The structure of a large social network

Balázs Szendrői

University of Washington and University of Utrecht

Gábor Csányi and Balázs Szendrői, *Structure of a large social network* cond-mat/0305580, to appear in Phys. Rev. E.

András Lukács and Dániel Varga, *research in progress*

Outline

1. The `wiw.hu` social network
2. The social network underlying the WIW system
3. Modelling the WIW network
4. Conclusions and outstanding issues

1. The **`wiw.hu`** social network

- Project started by a small group of young professionals in Budapest, Hungary in April 2002 with the aim to record social acquaintance
- Operation details
  - The network is invitation–only; new members join by an initial link connecting to the person who invited them
  - New acquaintance links are recorded between members after mutual agreement
  - Members use their real names; no proliferation of multiple pseudonyms
  - List of names fully searchable
  - Network options: "Shortest link", "Second neighbour" searches, "Map" functions
  - Additional services: email, chat rooms, topical forums

Szabó Márton | legutóbbi belépésed: 2004 febr. 01 20:28:57

ismerőseim: 191 | függő meghívóim: 0 | igazolatlan kapcsolatok: 11 | **felhasználók:** 2hó: 18940 (3220) | regisztrált: 46509 | kapcsolatok: 395517

navigáció

- **meghívó**
- **keresés (bővítve!)**
- **ismerősök**
- **függőben**
- **kit ismerhetek?**
- **térkép**
- **közös projektek**
- **WiW-Utazás telekocsi**
- **Witying**
- **fórum**
- **levelezés**
- **kilépés**

Egyedi belépések száma a mai napon: **3246**
Összes belépés száma: 5024

A WiW-en regisztráltak átlagéletkora: **25,6 év** (7542 emberből)
Egyéb adatokat a fejléc menüjében található módosítás menüpontban adhatsz meg!

A gépház üzen:

79. | 2004. jan. 23. | 15:44:51

Újra működik:

- térkép
- witying menüpont
- kit ismerhetek

további üzenetek...

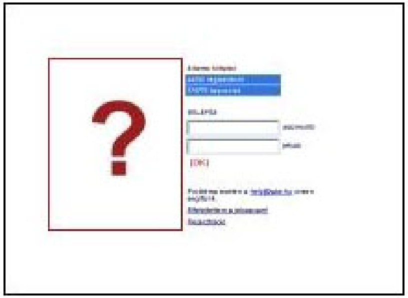
Hirdetések:

Rodigan

CIMLAPKÉP

A SZAVAZÁS EREDMÉNYÉT ITT LÁTOD

Aki még nem tudja mi ez: jan. 31-ig lehet szavazni a WiW bejelentkező oldali képére! A nyertes kép előreláthatólag 1 hónapig lesz kint a bejelentkező oldalon. A képeket felhívás alapján, WiW-esek készítették.



V Í R U S

A szerver levelező szerverének túlterheltsége miatt a levelek napokat késnek!
Január 27-e óta a WiW-es címekre érkezett gyanús levelek vírusosak!
A WiW-et, illetve a te gépedet nem tudja megfertőzni (mert a mellékleteket nem lehet megnyitni), ezért csupán arra jó, hogy a

Ithaka - ITEM - adatlap

Keszulo oldal tervei:
belepo oldal | projekt oldal

Kiskérdések eredménye

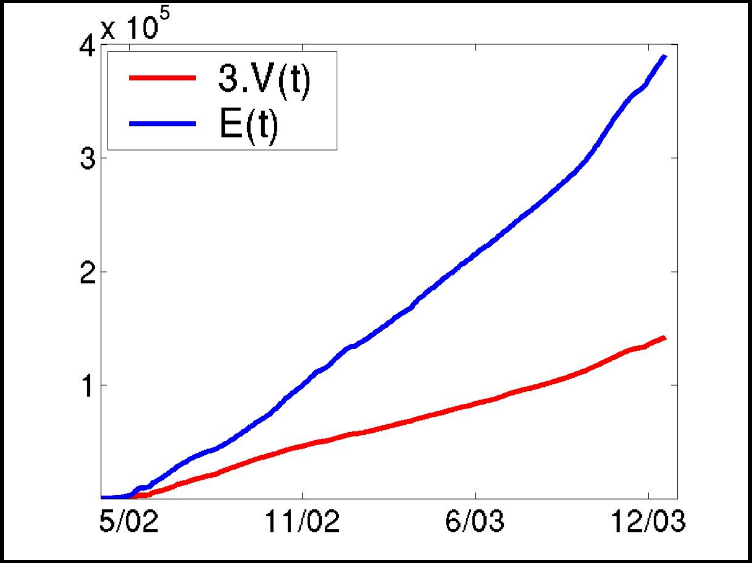
Admin topikok:
Item 2002

A legutolsó 30 feliratkozott: Ha ismersz valakit, klikkelj a névre!

