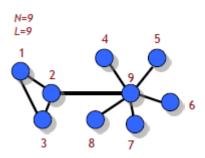
# METODE INTELIGENTE DE REZOLVARE A PROBLEMELOR REALE

Laura Dioşan Tema 8

#### Rețele complexe

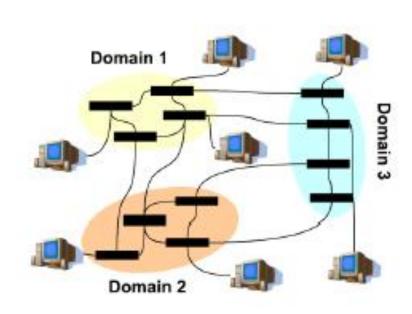
- Ce este o rețea complexă?
- Concepte de bază
- □ Tipuri de rețele complexe
- Aplicaţii

- O mulțime de elemente cu conexiuni între ele
- O rețea (un graf) G=(N,M) este format dintr-o mulțime de noduri  $N=\{n_1,n_2,...,n_N\}$  și o mulțime de conexiuni  $L=\{l_1,l_2,...,l_M\}$
- Un graf = o sbstractizare matematică a rețelei

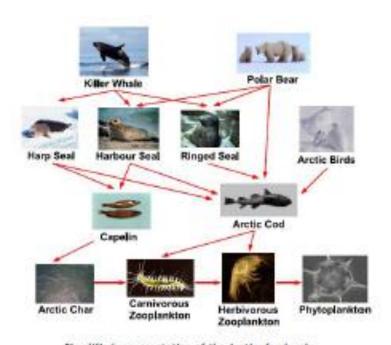


- Fiecare element este reprezentat prin
  - Locație (fizică)
  - Nod (informatică)
  - Actor (sociologie)
  - Vârf (teoria grafelor)
- Interacțiunea dintre 2 elemente este reprezentată prin
  - Legătură (fizică)
  - Link (informatică)
  - Relație (sociologie)
  - Muchie (teoria grafelor)

Nodurile şi legăturile pot proveni din contexte variate

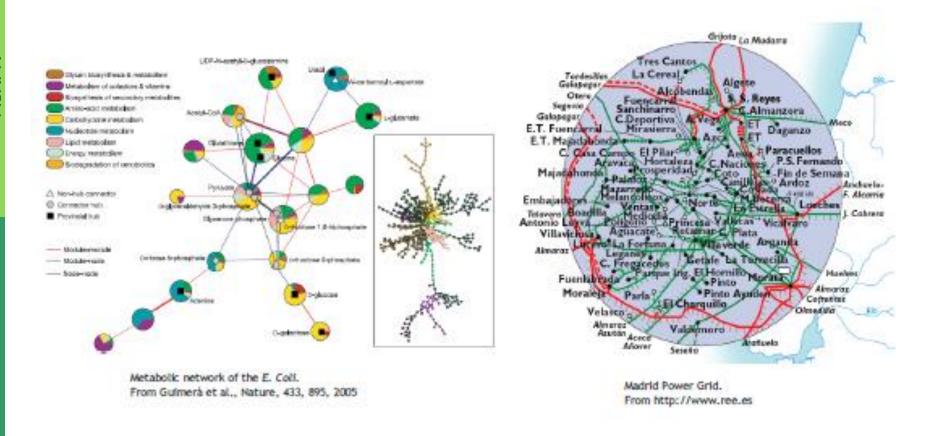


Schematic representation of a network of hosts and routers.

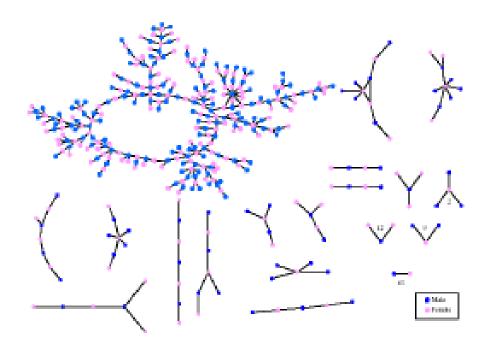


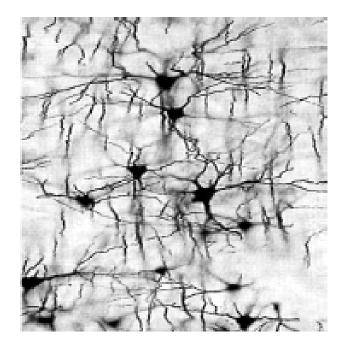
Simplified representation of the Arctic food web

#### Nodurile şi legăturile pot proveni din contexte variate



#### Nodurile şi legăturile pot proveni din contexte variate

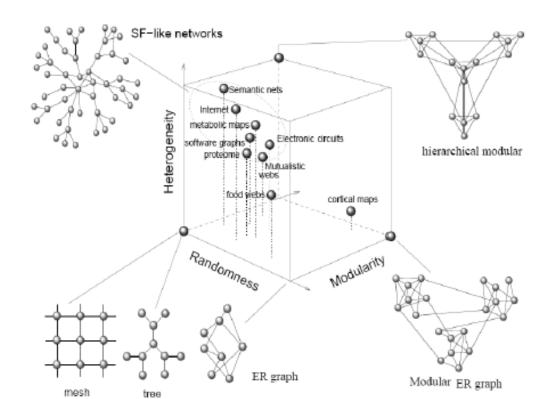




Structure of romantic and sexual contact at Jefferson High School From P.S. Bearman et al., AJS, 110, 44 (2004)

Neuron network

- O rețea complexă este o rețea cu
  - atribute topologice ne-triviale,
  - patternuri de conexiuni care nu sunt nici pur aleatorii, dar nici pur regulare



