

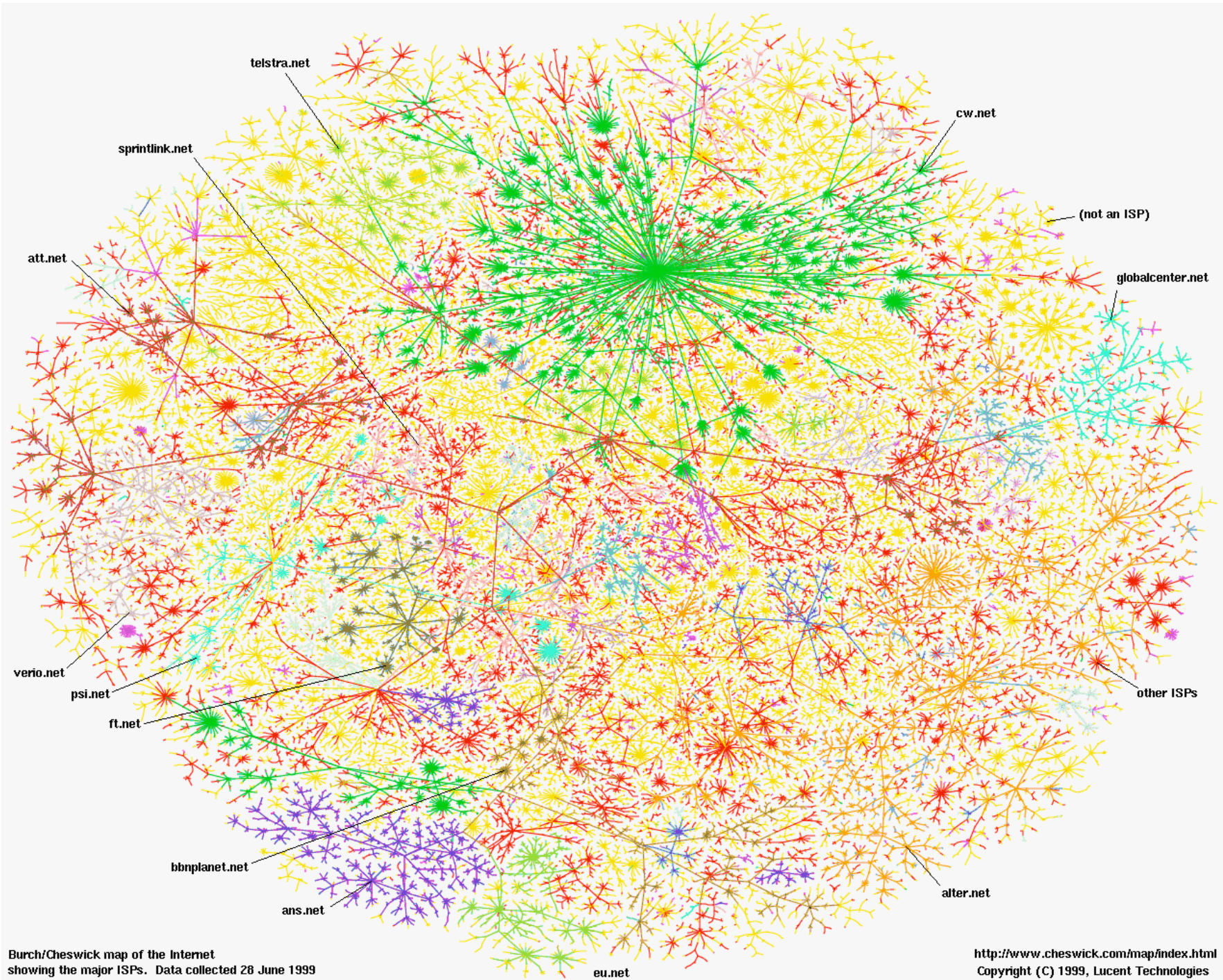
Internet ja muut informaatioverkostot

Pekka Orponen

Teknillinen korkeakoulu

Tietojenkäsittelyteorian
laboratorio





Burch/Cheswick map of the Internet showing the major ISPs. Data collected 28 June 1999