- Dömötör Viola [*vilchi*]
- Kemény Csaba [*lousmerz*]
- Élővilág Béla [*elovilag*]
- Rátkai Balázs [*lviz2*]
- Molnár László [*mlacuska2*]
- Molnár László [*mlacuska1*]
- Korondi Zsolt [*zkorondi*]
- Molnár László [*mlacuska*]
- Kocsis András [*akocsis*]
- Dombi Éva [*dombie*]
- Szűcs András [*bolkask8*]
- Kara Donát [*dongoh*]
- Tóth Viktor Márk [*tovima*]
- Vizkelety Dóra [*vizdori*]
- Karsay Ferenc [*hokiedzo*]
- Bölcsföldi Ádám [*hrcs*]
- Balai Ágota [*balaiagota*]

Growth of the WIW social network

- Growing graph $G = G(t)$: "snowball sampling"
- The number of vertices $V(t) = V(G(t))$ and links $E(t) = E(G(t))$ grew essentially linearly in time

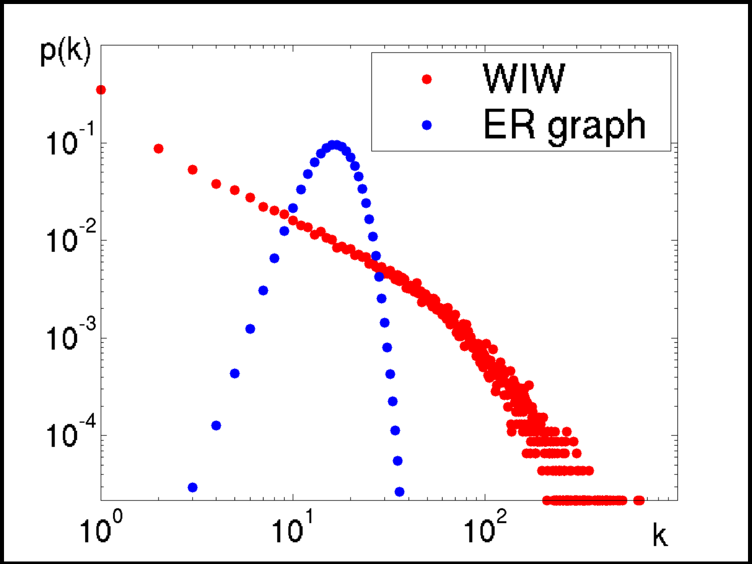


The current WIW graph

- $G = G(T)$: the WIW graph at fixed time $T = 30$ January 2004

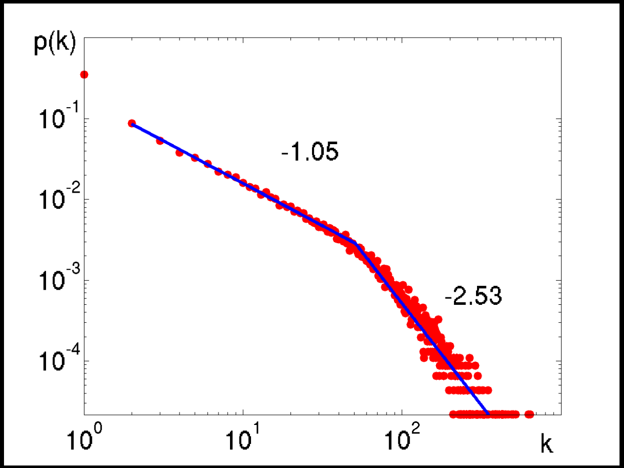
$$V = V(G) = 45886, \quad E = E(G) = 393797, \quad \langle k \rangle_G = 17.16$$

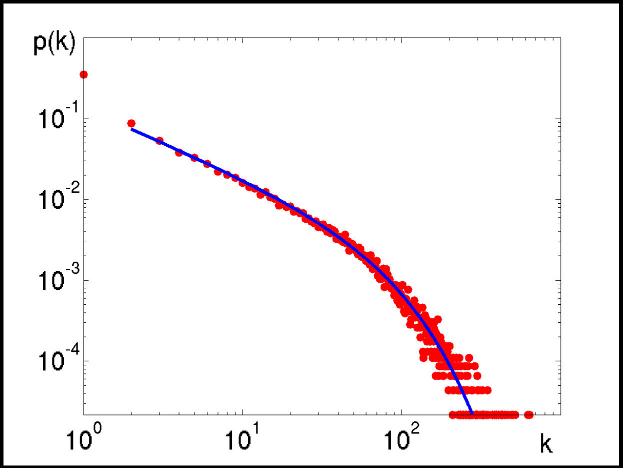
- Degree distribution $p(k) = V^{-1} \cdot \#\{v \in G : \deg(v) = k\}$



Analysis of the degree distribution

- Two competing distribution patterns





$$p(k) = \begin{cases} c_1 k^{-\gamma_1} \text{ for } k < k_{\text{crit}} \\ c_2 k^{-\gamma_2} \text{ for } k > k_{\text{crit}} \end{cases} \quad (1)$$

$$p(k) = c k^{-\gamma} e^{-\alpha \cdot k} \quad (2)$$

- A detailed statistical analysis slightly favours the "exponentially truncated power law" (2) over the "double power law" (1).