- Descrierea rețelei (cu // noduri și // legături) prin matrici
  - Matricea ponderilor fiecărei legături

     Matricea ponderilor fiecărei legături

     Matricea ponderilor fiecărei legături

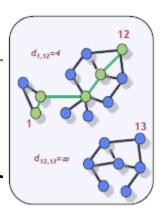
     Matricea ponderilor fiecărei legături

     Matricea ponderilor fiecărei legături

Matricea de adiacență A

- Matricea Laplaciană L = A K
   K matrice diagonală cu
  - $k_{ii} = \sum a_{ii}$

- $\square$  Cel mai scurt drum între 2 noduri ( $d_{ij}$ )
  - Cel mai scurt drum (ca nr de muchii)
  - Cel mai scurt drum (ca sumă a ponderilor mu

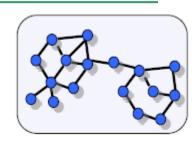


#### Drum mediu /

- Media (aritmetică sau armonică) tuturor celor mai scurte drumuri între oricare 2 noduri ale rețelei
- □ Diametru (*D*)
  - Cel mai lung dintre toate drumurile cele mai scurte
- Componentă
  - Mulţimea tuturor nodurilor în care se poate ajunge plecând dintr-un anumit nod

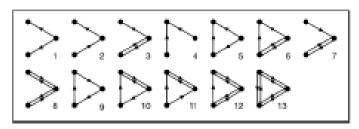
- $\square$  Gradul unui nod  $(k_i)$ 
  - Numărul de legături ale unui nod
- $\square$  Puterea unui nod  $(s_i)$ 
  - Suma ponderilor conexiunilor unui nod  $s_i = \sum w_{ij}$
- □ Apropierea (*b<sub>i</sub>*) unui nod sau a unei legături
  - Numărul celor mai scurte drumuri care trec prin acel nod/acea legătură

- Nodul cu cel mai mare grad e şi cel mai puternic?
- □ Care e nodul cu cea mai mare apropiere?



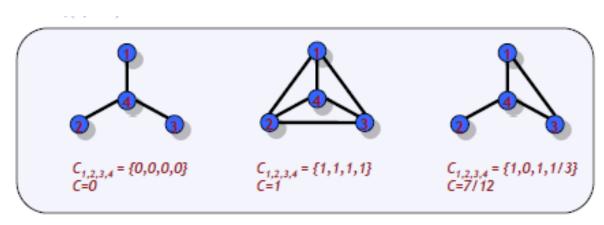
- Motive în rețea
  - Componente (sub-grafe)
     care apar în rețea mai
     des decât ne-am aștepta
     (în caz aleator)
  - Fiecare motiv poate "coda" informaţii specifice

Example: all 13 types of three-node connected subgraphs:

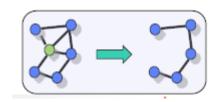


Network	Nodes	Edges	$N_{ m real}$	$N_{\mathrm{rand}} \pm \mathrm{SD}$	Z score
Gene regulati (transcription			>	X V Y V Z	Feed- forward loop
E. coli S. cerevisiae*	424 685	519 1,052	40 70	$7 \pm 3$ $11 \pm 4$	10 14
Neurons			>	X W Y W Z	Feed- forward loop
C. elegans†	252	509	125	90 ± 10	3.7
Electronic circuits (digital fractional multipliers)			/ Y←	_ z	Three- node feedback loop
s208 s420 s838‡	122 252 512	189 399 819	10 20 40	1 ± 1 1 ± 1 1 ± 1	9 18 38
World Wide V	Veb			X Y Y Z	Feedback with two mutual dyads
nd.edu§	325,729	1.46e6	1.1e5	$2e3 \pm 1e2$	800

- □ Coeficient de clusterizare C → are my friends, friends of my friends?
  - Ia în calcul numărul de triunghiuri din rețea
  - Coeficentul unui nod = raportul dintre numărul legăturilor care conectează vecinii nodului și numărul de legături posibile între acești vecini
  - C- media coeficienților tuturor nodurilor din rețea



- Eficiență globală
  - Media armonică a drumurilor optime între toate nodurile rețelei
- □ Eficiență locală (a unui nod) → coeficientul de clusterizare
  - Lungimea celui mai scurt drum între mulțimea vecinilor nodului (fără a considera nodul)



#### Spectrul unui graf

- Mulţimea valorilor proprii a matricii de adiacenţă sau a celei Laplaciene
  - Un graf cu N noduri are N valori proprii #=(#, #2..., #w) şi N vectori proprii V<sub>f</sub>=(V<sub>1</sub>, V<sub>2</sub>,..., V<sub>w</sub>).
  - importante pentru atributele topologice
    - Diametru
    - Număr de cicluri
    - Propagarea informației
  - importante pentru atributele de conectivitate
- Densitatea spectrală

$$\rho(\mu) = \frac{1}{N} \sum_{i=1}^{N} \delta(\mu - \mu_i)$$

#### Structura comunitățiilor

- Comunitate în graf
  - Un sub-graf ale cărui noduri sunt puternic conectate (sau cel puțin mai puternic conectate decât în cazul unei rețele aleatoare)

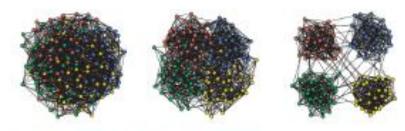
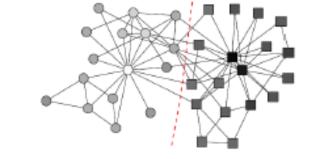


Figure from: Guimerà et al., Nature, 433, 895(2005)

