„Infokommunikációs technológiák és a jövő társadalma (FuturICT.hu)” projekt

TÁMOP-4.2.2.C-11/1/KONV-2012-0013

Graph Models of Social Networks

Dr. Kiss Attila kiss@inf.elte.hu
Eötvös Loránd University,
Budapest, Hungary

This work was partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013) and with the support of the Hungarian and Vietnamese TET (grant agreement no. TET 10-1-2011-0645).

Vietnam, 2014

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

What do the following things have in common?

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

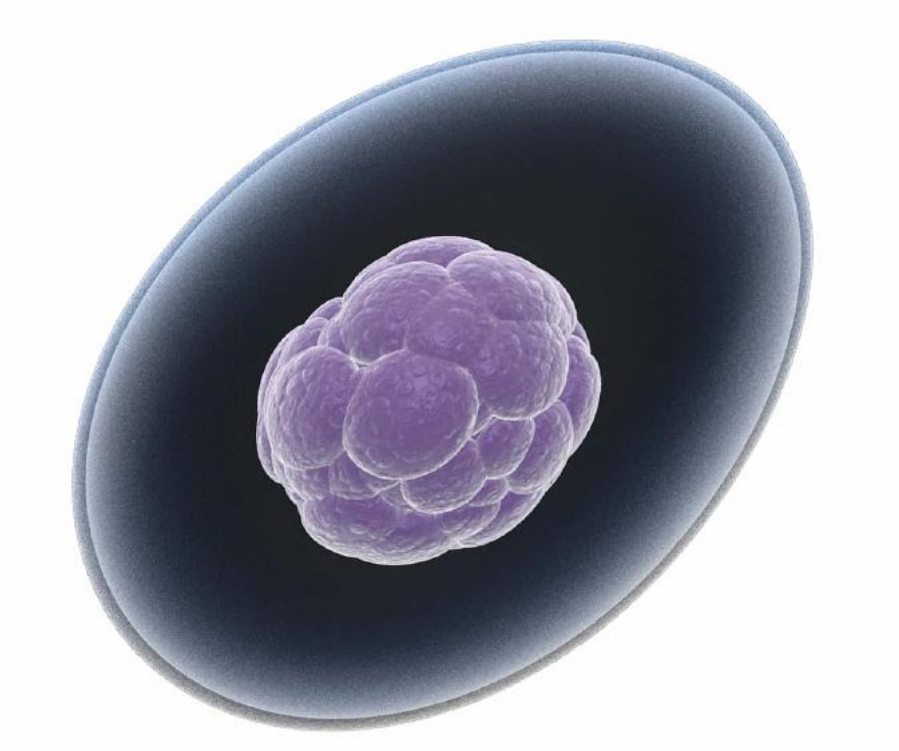
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