Features of the degree distribution

- Power law in degree distribution: present in *growing systems*
  - World Wide Web (Albert–Jeong–Barabási 1999)
  - Preferential attachment model (Barabási–Albert 1999, Bollobás–Riordan–Spencer–Tusnády 2001)
- Exponential decay in the tail: evidence of a natural bound on $\deg(v)$ (Newman, Lukács–Varga)
  - The number of one's acquintances is likely to be bounded from above
  - This constraint is not present in the case of systems such as the Web
- The mixed distribution of WIW resembles degree distributions observed in other systems, such as scientific collaboration network data.

Small-world features following Watts–Strogatz 1998

- *Path length* $l_{v,w}$: the length of the shortest path from v to w.

  WIW average path length

  $$\langle l\rangle_G = \binom{V}{2}^{-1} \sum_{v,w} l_{v,w} = 4.5,$$

  only slightly larger than the average path length in a random graph of the same size $\langle l\rangle_{\text{random}} = 3.8$ (*"small world effect"*)

- *Clustering coefficient* $C(v) = \dfrac{\#\{\text{triangles including } v\}}{\binom{\deg(v)}{2}}$

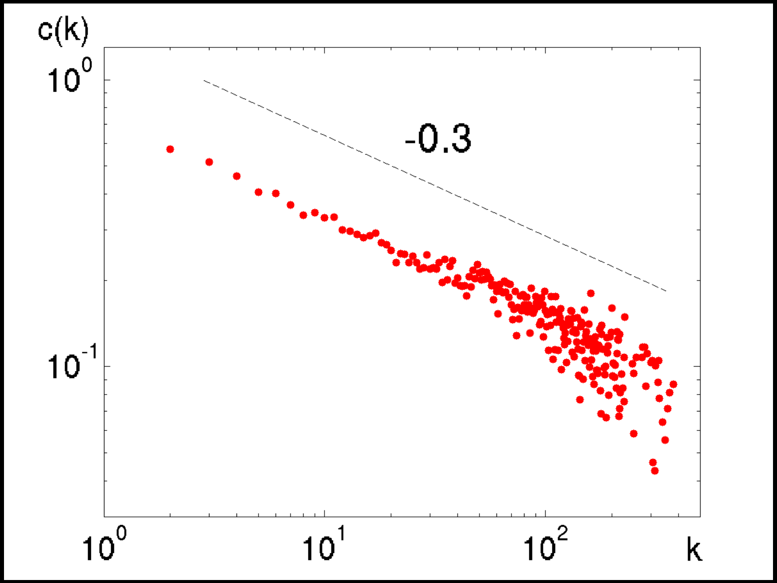
  For the WIW graph, the average clustering

  $$\langle C\rangle_{\mathrm{G}} = V^{-1} \sum_{v} C(v) = 0.19,$$

  which is several orders of magnitude larger than the clustering coefficient of a random graph of the same size and edge density $\langle C\rangle_{\text{random}} = 3.7 \cdot 10^{-4}$.

Degree-clustering correlation

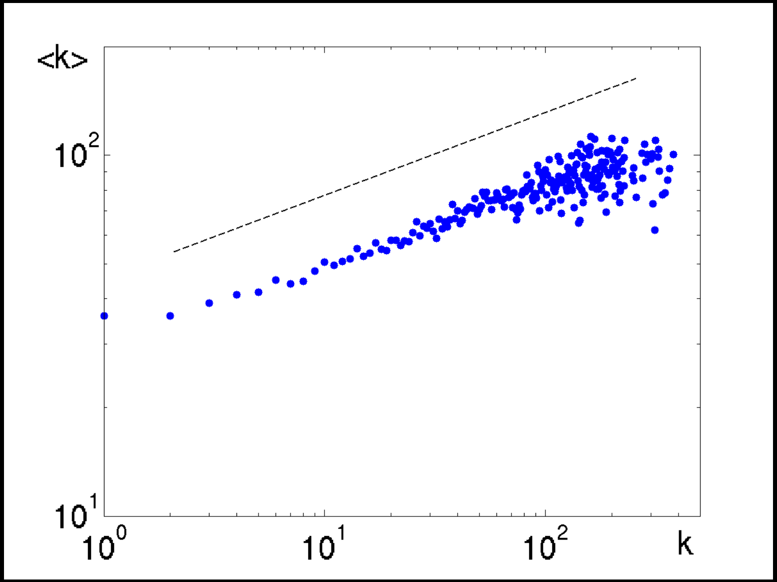
The dependence of the local clustering coefficient $C(v)$ on the degree k



- The power law $C(k) \sim k^{-\alpha}$ has been interpreted by (Ravasz–Barabási 2002) as evidence for the presence of *hierarchical architecture* in the network