- Evaluare
  - Modularitate M



Zachary Karate Club

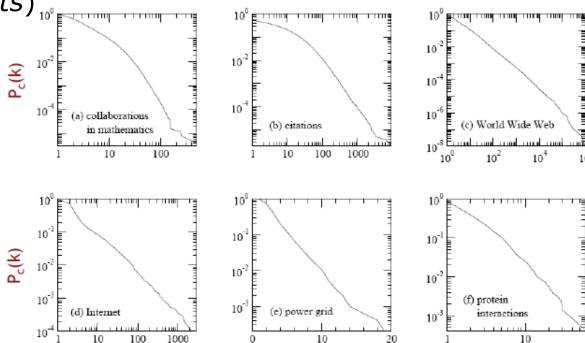
$$M = \sum_{s=1}^{N_M} \left[ \frac{l_s}{L} - \left( \frac{d_s}{2L} \right)^2 \right]$$

where  $N_M$  is the number of modules, L is the number of links in the network,  $l_s$  is the number of links between nodes in module s, and  $d_s$  is the sum of the degrees of the nodes in module s.

- Distribuţia gradelor
  - Distribuția (cumulată a) gradelor
    - Proporția nodurilor în rețea cu grad (mai mare sau) egal cu un prag k

Poate fi exponențială (random nets) sau lege putere

(scale-free nets)



degree k

#### Clustering distribution

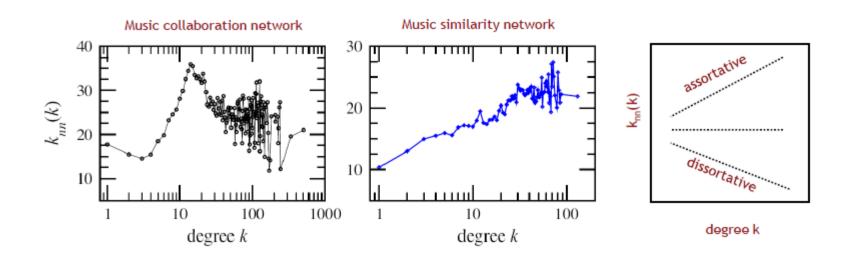
Figure: Clustering distribution three organisms: Aquidex aeolicus (archaea) Escherichia coli (bacterium) and Saccharomyces cerevisiae (eukaryote) (E). (F) The C(k) curves averaged over all 43 organisms is shown, and the inset displays all 43 species together. Lines correspond to C(k)~k-1, and diamonds represent the C(k) value expected for an equivalent scale-free network, indicating the absence of scaling

Aquitex apolicus

Escharichia coli

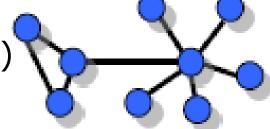
From E. Ravasz et al., Science, 297, 1551 (2002).

- Nearest neighbor degree k<sub>nn</sub>(k) şi assortativity
  - k<sub>nn</sub>(k) măsoara gradul vecinilor, fiind un indicator al asortării rețelei



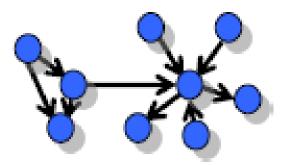
- Direcția legăturilor
  - Neorientate
  - Orientate
- Tipul legăturilor
  - Neponderate
  - Ponderate
- Diferențele între noduri
  - Simple
  - Bipartite
- Topologie
  - Statice
  - Evolutive
- Dinamica nodurilor
  - Fără dinamică
  - Cu dinamică

- □ Direcţia legăturilor → procesele dinamice din reţea (propagarea informaţiei, sincronizarea, robusteţea)
  - Neorientate (legături simetrice)



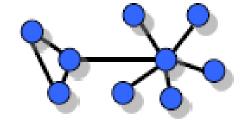
Examples: router network, power grid.

Orientate (legături asimetrice)

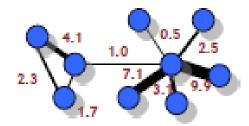


Examples: Internet, food webs, e-mail/telephone networks, etc...

- □ Tipul legăturilor → procesele dinamice din rețea (propagarea informației, sincronizare, robustețe)
  - Neponderate (omogene)



Examples: citation network, internet, etc...

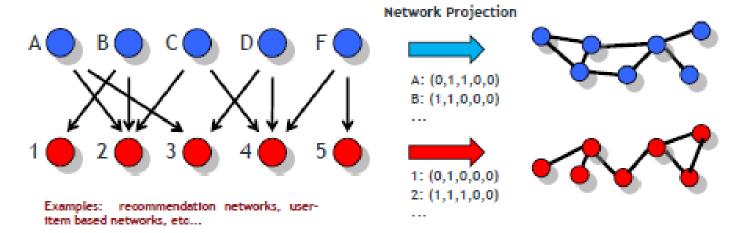


Ponderate (eterogene)

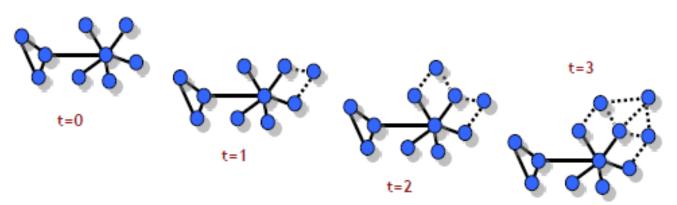
#### Diferențele între noduri

Simple

■ Bipartite → rețele cu noduri de 2 sau mai multe tipuri și legături doar între nodurile de același



- Topologie → rețelele nu "apar" instantaneu;
  - Statice (ca și structură)
  - Evolutive (ca și structură)
    - Posibile întrebări:
      - Care sunt reguli ce guvernează evoluția?
      - Care sunt consecințele aplicării acestor reguli asupra topologiei finale a rețelei?