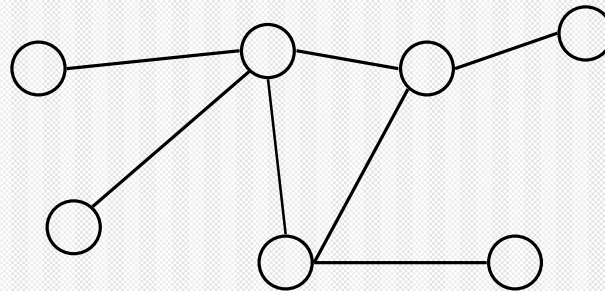
Sisällys

- Verkostoja
- Verkostomalleja
- Verkostoprosesseja

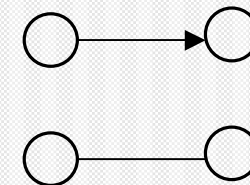


Peruskäsitteitä

- *Verkko* t. *verkosto*: kokoelma *solmuja* ja niitä yhdistäviä *kaaria*.



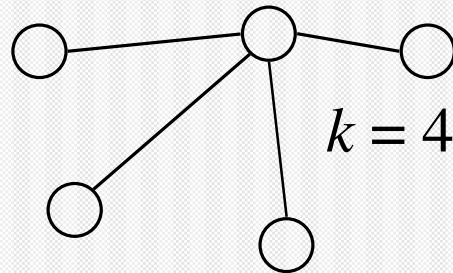
Kaaret voivat olla *suunnattuja* tai *suuntaamattomia*



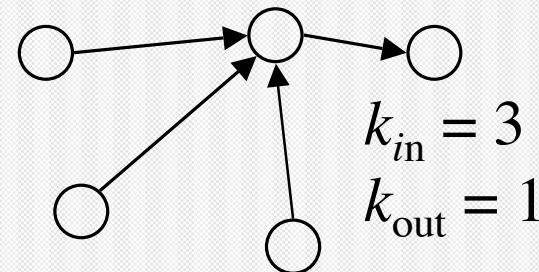


Peruskäsitteitä (jatk.)

- Solmun *aste* k on sen välittömien naapureiden (= solmuun liittyvien kaarien) määrä:



- Suunnatuissa verkoissa solmuilla on erikseen *tulo-* ja *lähtöaste*:





Verkostoja: Internet ja WWW

- Internet:
 - n. 15.000 “autonomista osaverkkoa”
 - n. 250.000(?) reittipalvelinta
 - laitteiden fyys. yhteyksien mukaisen suuntaamattoman verkon keskim. asteluku osaverkkotasolla $\langle k \rangle \sim 3.5$, reititintasolla $\langle k \rangle \sim 2.6$
- WWW
 - n. 10 miljardia (??) WWW-sivua
 - Sivulla keskim. 7 linkkiä muihin sivuihin -> suunnatun linkkiverkon keskim. lähtöasteluku $\langle k \rangle \sim 7$



Verkostoja: vertais-, spontaani- ja sensoriverkot

- Lähitulevaisuudessa tulevat yleistymään myös yksittäisten tietokonelaitteiden suoraan muodostamat infrastruktuurittomat verkot.
- Vertaisverkko (“peer-to-peer”, p2p-verkko).
 - Verkko, jossa laitteet jakavat resursseja (tiedostoja, koneaikaa, palveluja) keskenään ilman keskitettyjä resurssipalvelimia
- Spontaaniverkko (“ad hoc” -verkko)
 - Langattomien laitteiden keskenään, ilman kiinteän verkon tukea, muodostama tietoverkko
- Sensoriverkko:
 - Ympäristön havainnointiin suunniteltujen pienten ja halpojen erikoislaitteiden muodostama spontaaniverkko



Verkkomalleja: miksi?

- Halutaan ymmärtää ja ennustaa em. tapaisten verkostojen rakennetta ja toimintaa yleisin perustein, esim.:
 - Miten verkon solmujen väliset etäisyydet kasvavat verkon kasvaessa? Voiko etäisyydet pitää lyhyinä isossakin verkossa lisäämällä kaaria strategisiin paikkoihin? Mihin?
 - Miten haavoittuva verkko on solmujen tai kaarien vikaantumiselle? Montako solmua yhtenäisestä verkosta voidaan poistaa ennen kuin se hajoaa pienempiin osiin?
 - Mikä olisi paras tapa välittää sanomia verkon solmusta toiseen, jos verkon rakenne tunnetaan vain paikallisesti?
 - Voiko verkostosta (esim. WWW) tunnistaa pelkän solmujen naapuruustiedon perusteella sisällöllisesti tärkeitä solmuja tai yhteenkuuluvien solmujen ryhmiä?