Degree-degree correlation and assortativity

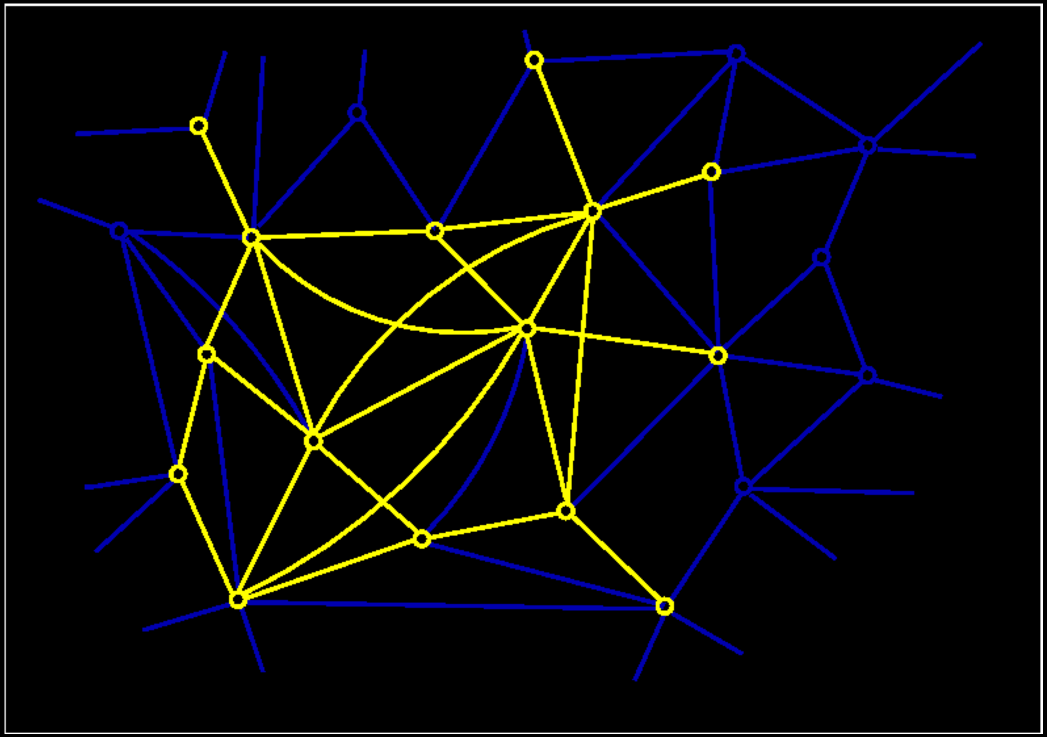
The dependence of the average degree of neighbours $\langle k \rangle$ on the degree k:



- The average neighbour degree increases with the degree: *assortativity*
- Newman's assortativity coefficient: $r = 0.196$.

2. The social network underlying the WIW system

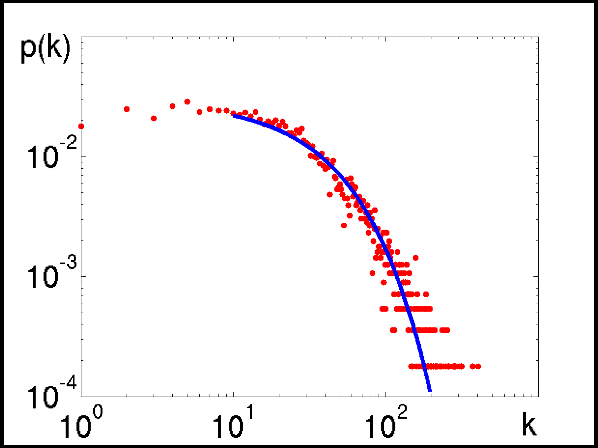
The WIW system grows on an underlying social network



WIW growth and the underlying network

Several features of the WIW graph are likely to be consequences of the growth on an underlying network.

- Linear growth of $V(t)$
- Exponential cutoff in the degree sequence
  - Degree distribution of the graph of "active" users has stronger exponential cutoff (Lukács–Varga)