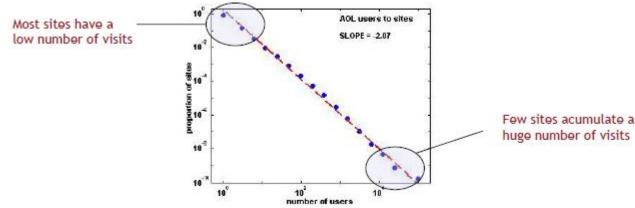
- Dinamica nodurilor
  - poate fi influențată prin matricea conexiunilor;
  - se poate analiza influența topologiei rețelei în procesele dinamice care apar în rețea
    - sincronizări,
    - procese stocastice
  - dar şi vice-versa (influenţa proceselor asupra topologiei reţelei)
  - Fără dinamică
  - Cu dinamică → nodurile sunt sisteme dinamice (cuplate):
    - oscilatoare periodice ex. Pendulul lui Foucault,
    - sisteme excitabile,
    - oscilatoare haotice ex. criptări,
    - □ sisteme bistabile ex. *semiconductor memory*

#### Rețele sociale

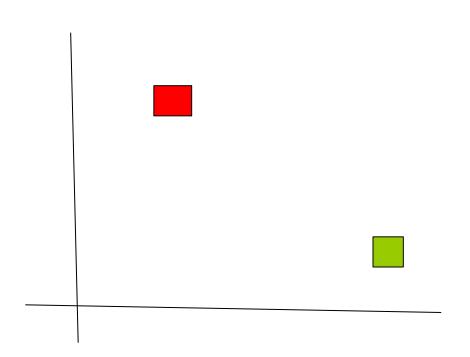
- Proprietăți
  - Small-world
  - □ Distribuţii de tip power-low → noduri foarte conectate (hub-uri)
  - Coeficient de clusterizare mare (relativ la rețelele aleatoare)
  - Amestec de asortare
    - Ex: most connected nodes like to be together

#### ■ Rețele sociale

- Vilfredo Pareto (1848 1923)
  - Distribuția bunăstării (venituri, pământuri, etc.) în diferite țări
    - Au aceeași distribuție *power-law*  $p(X \ge x) \sim x^{-\beta}$
  - Principiu Pareto (regula 80-20):
    - 80% din efecte provin din 20% din cauze
      - Consecință a distribuției cumulate de tip power law
    - Aplicat în diferite domenii (economie ... sociologie)

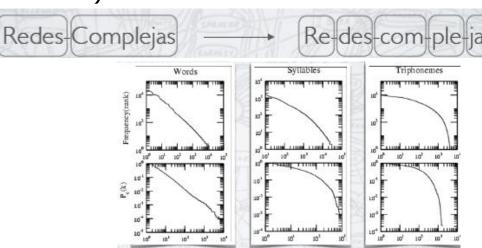


AOL users' visits to various sites on a December day in 1997. From L.A. Adamic



#### ■ Rețele sociale

- George K. Zipf (1902 1950)
  - Frecvenţa apariţiei cuvintelor în limba engleză urmează o distribuţie a rangurilor (r) de tip power-law
    - $N(r) \sim r^{\gamma}$
  - Contexte (la diferite scări)
    - Cuvinte
    - Silabe
    - Foneme



Simon model: "as words are

successively added to the text, a context is created. As the context emerges, it favors the later appearance of certain words -in

particular, those that have already appeared- and inhibits the use of

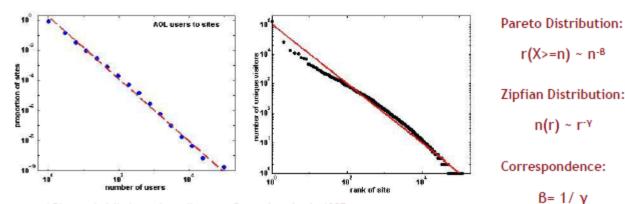
Depend on the context

Depends on semantics

others."

- Distribuţia se schimbă atunci când se pierde semantica
  - Frecvenţa şi gradul devin similare

- Rețele sociale
  - George K. Zipf (1902 1950)
    - Alte domenii
      - Lingvistică
      - Distribuţia populaţiei
      - Rangurile veniturilor
    - Legătura cu Pareto
      - al r-lea cel mai larg oraș are n locuitori (Zipf)
      - r orașe au n sau mai mult de n locuitori (Pareto)

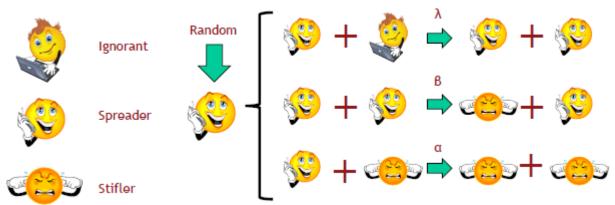


AOL users' visits to various sites on a December day in 1997.

From L.A. Adamic, Zipf, Power-laws, and Pareto - a ranking tutorial.