World economy

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

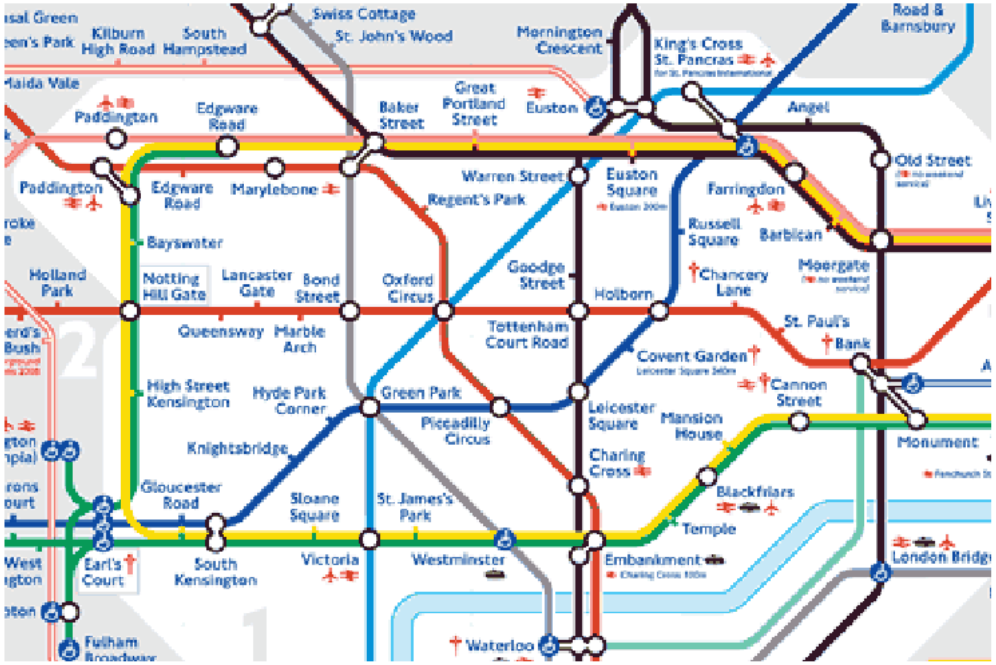


Human cell

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Railroads

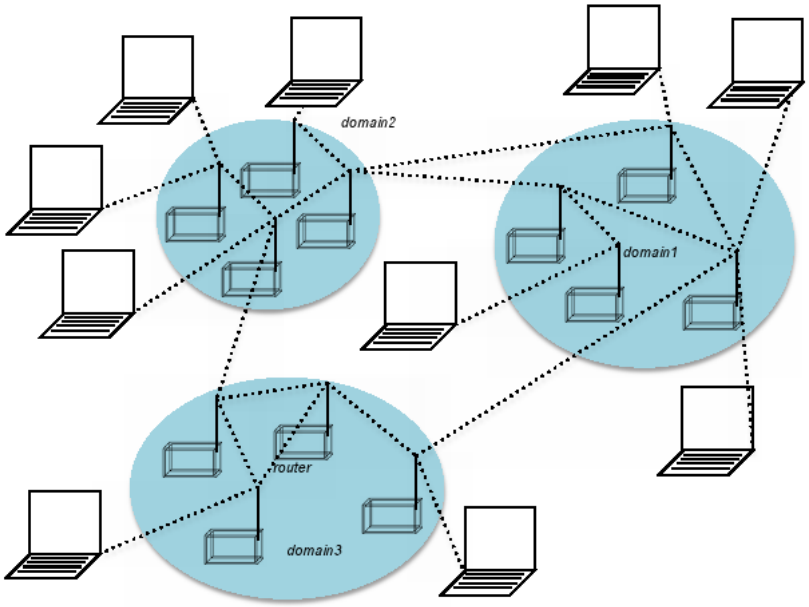
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Brain

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Internet

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

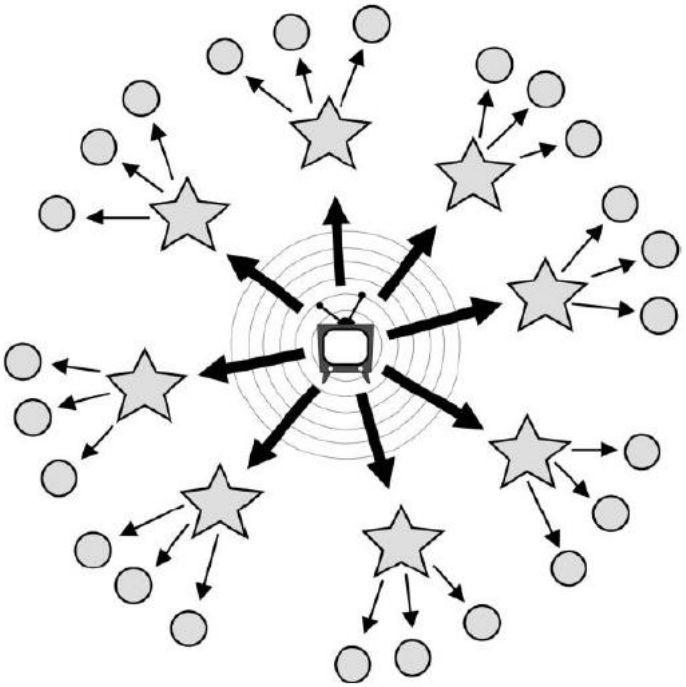
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Friends & Family