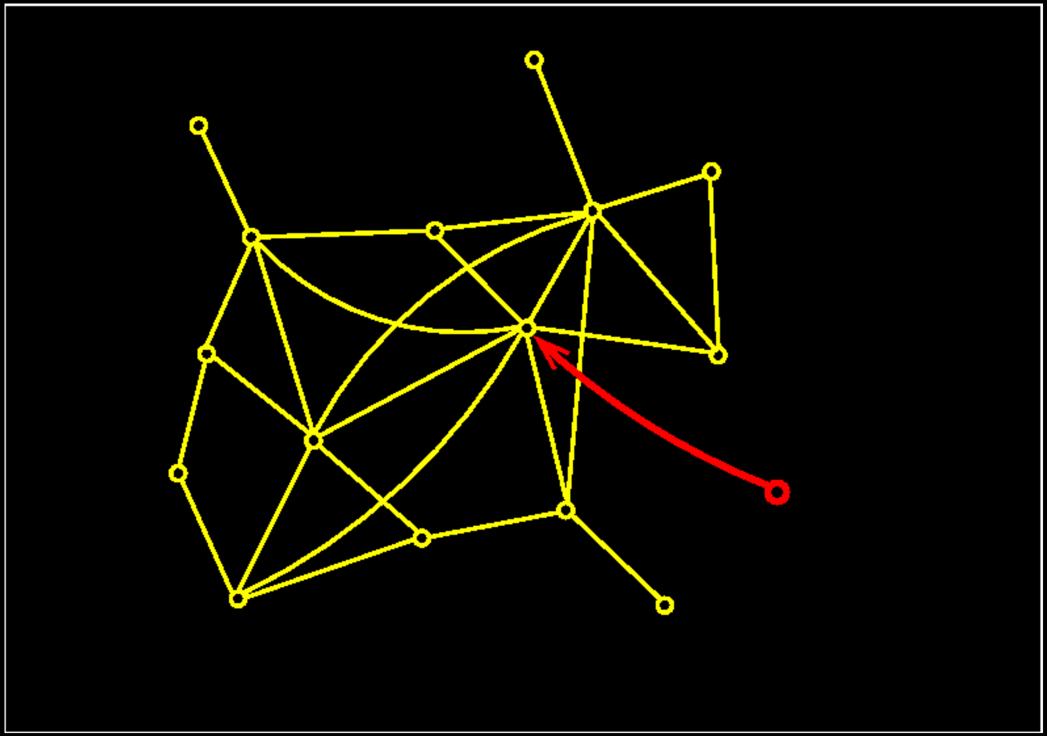
3. Modelling the WIW network

- Model should reproduce observed features, such as
  - highly skewed degree distribution
  - high average clustering $\langle C \rangle_G$
  - short average path length $\langle l \rangle_G$
- Ideally, it should be a two-step process
  1. Model underlying social network
  2. Model growth process
- We do not have a good model for the underlying social network!

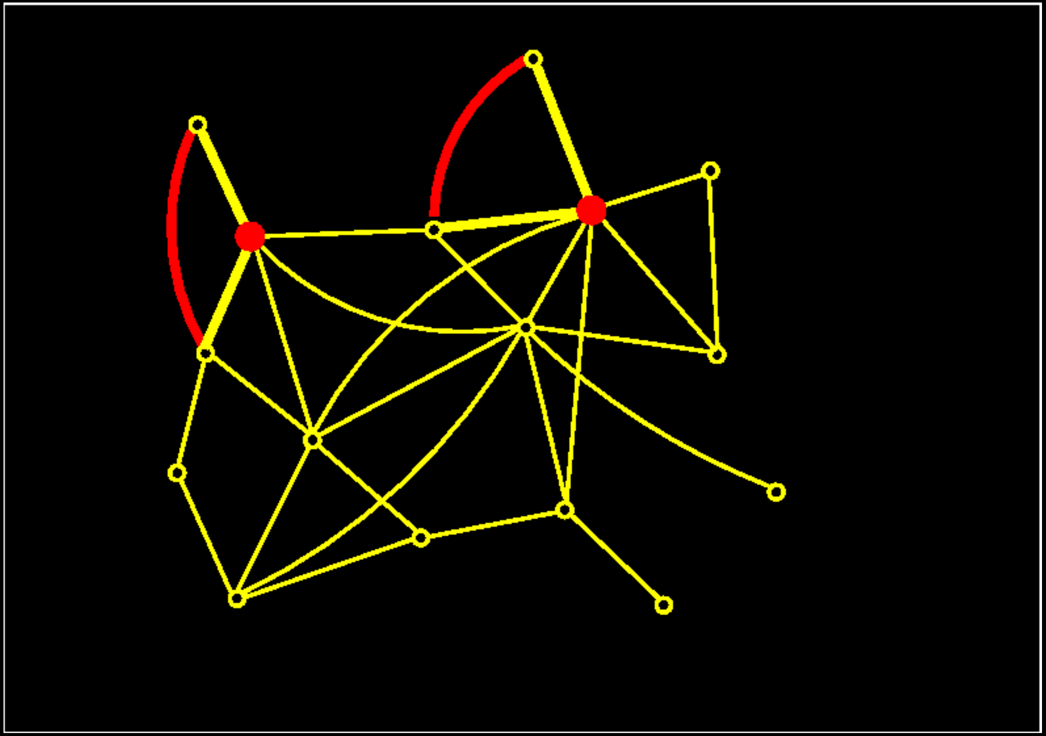
Our model of the WIW network

- *Growing network*: new nodes arrive at a rate of one per unit time.
- *New nodes* attach to an earlier node chosen with a probability distribution giving weight k^q to a node of degree k
  - Exponent $0 < q < 1$ is a parameter; idea motivated by experimental results on scientific collaboration networks (Barabási et al 2001, Newman 2001)
- *Internal edges* are created by a strictly local mechanism: a new edge is introduced between two unconnected neighbours of a randomly chosen node
- Constant edge/node ratio imposed