#### ■ Rețele sociale

- Răspândirea zvonurilor/bolilor
  - Modelul Ignorant-Spreader-Stifler (modelul Daley and Kendal)
    - La fiecare moment de timp un informator aleator i este selectat; acesta contacteaza unul din vecinii săi j
      - Dacă j este ignorant, j se transformă în informator
      - Daca j este un informator sau un ne-interesat, i se transformă în ne-interesat



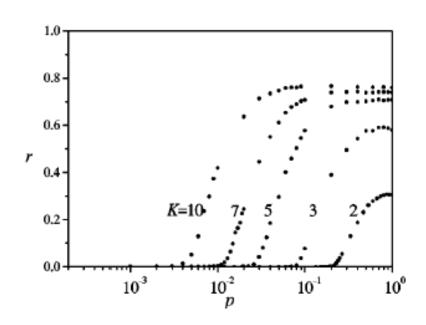
#### ■ Rețele sociale

- Răspândirea zvonurilor/bolilor într-o rețea small-world
  - Modelul Ignorant-Spreader-Stifler (modelul Daley and Kendal)

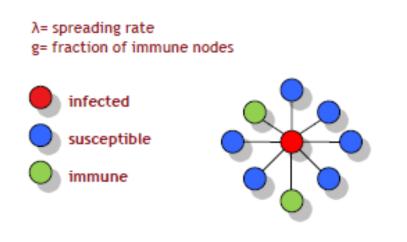
After a certain transient, the rumor stops:

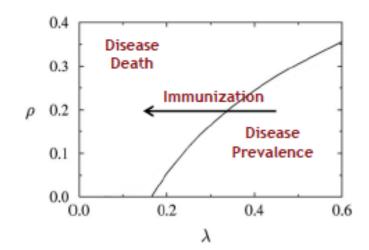
N= total number of nodes
r = N<sub>stiflers</sub>/N
P= probability of rewiring
λ = β = α (for the case of the figure)

- Below a certain p<sub>c</sub> the rumor does not reach a significant part of the network.
- Highest degree, highest probability for the onset of the rumor propagation.



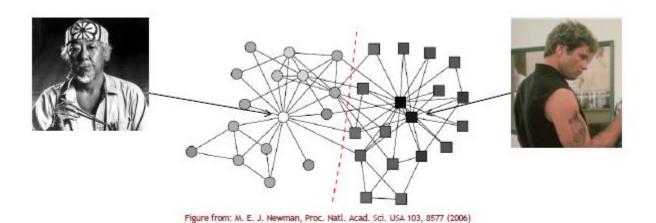
- Rețele sociale
  - Răspândirea zvonurilor/bolilor într-o rețea small-world
    - Cum să luptăm împotriva răspândirii?





Effective Spreading Rate:  $\lambda_g \rightarrow \lambda$  (1-g)  $\leq \lambda_c$ 

- Rețele sociale
  - Comunitățile și rolul lor
    - Clubul Zachary (de karate)

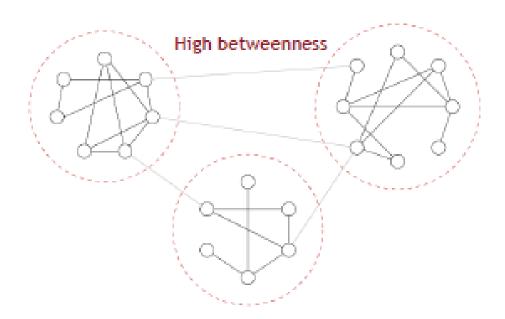


- □ Rețele sociale → Comunitățile și rolul lor
  - Structura modulară a rețelelor complexe
    - Existența comunităților nu se reflectă in distribuția gradelor, clustering sau assortativity
    - Comunitățile sunt relaționate de funcțiile nodurilor
    - Detecția comunităților = problemă cu multiple soluții
      - Algoritmi
        - Metode de eliminare a legăturilor
        - Metode aglomerative
        - Algoritmi bazaţi pe modularitate
        - Metode spectrale

- □ Rețele sociale → Comunitățile și rolul lor → Detecția comunităților
  - Metode de eliminare a legăturilor
    - Se tot elimină legăturile slabe până când rețeaua începe să se rupă

#### Example:

The weakest links are those with the highest betweenness.

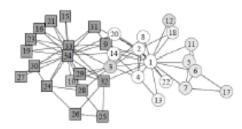


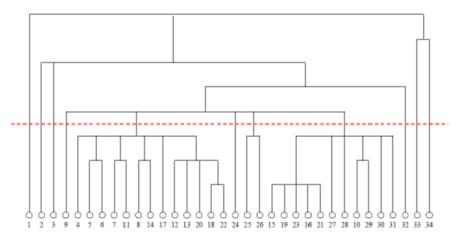
- □ Rețele sociale → Comunitățile și rolul lor → Detecția comunităților
  - Metode aglomerative (Bottom-up)
    - Inițial fiecare nod aparține unei comunități
    - Nodurile similare

#### Example:

The similarity is based on the structural equivalence

$$x_{ij} = \sqrt{\sum_{k \neq i,j} (A_{ik} - A_{ij})^2}$$





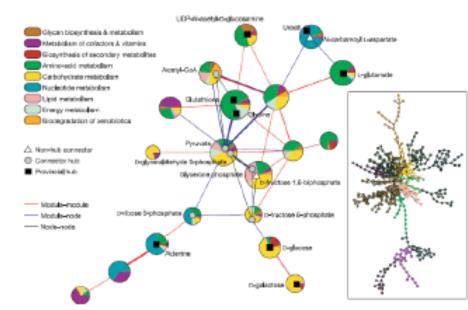
Hierachical clustering partition of the Zachary Karate Club.

- □ Rețele sociale → Comunitățile și rolul lor → Detecția comunităților
  - Algoritmi bazați pe modularitate
    - Modularitatea = numărul de elgături dintre grupuri numărul estimat de legături între grupuri într-o rețea aleatoare

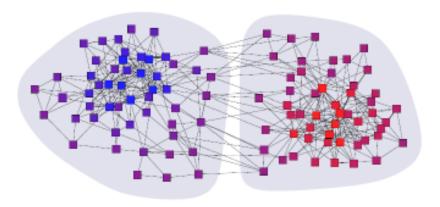
#### Example:

The simulated annealing algorithm is an stochastic algorithm that searches for the maximum modularity by allowing negative (decreasing) variations of the modularity.