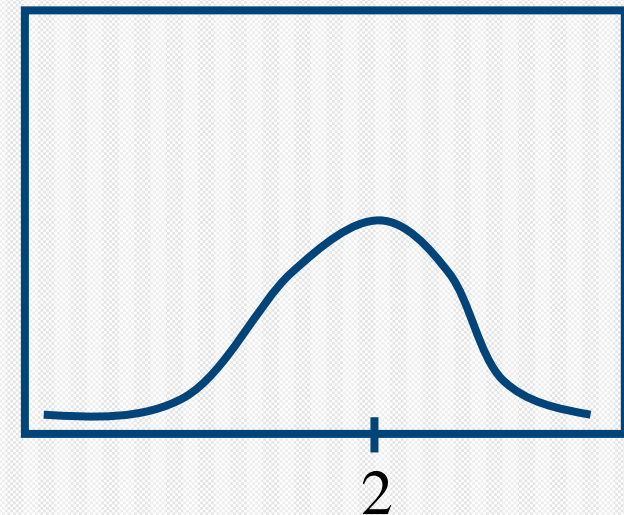
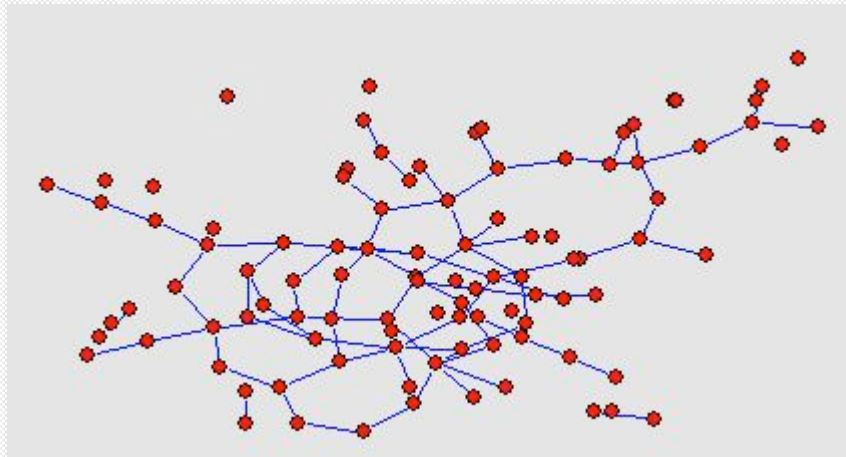
Erdős-Rényi -satunnaisverkot

- Klassinen, paljon tutkittu suuntaamattomien satunnaisverkkojen malli
- Erdős & Renyi (1959), Solomonoff & Rapoport (1951)
- Verkossa n solmua, kahden satunnaisesti valitun solmun välillä kaari tasaisella todennäköisyydellä p
- Seuraus: solmujen keskimääräinen asteluku $\langle k \rangle = p(n-1) \sim pn$. Suurilla n ja pienillä p astelukujen jakauma vahvasti keskittynyt tämän arvon ympärille (ns. Poisson-jakauma).



ER-verkot ja Internet

- Esim. ER-prosessilla tuotettu verkko, jossa $n = 100$, $\langle k \rangle = 2$

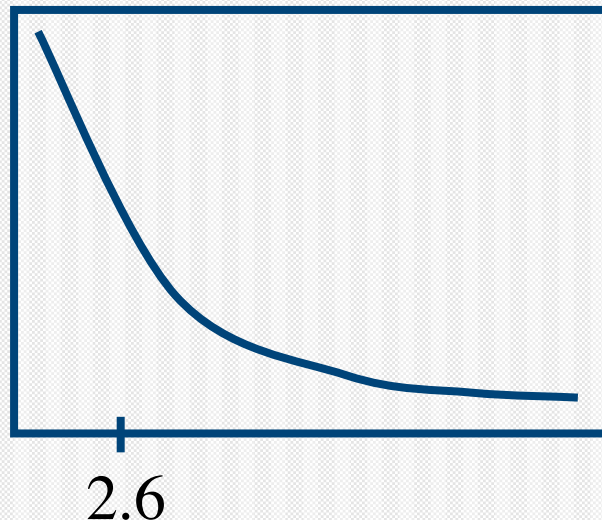


- Onko Internet ($\langle k \rangle \sim 2.6$) tällainen?