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

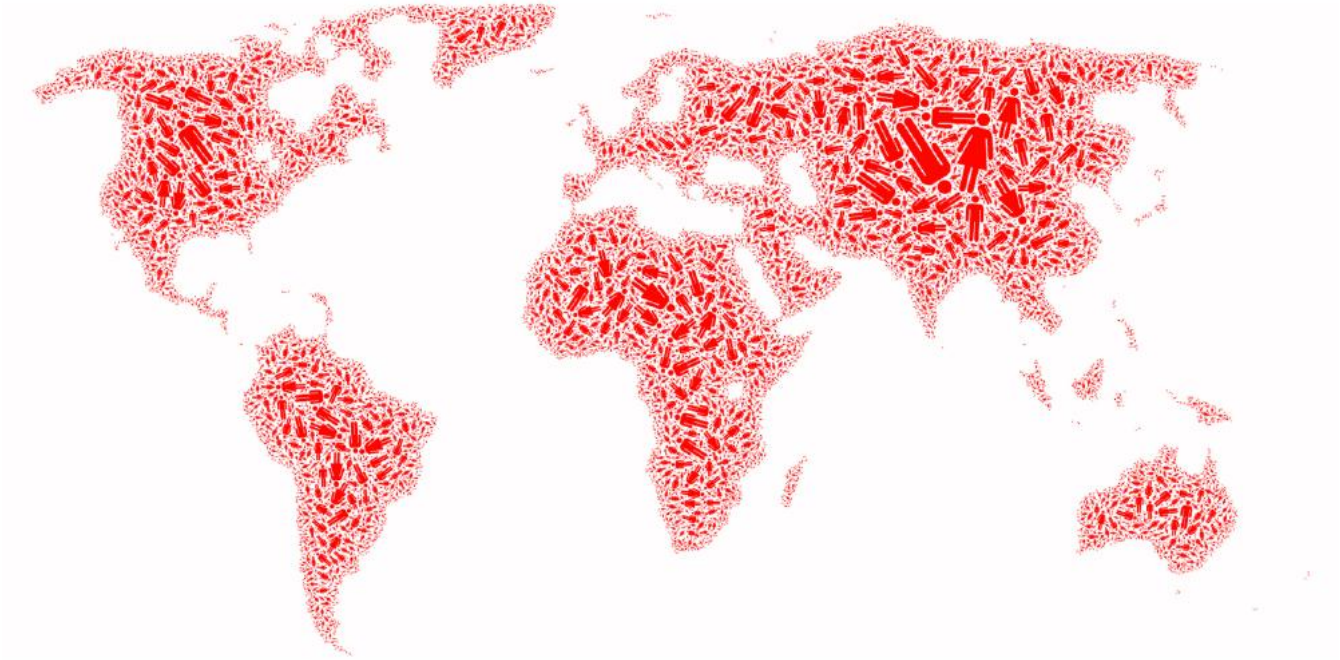


Media & Information

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

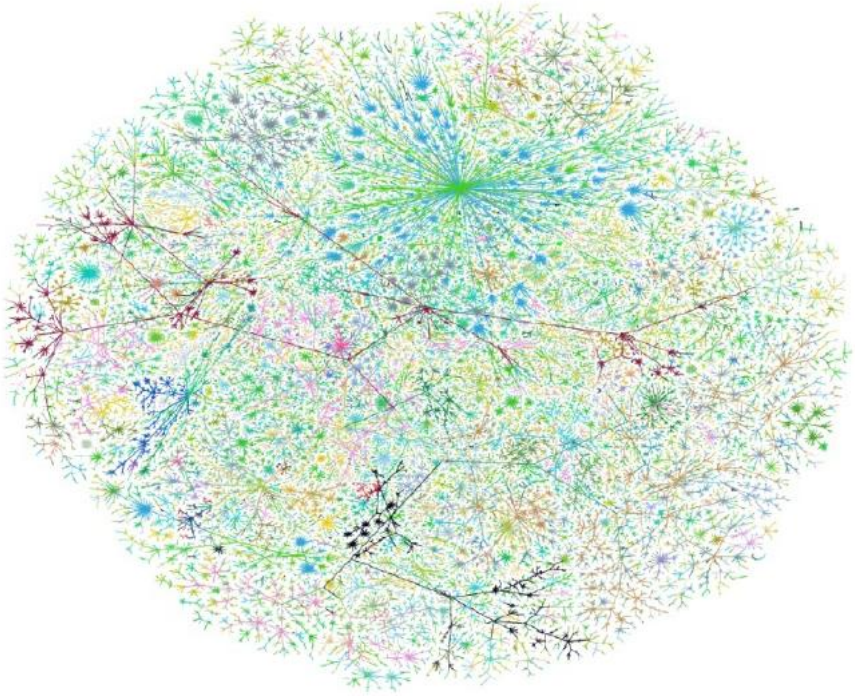


Society

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



The Network!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Networks!!