Figure: Partition of the metabollic network of E. Coli using AS algorithm. From Guimerà et al., Nature, 433,895 (2005).



- □ Rețele sociale → Comunitățile și rolul lor → Detecția comunităților
  - Metode spectrale
    - Separarea rețelei în
      - 2 (sau mai multe) componente
      - prin analiza spectrală a matricii Laplaciene
        - care conţine toate informaţiile topologice despre reţea



#### ■ Rețele biologice

- Proprietăți
  - Small-world
  - Dissortative mixing
    - multe dintre nodurile conectate nu sunt conectate preferențial unele cu altele
  - Organizate în sub-module
    - Modularitate ridicată
    - Structuri comunitare
- Pionieri
  - Watts şi Strogatz



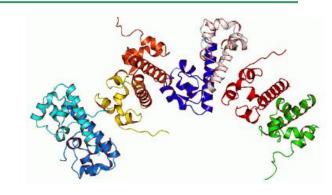
	Lactual	Lrandom	$C_{ m actual}$	$C_{\rm random}$
Film actors Power grid	3.65 18.7	2.99 12.4	0.79 0.080	0.00027 0.005
C. elegans	2.65	2.25	0.28	0.05

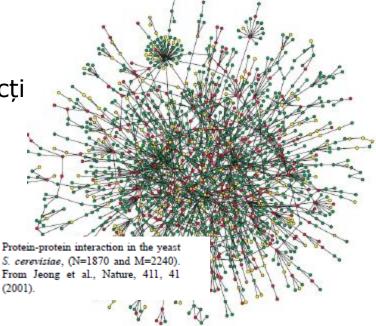
The small-world of *C. Elegans* neural network, with an edge joining two neurons if they are connected by either a synapse or a gap junction (n= 282, <k>= 14.). Table from Watts et al., 393, 440 (1998)

- Rețele biologice
  - Rețele metabolice
  - Rețele proteice
    - Drojdia
      - 6000 proteine
      - 3 interacțiuni per proteină
      - 20 000 interacțiuni
    - Corpul uman
      - De ordinul 100 000 de interacți
  - Rețele genetice



Figure from Thanos, et al., Science, 283, 833 (1999)





#### Rețele biologice

- Rețele neutre ale acidului ribonucleic (RNA)
  - Transformarea secvență structură este degenerată
    - Aceeași structură poate fi obținută cu o multitudine de diferite lanțuri

```
A: adenina

C: citosina

G: guanina

U: uracilo | Gecceccegugacgaaucgaaucuaagcu

Gecaauugcuaaaaggaucuaaucuagcu

Gecceccgugacgaacgaaucuaagcucaagcu

Gecceccgugacgaaucuaaagguccaagcu

Gecceccgugacgaaucuaaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaaagcucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucuaacucaagcu

Gecceccgugacgaaucucaagcu

Gecceccgugacgaaucucaagcu

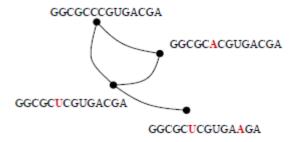
Gecceccgugacgaaucucaagcu

Gecceccgugacgaaucucaagcu

Geccecccgugacgaaucucaagcu

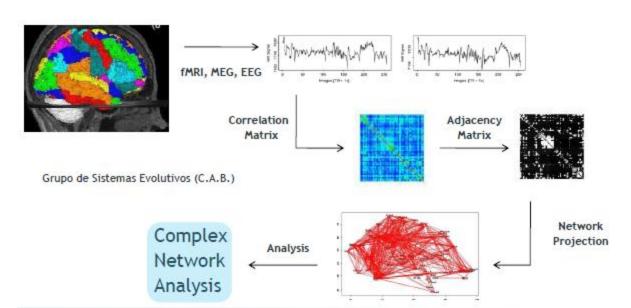
Gecceccgugacgaaucucaagcu

Gecceccucaagcu
```



#### Reţele biologice

- Rețele funcționale ale creierului
  - RMN funcțional
    - Măsoară cum diferite părți ale creierului răspund la stimuli exteriori sau la activități pasive într-o stare de repaus
  - Electroencefalogramă
    - înregistrarea biocurenților câmpurilor transcranieni la nivelul tegumentelor și înregistrarea grafică a acestora în timp sub forma de unde
  - Magnetoencefalogramă
    - tehnică de imagistică neurală funcțională pentru cartografierea activității creierului prin înregistrarea câmpurilor magnetice produse de curenții electrici ce apar în mod natural în creier,









#### ■ Rețele biologice

- Rețele funcționale ale creierului
  - Alzeheimer
  - Mild Cognitive Impairment
  - Schizofrenie
  - Epilepsie

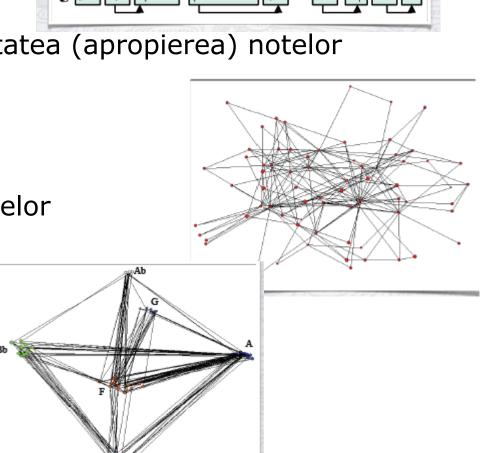


- Rețele muzicale
  - Rețele ale notelor
  - Rețele ale melodiilor
  - Rețele ale artiștilor
  - Rețele ale utilizatorilor

- Rețele muzicale
  - Rețele ale notelor
    - Nota = nod
    - Legătura = proximitatea (apropierea) notelor

Rețeaua duratei notelor

Rețeaua notelor



- □ Rețele muzicale de note
  - Legea Zipf

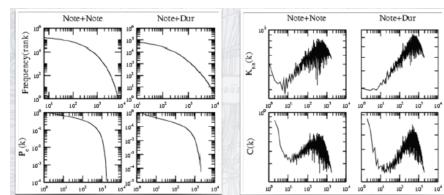
 Context = o ierarhie a modelelor (şabloanelor) care apar în diferite tempo-uri (progresie armonică,

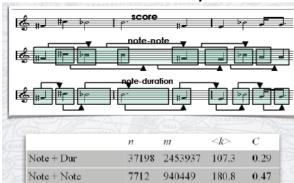
melodie, ton, ritm, ...)