Our model of the WIW network



Our model of the WIW network



Degree distribution of our model

- Degree distribution: depends on the parameter q
  - For large k, we obtain a power law decay

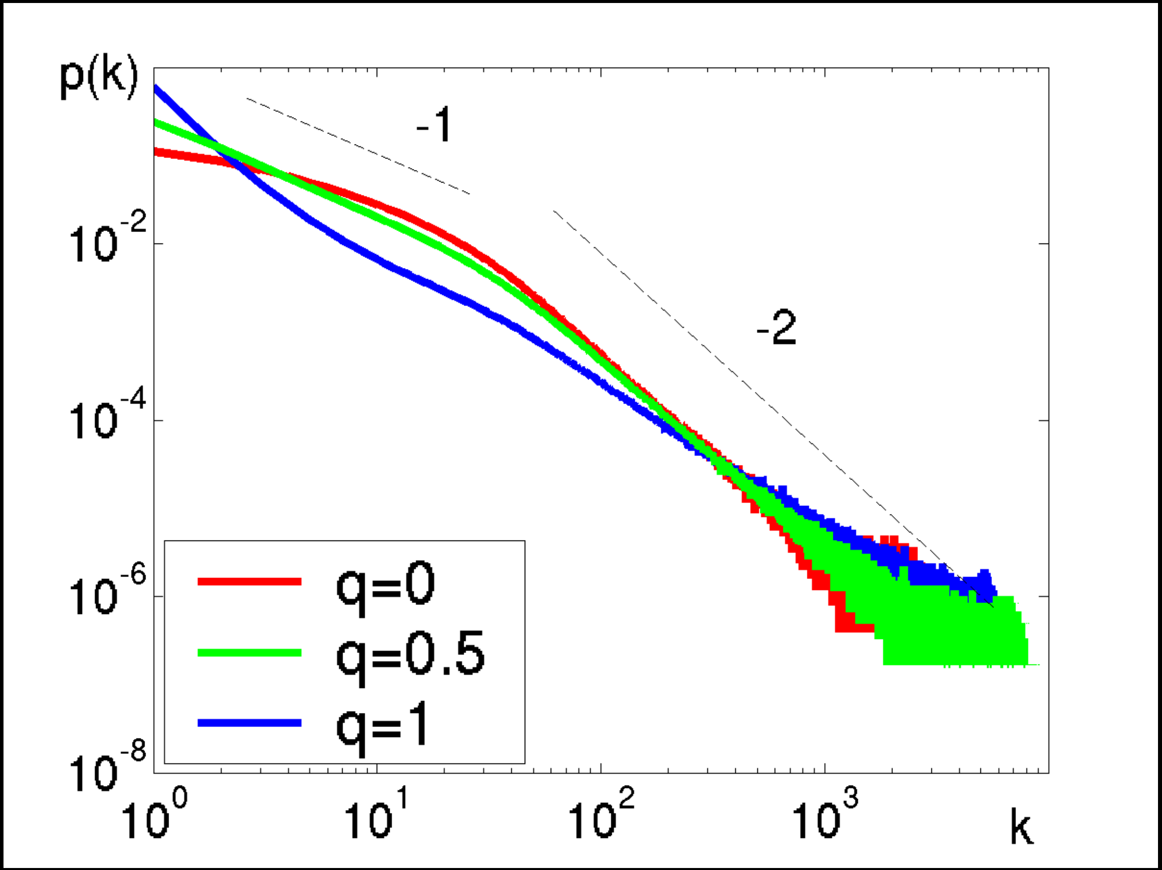
$$p(k) \sim k^{-2},\ k >> 1$$

    * This can be demonstrated by a “mean-field” type argument
  - Reproduces small-k power law