Internetin astelukujakauma

- Internet *ei ole* ER-verkko (Faloutsos³ 1999)
- Astelukujakauma on pikemminkin tällainen:



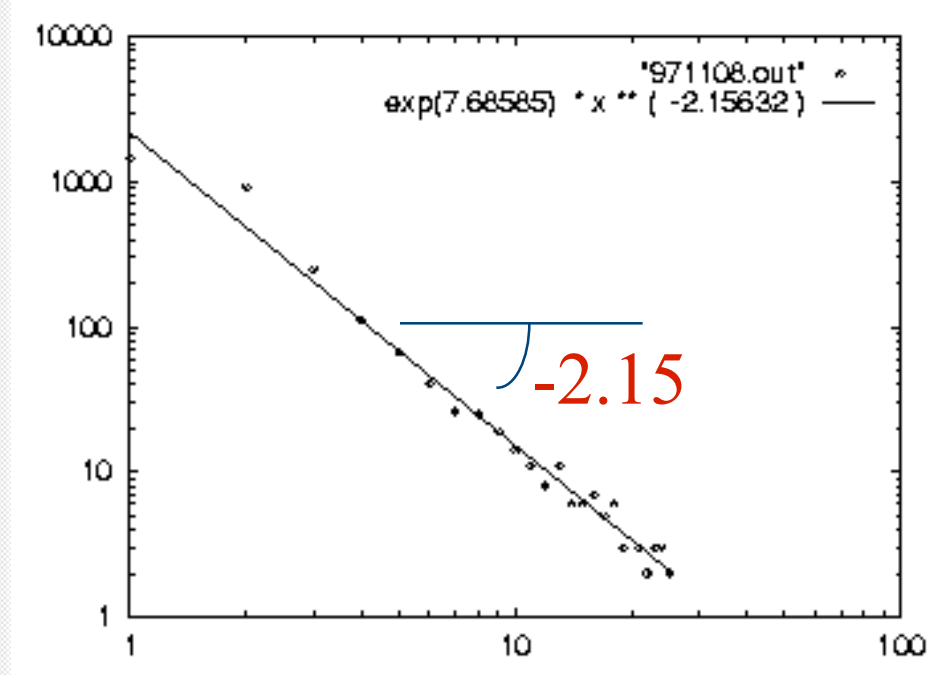
- Useimpien solmujen (reitittimien) asteluku on 1 tai 2, mutta joukossa on myös hyvin korkea-asteisia



AS-Internet (Faloutsos³ 1999)

- Internetin autonomisten osaverkkojen astelukujakauma noudattaa “potenssilakia”:
astetta k olevien solmujen määrä $\sim 1/k^{2.15}$