Note + durată

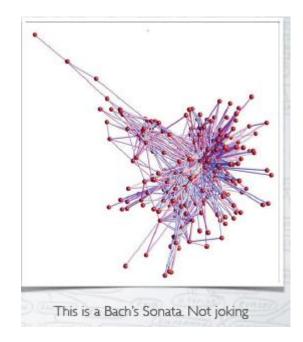
Note + note

- Rețelele de note
  - sunt assortative
  - au o distribuţie de clusterizare specifică

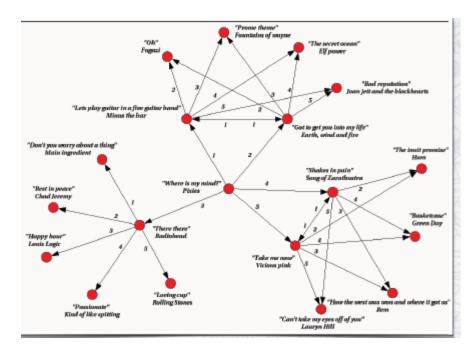




- □ Rețele muzicale de note
  - De ce?
    - Sa creăm muzică pornind de la proprietățile rețelei și folosind procese aleatoare ghidate



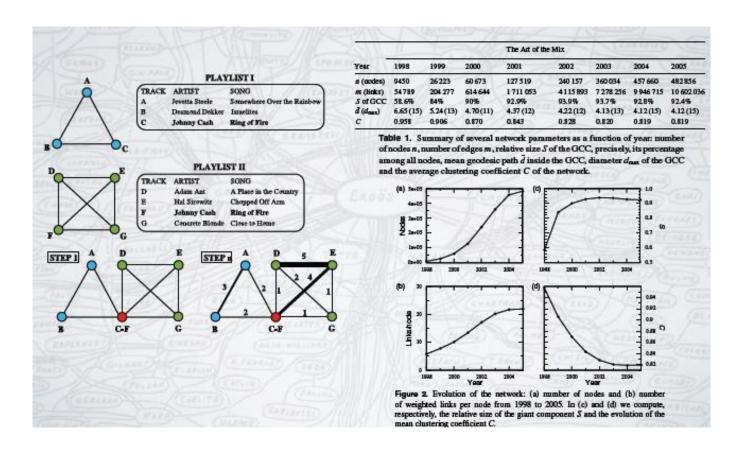
- Rețele muzicale
  - Rețele ale melodiilor
    - □ Nod = melodie
    - Legătura = diferite relații
      - Ex. Apariția în playlist-uri



From "The complex network of musical tastes", J.M. Buldú, P. Cano, M. Koppenberger, J.A. Almendral and S. Boccaletti, New J. Phys. 9, 172 (2007).

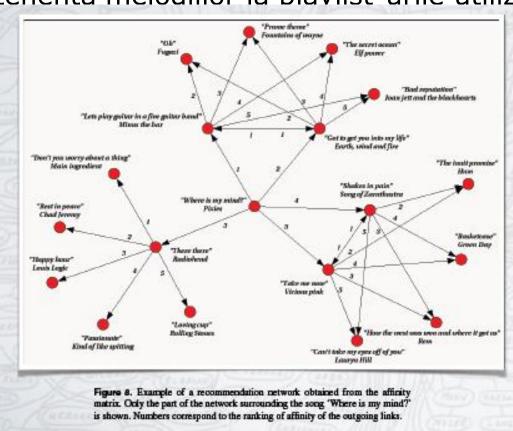
- Rețele muzicale ale melodiilor
  - De ce?
    - Analiza structurii
      - Drumuri, module
    - Detectarea celei mai influente melodii
    - Clasificare (etichetare)
    - Proiectarea unor sisteme automate de recomandare eficiente

- Rețele muzicale de melodii
  - Evoluează în timp

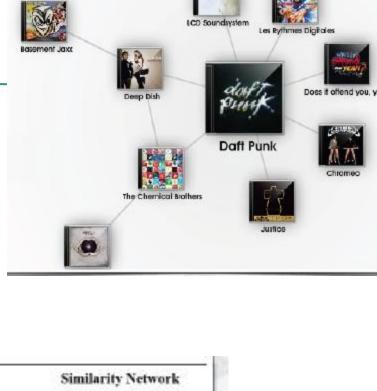


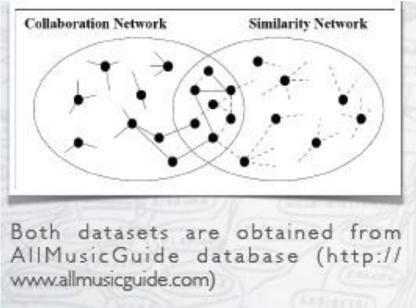
- Rețele muzicale de melodii
  - Gusturi muzicale