$$p(k) \sim k^{-1},\ k < k_{\text{crit}}$$

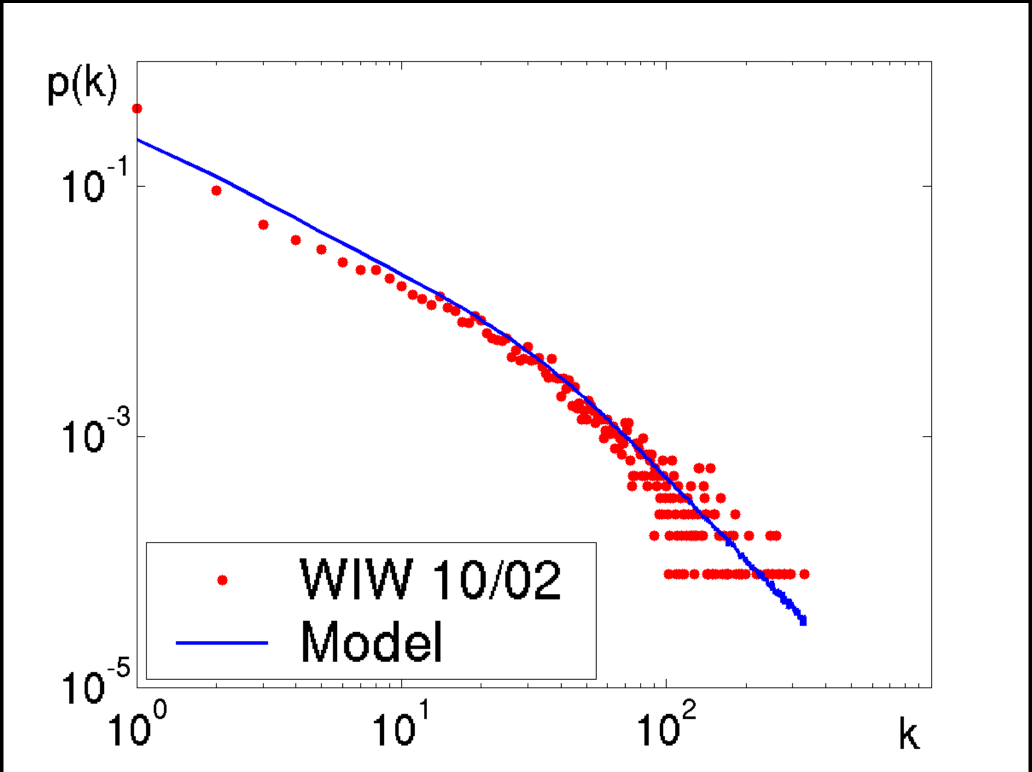
    for a critical value $q = q_{\text{crit}}$.

Degree distribution of our model



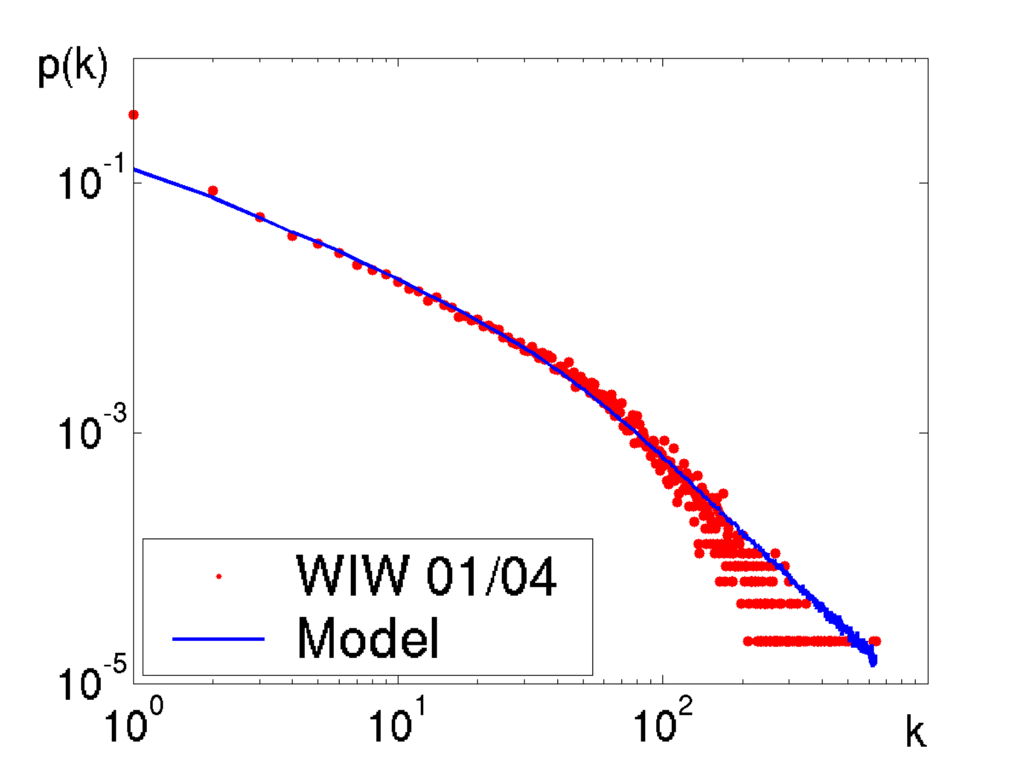
The fit of the degree distribution of our model

- The fit in October 2002



The fit of the degree distribution of our model

- The fit in January 2004 - degree bound has stronger effect?