Behind each such system there is an intricate wiring diagram, **a network**, that defines the **interactions** between the components

We will never understand these systems unless we understand the networks behind them!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Why Networks? Why Now?

- **Universal language for describing complex data**
  - Networks from science, nature, and technology are more similar than one would expect
- **Shared vocabulary between fields**
  - Computer Science, Social science, Physics, Economics, Statistics, Biology
- **Data availability (/computational challenges)**
  - Web/mobile, bio, health, and medical
- **Impact!**
  - Social networking, Social media, Drug design

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Networks: Size Matters

- **Network data: Orders of magnitude**
  - 436-node network of email exchange at a corporate research lab [Adamic-Adar, SocNets ‘03]
  - 43,553-node network of email exchange at an university [Kossinets-Watts, Science ‘06]
  - 4.4-million-node network of declared friendships on a blogging community [Liben-Nowell et al., PNAS ‘05]
  - 240-million-node network of communication on Microsoft Messenger [Leskovec-Horvitz, WWW ’08]
  - 800-million-node Facebook network [Backstrom et al. ‘11]

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Reasoning about Networks

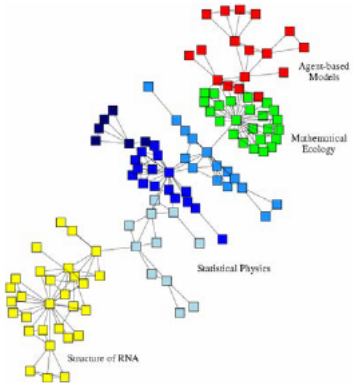
- **How do we reason about networks?**
  - **Empirical:** Study network data to find organizational principles
    - How do we measure and quantify networks?
  - **Mathematical models:** Graph theory, statistical models
    - Models allow us to understand behaviors and distinguish surprising from expected phenomena
  - **Algorithms** for analyzing graphs
    - Hard computational challenges

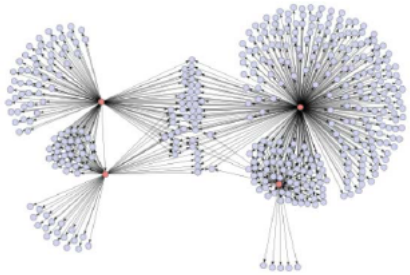
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Networks: Structure & Process

What do we study in networks?

- **Structure and evolution:**
  - What is the structure of a network?
  - Why and how did it became to have such structure?
- **Processes and dynamics:**
  - Networks provide “skeleton” for spreading of information, behavior, diseases
  - How do information and diseases spread?