Apartenenta melodiilor la plavlist-urile utilizatorilor

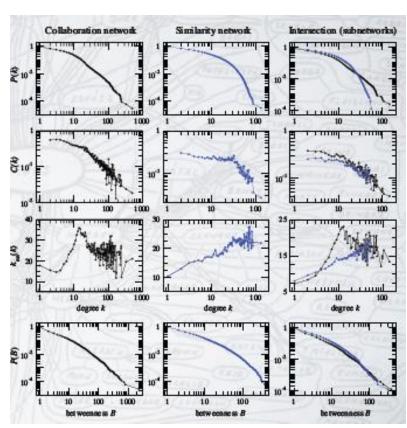


- Rețele muzicale
  - Rețele ale artiștilor
    - □ Nod = artist
    - Legătură = diferite relații
      - Similaritate
      - Colaborare
      - Afinitate

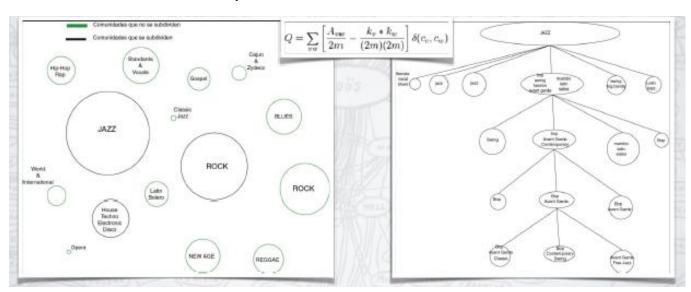




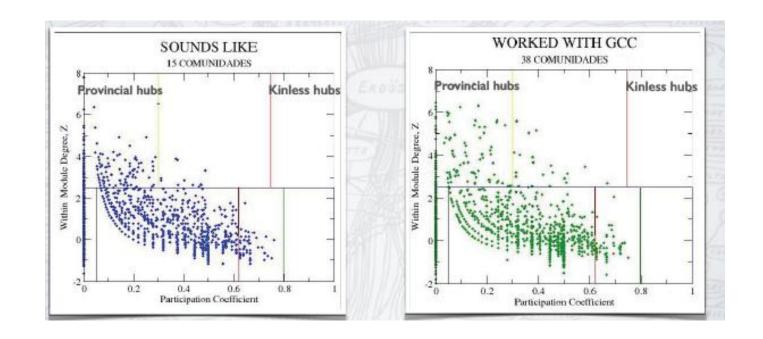
- Rețele muzicale
  - Rețele ale artiștilor
    - □ Nod = artist
    - Legătură = diferite relații
      - Similaritate
      - Colaborare
      - Afinitate
    - Au acelaşi număr de noduri
    - Au structuri diferite



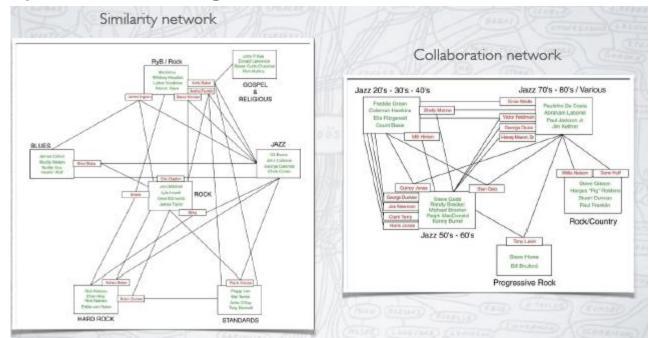
- Rețele muzicale ale artiștilor
  - Detecția comunităților oferă informații despre rețea
    - Separarea rețelei pe baza modularității
      - În comunități
      - În sub-comunități



- Rețele muzicale ale artiștilor
  - Detecția comunităților oferă informații despre rețea
    - Separarea rețelei pe baza modularității
    - Se pot identifica hub-uri

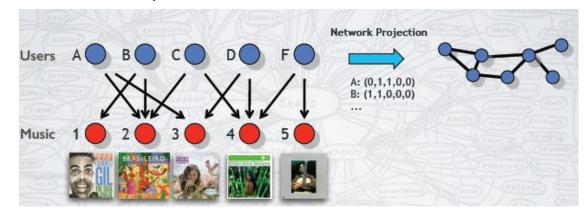


- Rețele muzicale ale artiștilor
  - Detecția comunităților oferă informații despre rețea
    - Separarea rețelei pe baza modularității
    - Se pot identifica hub-uri
    - Se pot face cartografieri ale hub-urilor



#### Rețele muzicale

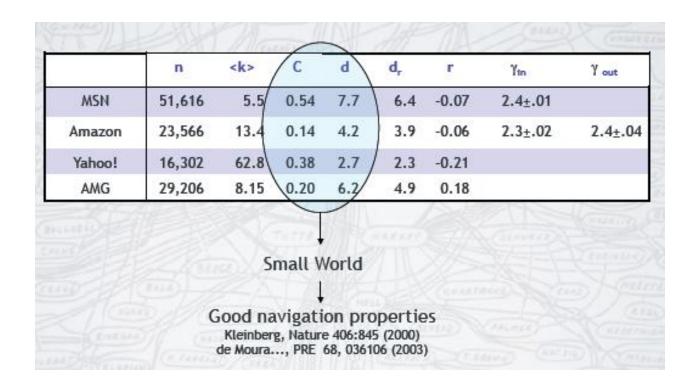
- Rețele ale utilizatorilor
  - Nod = utilizator (consumator)
  - Legătură = diferite relații



- De ce?
  - □ Folosite în sistemele de recomandare



- Rețele muzicale ale utilizatorilor
  - De ce e importantă topologia rețelei?



- Rețele muzicale
  - O diversitate de a proiecta muzica într-o rețea
  - Care model de rețea este cel mai potrivit?