$\log(l_{km})$

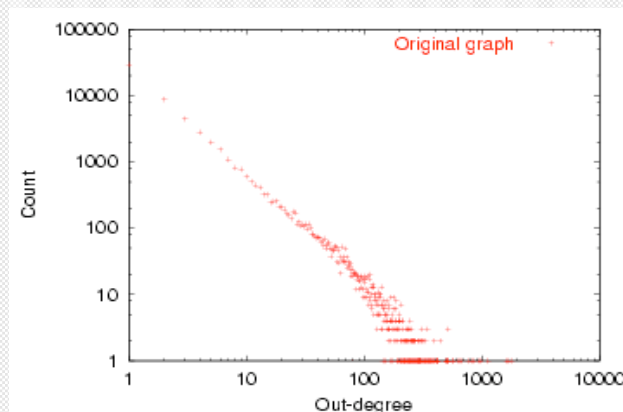


$\log(\text{aste})$

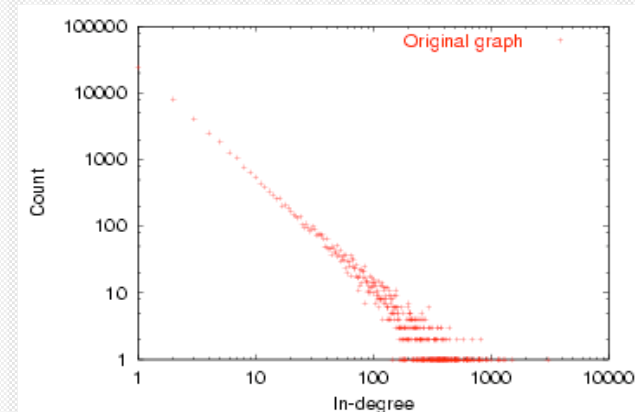


Muita “skaalattomia” verkostoja

- Internet-reititinverkko: astetta k olevien reitittimien lkm $\sim 1/k^{2.48}$ (Faloutsos³ 1999)
- WWW (Albert et al. 1999, Kumar et al. 1999):
 - Verkkosivut, joissa on k linkkiä: lkm $\sim 1/k^{2.4}$
 - Verkkosivujen, joihin osoittaa k linkkiä: lkm $\sim 1/k^{2.1}$



lkm vs. lähtöaste



lkm vs. tuloaste



Kasvavat satunnaisverkot

- Skaalattomia astelukujakaumia voidaan tuottaa *kasvattamalla* satunnaisverkko painotetusti.
- “Preferential attachment” (Barabási & Albert 1999); “cumulative advantage” (De Solla Price 1965); “Matthew effect” (Simon 1955).
- Verkostoa kasvatetaan solmu kerrallaan niin, että kunkin uuden solmun kaaret (linkit) kytketään vanhoihin solmuihin todennäköisyyksin, jotka ovat suhteessa niiden senhetkisiin astelukuihin. Näin “suositut” solmut lisäävät suosiotaan ja “epäsuositut” jäävät syrjään.
- De Solla Price: suunnatut verkot (esim. WWW); Barabási & Albert: suuntaamattomat (esim. Internet)



Kasvavat satunnaisverkot (jatk.)

- Kasvatusprosessi tuottaa verkkoja, joiden astelukujakaumat muistuttavat tunnettuja; esim. Pricen mallissa viittausten (tuloasteen) jakauma, kun keskim. lähtöaste on $\langle k \rangle$:

$$p_k \sim 1/k^\alpha, \text{ missä } \alpha = 2 + 1/\langle k \rangle$$

- Esim. WWW:n havaittu lähtöasteluku $\langle k \rangle \sim 7$ antaa Pricen mallissa $\alpha \sim 2.14$, mikä vastaa havaintoja ($\alpha \sim 2.1$) erittäin hyvin
- Kasvumalleista on myös parametroituja versioita, joita voidaan sovittaa vapaammin dataan