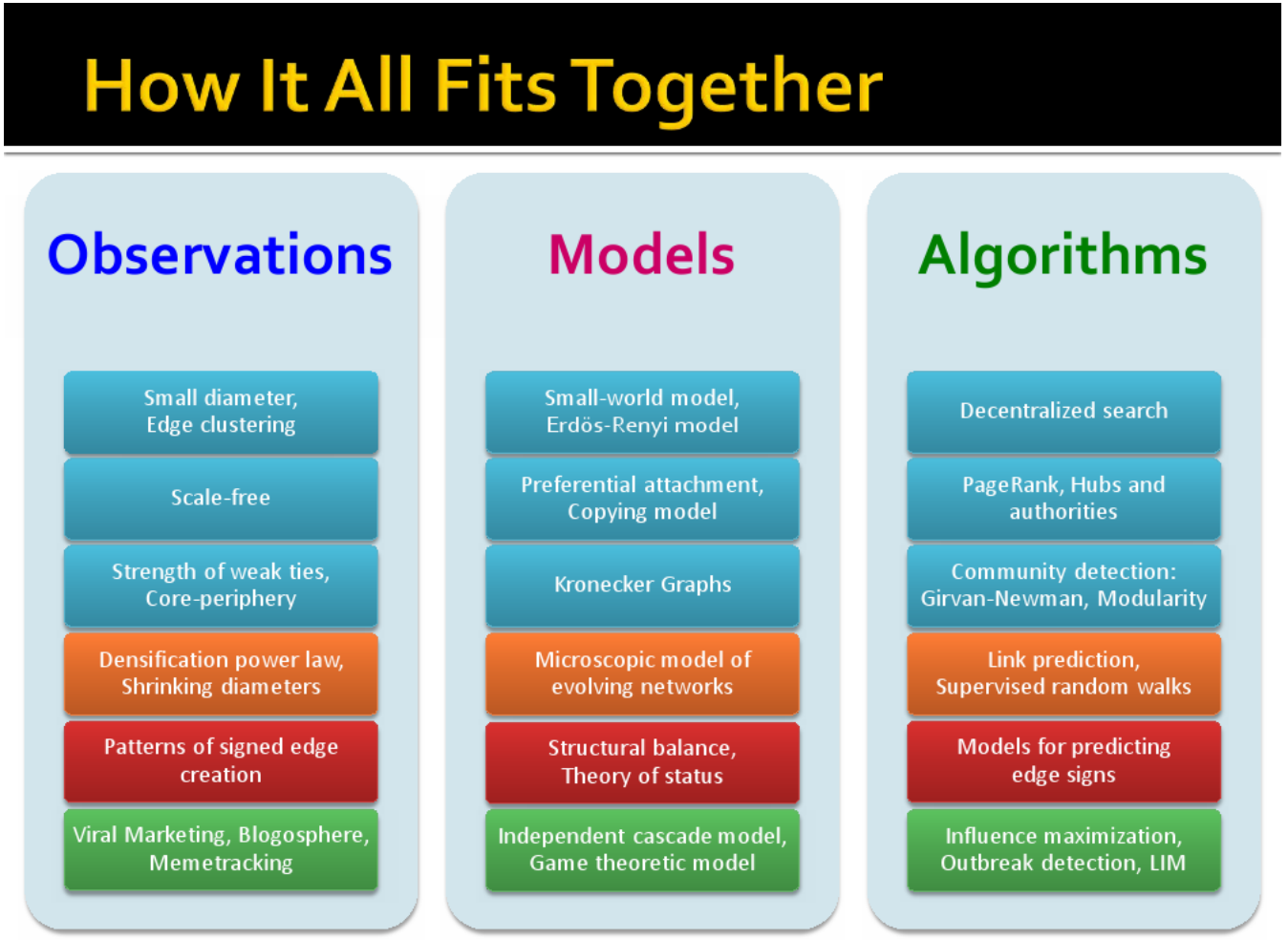




Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

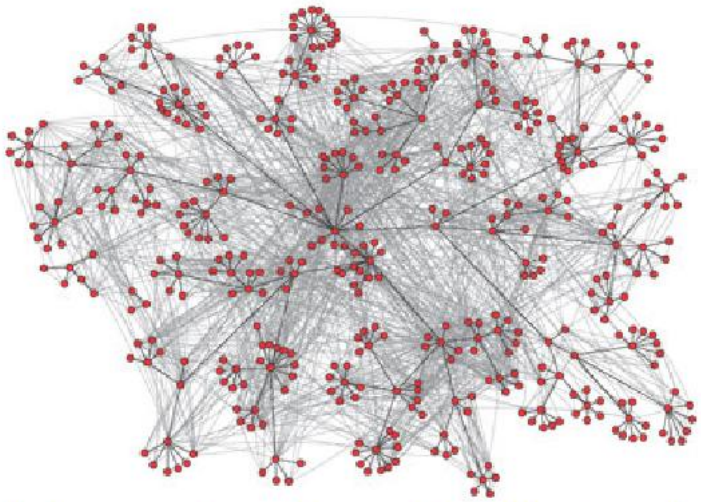
Starter Topic: Structure of the Web Graph

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Structure of Networks?



