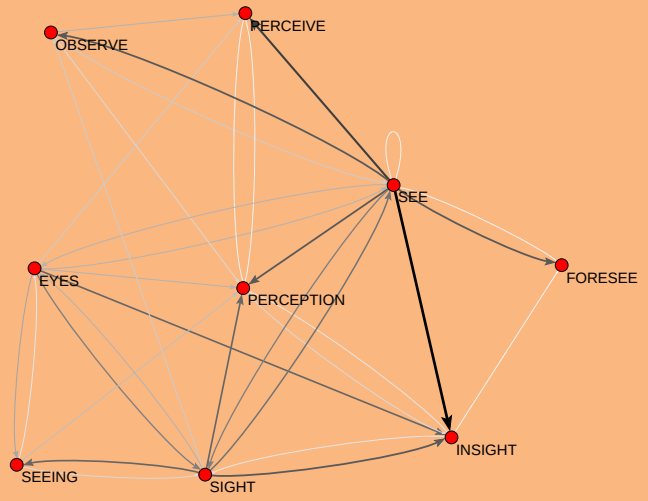


Izbrane teme v EAT

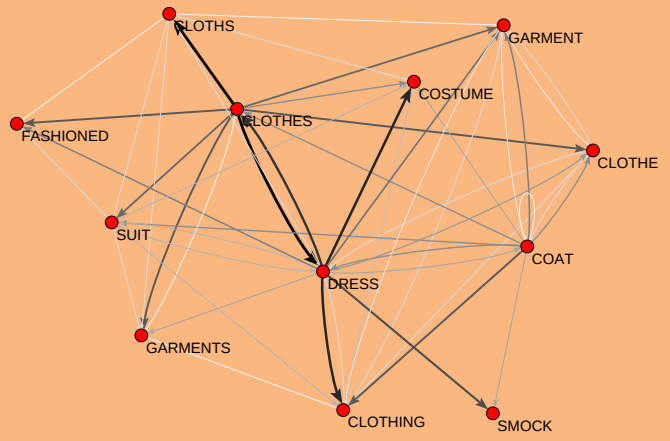


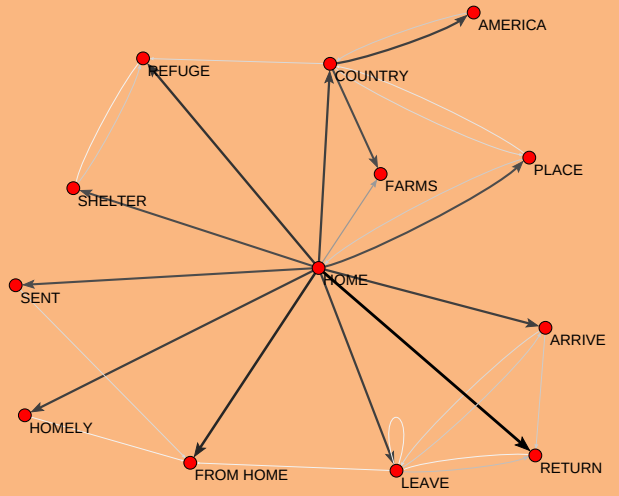


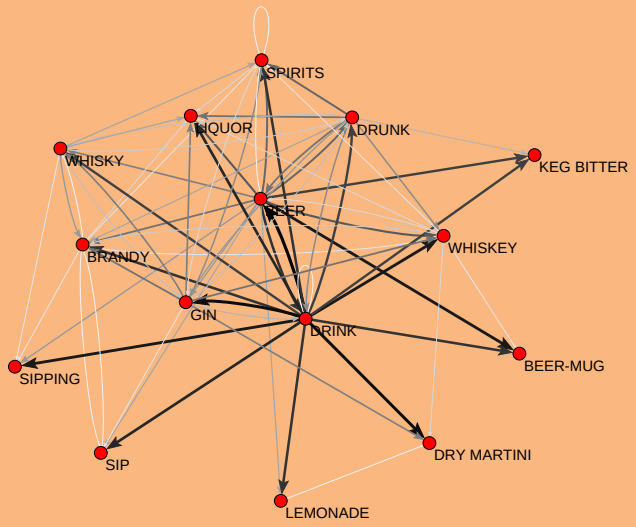


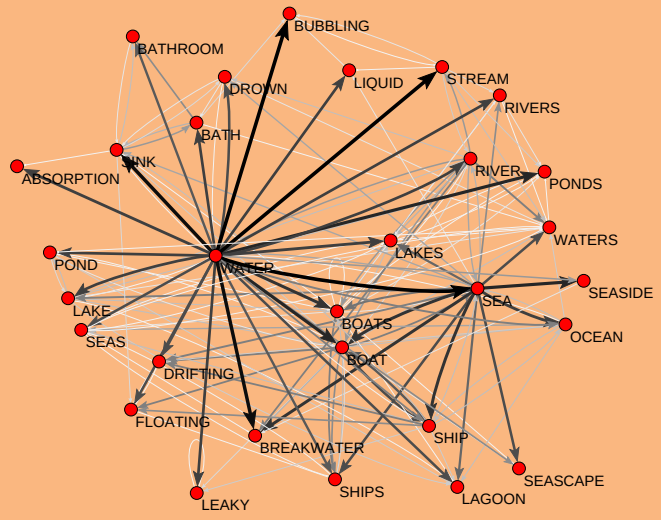


Izbrane teme v EAT

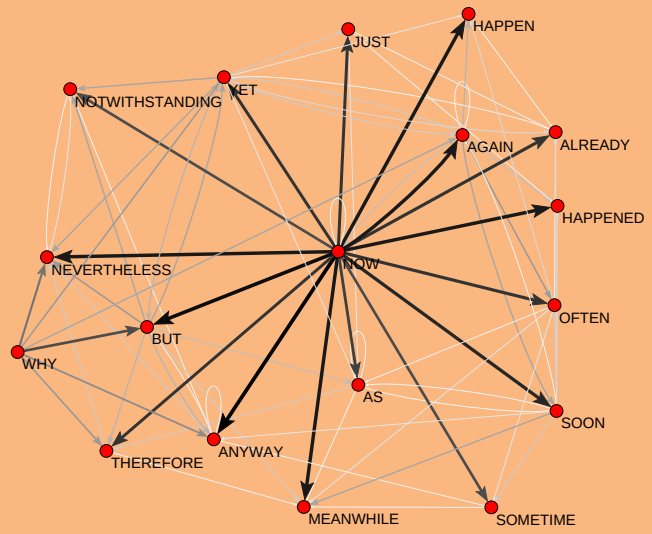


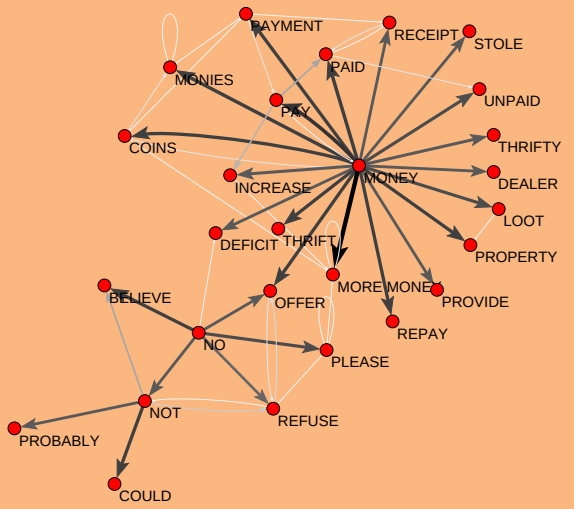


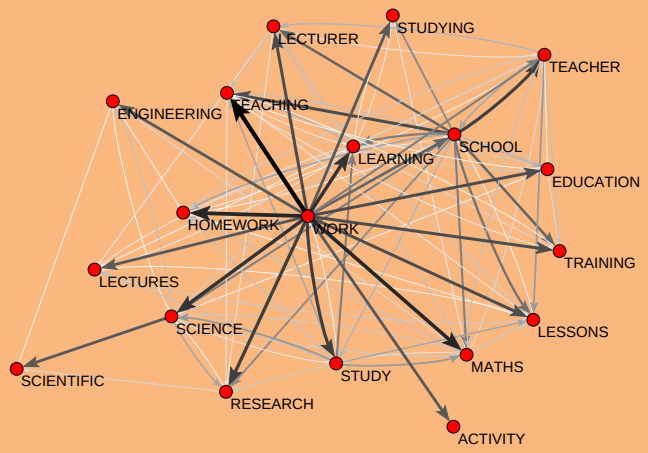