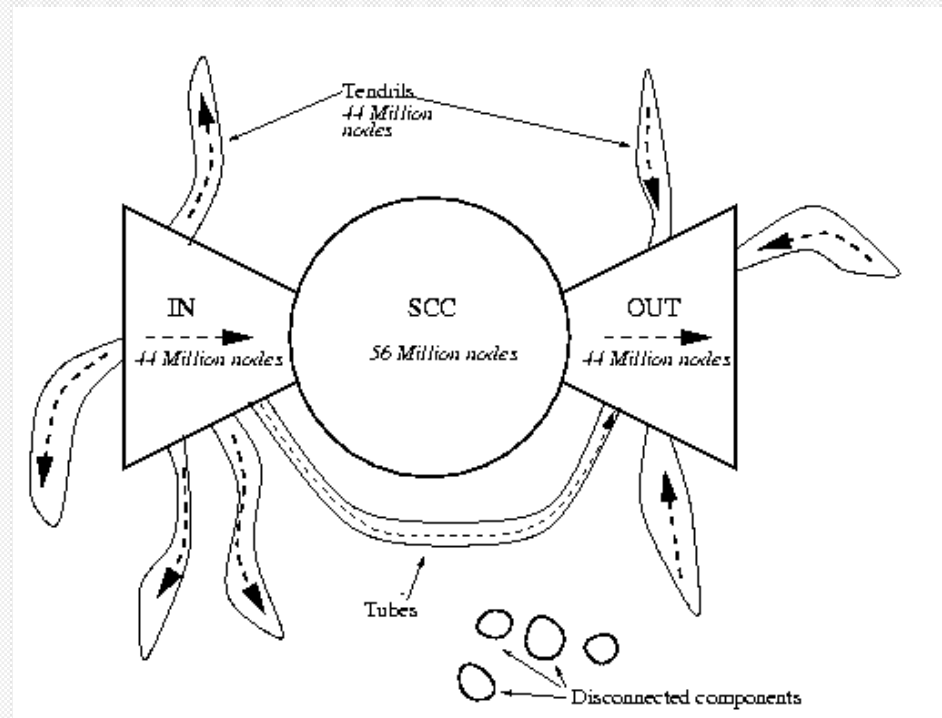
WWW-verkon muita ominaisuuksia

- Pieni halkaisija: satunnaisesta solmusta pääsee toiseen keskimäärin 19 linkin kautta
 - Internet: AS-halkaisija ~ 4 ; reititin-halkaisija ~ 9
 - Kasvumallit selittävät nämä ominaisuudet hyvin
- Klusteroituminen: naapurisolmut ovat usein naapureita myös keskenään
 - Voidaan mallintaa kasvumalleilla, jotka “kopioivat” naapurustoja -- ei vielä aivan luontevaa mallia



WWW-verkon ominaisuuksia (jatk.)

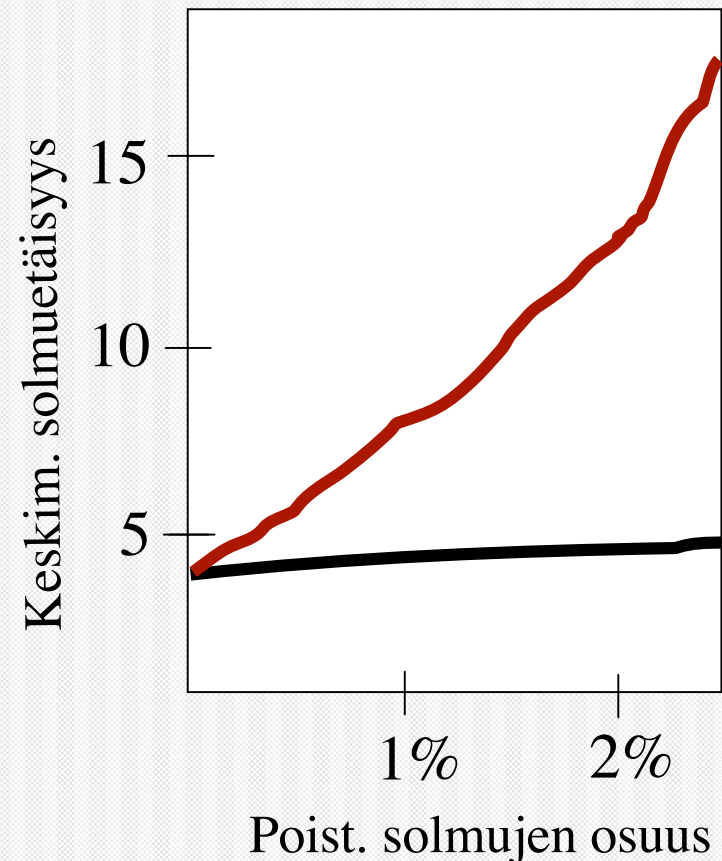
- “Rusettirakenne” (Broder et al. 1999):





Verkostojen haavoittuvuus

- Skaalattomissa verkoissa ovat harvalukuiset korkea-asteiset solmut (“hubs”) keskeisessä asemassa kommunikoinnin kannalta.
- Albert et al. (2000): keskim. solmuetäisyyden piteneminen AS-Internetissä kun solmuja poistetaan (a) satunnaisesti, (b) **asteluvun mukaisessa järjestyksessä**:





Virusepidemiat verkostoissa

- Epidemiologiset mallit:
 - SIR: Susceptible -> Infective -> Recovered
 - SIS: Susceptible -> Infective -> Susceptible
- Perinteisissä epidemiologisissa tarkasteluissa saadaan yl. epidemian ekstensiiviselle leviämiselle (SIR) tai pysyydelle (SIS) kynnyisarvo, joka riippuu taudin tarttuvuus- ja parantuvuusnopeuksista.
- Skaalattomissa verkostoissa tällaista kynnyisarvoa ei kuitenkaan ole (jos $\alpha < 3$), vaan epidemia leviää *aina* ekstensiiviseen osuuteen solmuista (Pastor-Satorras & Vespignani 2001).