Small-world properties of our model

- *Path length* Average path length of the model (with fixed average degree and arbitrary q) scales with the number V of vertices as

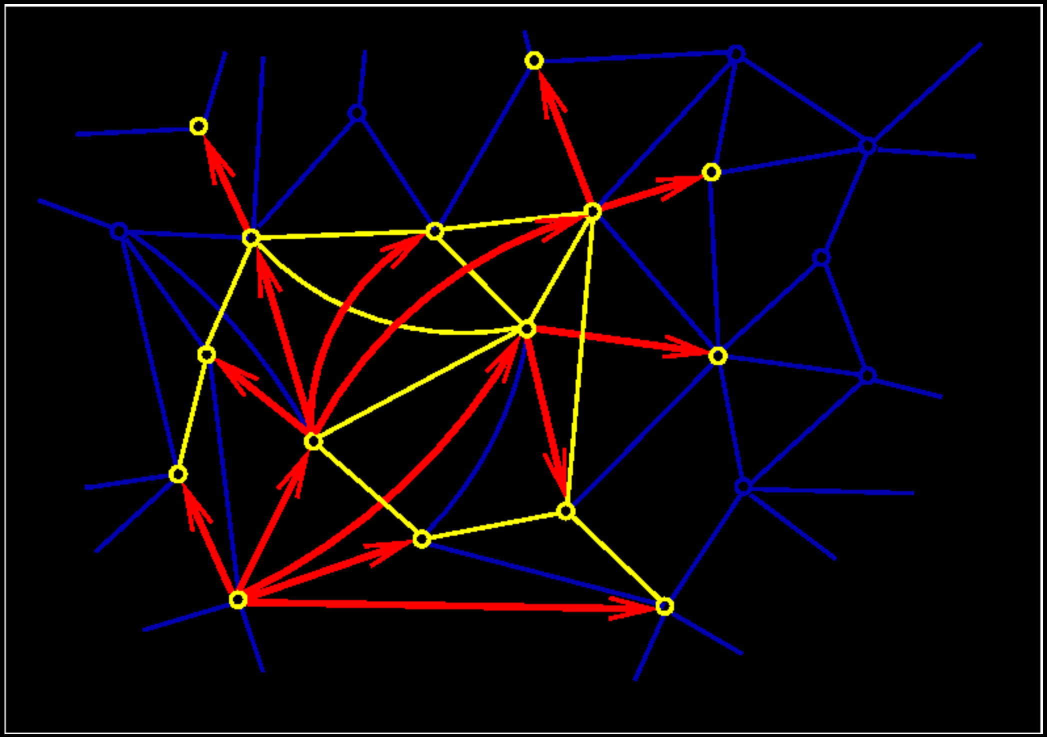
$$\langle l \rangle \sim \log V$$

  which is the natural notion of "small-world" behaviour in growing systems.

- *Clustering* Model produces graphs with high average clustering $\langle C \rangle$, independent of system size V.

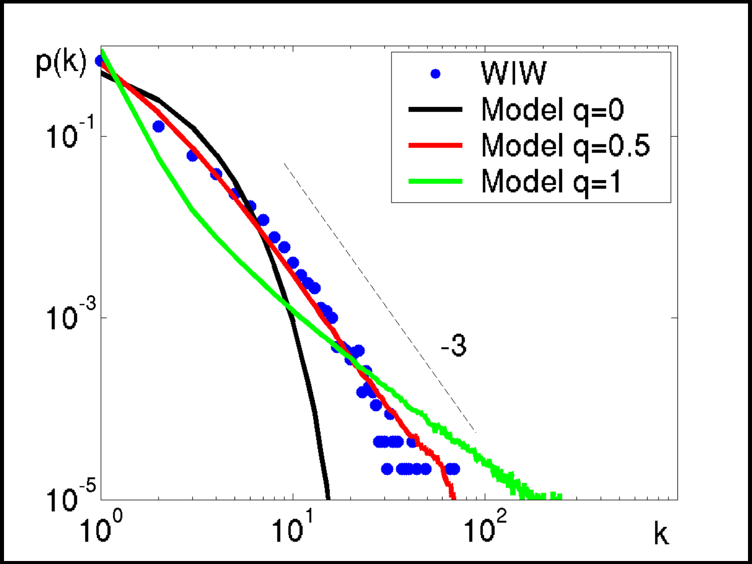
The invitation tree

Both WIW graph and model contain distinguished *invitation tree*



The degree distribution of the invitation tree

- Degree distribution of (oriented) invitation tree highly skewed, *different* from original distribution
- At $q = q_{\text{crit}}$ the model invitation tree has the right distribution



4. Conclusions

- Skewed power-law-type degree distribution in a growing social network
- High clustering in a social network
  - Likely to have an influence on spreading phenomena
- Triangle mechanism as a fundamental way of modelling contact
- Invitation mechanism crucially influences network structure
- Different networks present in a social network
  - "Strong links" along which invitation spreads
  - "Weak links" (Granovetter) along which acquintanceship is acknowledged

Outstanding issues

- Models based on local triangle creation are very natural in the context of social networks
  - Theoretical results about such models are lacking
- Find more precise graph models of social networks
  - Models should admit polynomial spreading/percolation processes
    - Lattices would work, but clearly constitute very poor models
  - Model should produce assortative graphs
    - High-degree vertices cluster together
- Analyze hierarchical structure of real-world networks
  - Clusters and groups