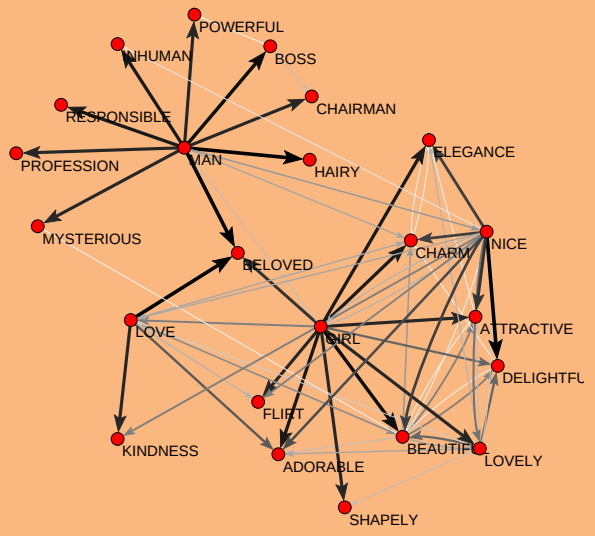


Izbrane teme v EAT









Primer: The Knuth's English Dictionary

- Knuthov slovar angleških besed je omrežje, ki ima:
  - 52652 točk – angleške besede dolžine od 2 do 8 črk
  - 89038 povezav – besedi sta sosedni, če lahko dobimo eno iz druge, tako da zamenjamo, odstranimo ali dodamo eno črko

Izbrane skupine v Knuthovem slovarju