Virusepidemiat (jatk.)

- Syynä epidemioiden leviämiseen ovat jälleen verkon korkea-asteiset napasolmut (“hubit”).
- Toisaalta voidaan osoittaa, että epidemiat voidaan myös tehokkaasti estää/pysäyttää immunisoimalla juuri napasolmut (Pastor-Satorras & Vespignani 2001).
- Miten napasolmut löydetään isosta ja/tai tuntemattomasta verkosta?



Napasolmujen määrittäminen

- Void. todeta, että (suuntaamattomassa, yhtenäisessä) verkossa suoritettu pitkä satunnaiskävely viettää kussakin verkon solmussa u ajan, joka on verrannollinen u :n astelukuun.
- Napasolmujen rokottamiseksi suoritetaan siis vain pitkä satunnaiskävely ja rokotetaan joka N :s solmu.
- Yksinkertaistettu versio: valitaan satunnaisesti jokin solmu ja rokotetaan sen kaikki *naapurit*. Toistetaan riittävän monta kertaa.



Tiedonhaku verkostoissa

- Rakenteeltaan tuntemattomasta P2P-verkosta pitää löytää tietty tiedosto. Miten menetellä?
- Useimpien solmujen naapurustossa on jokin napasolmu -> Jos kukin solmu tietää omien tiedostojensa lisäksi kaikkien naapureiden tiedostot, niin napasolmuista tulee luonnostaan verkoston “indeksejä”.
- Tiedoston hakemiseksi käynnistetään verkossa satunnaiskävely. Kun jokin kohdattu solmu kertoo, että tiedosto löytyy sen naapurista, käydään noutamassa sieltä.



“Arvokkaiden” solmujen määrittäminen

- Google: PageRank-algoritmi (Page & Brin 1998)
 - Suuntaamattomille verkoille; yleistys suunnatuille verkoille HITS (Kleinberg 1999)
- Idea: solmu on “arvokas” jos sillä on paljon “arvokkaita” naapureita -> kunkin solmun i arvo $0 \leq u_i \leq 1$ on verrannollinen sen naapureiden painotettuun arvosummaan:

$$u_i = (1-\varepsilon)(u_{i1}/k_{i1} + \dots + u_{ik}/k_{ik}) + \varepsilon/n,$$

missä ε on solmujen yhteinen “tasapainotusparametri”.



“Arvokkaat” solmut (jatk.)

- Voidaan osoittaa, että jos verkko on yhtenäinen, niin yhtälöryhmällä:

$$u_1 = (1-\varepsilon)(u_{11}/k_{11} + \dots + u_{1k}/k_{1k_1}) + \varepsilon/n$$

$$u_2 = (1-\varepsilon)(u_{21}/k_{21} + \dots + u_{2k}/k_{2k_2}) + \varepsilon/n$$

⋮

$$u_n = (1-\varepsilon)(u_{n1}/k_{n1} + \dots + u_{nk}/k_{nk_n}) + \varepsilon/n$$

on yksikäsitteinen positiivinen ratkaisu. (Matemaattisin termein verkon yhteysmatriisin Perron-Frobenius - ominaisvektori).

- Tämä voidaan laskea tunnetuin numeerisin menetelmin ja antaa siis verkon solmujen “Google-arvot”.



Tutkimushaasteita

- WWW-verkon ja muiden kiinnostavien verkostojen ominaisuuksien luonteva mallinnus (esim. skaalattomuus + klusteroituminen; aikakehitys ja sen vaikutukset)
- Mallien ominaisuuksien analyttinen ratkaiseminen (esim. kuinka nopeasti epidemiat leviävät skaalattomissa verkostoissa)
- Verkostojen rakenteen hyödyntäminen algoritmeissa (reititys, klusterointi, otanta)