Network is a collection of objects where some pairs of objects are connected by links

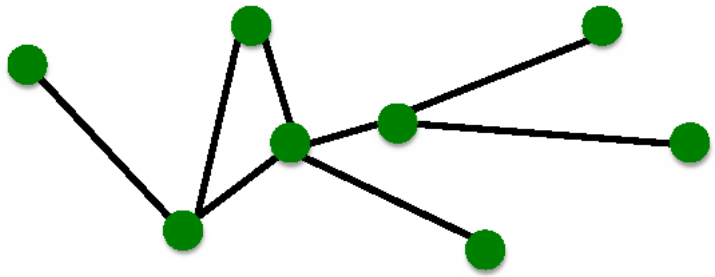
What is the structure of the network?

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Components of a Network



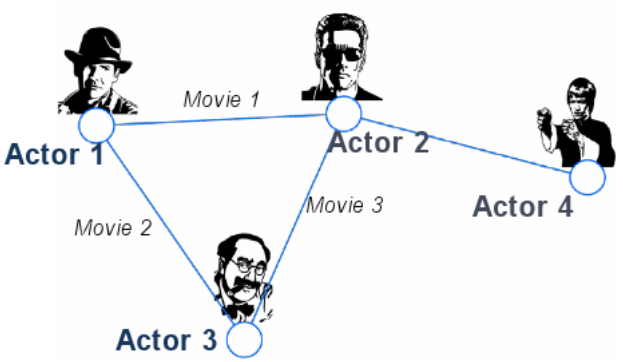
- **Objects:** nodes, vertices — N
- **Interactions:** links, edges — E
- **System:** network, graph — $G(N,E)$

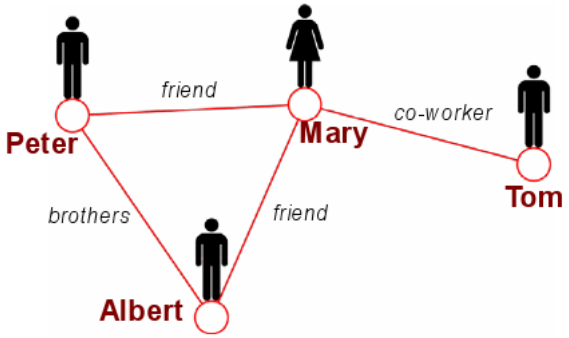
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

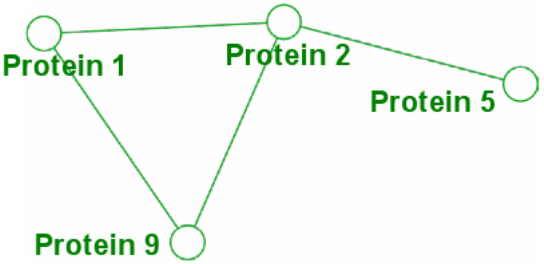
MAGYARORSZÁG MEGÚJUL

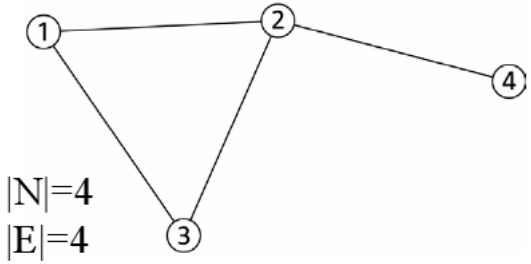
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Networks: Common Language









Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

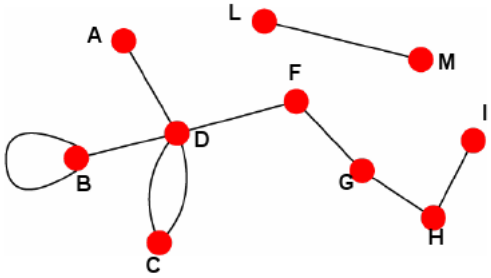
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Undirected vs. Directed Networks

Undirected

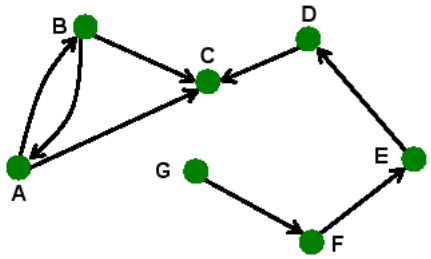
- **Links:** undirected (symmetrical, reciprocal)



- **Examples:**
  - Collaborations
  - Friendship on Facebook

Directed

- **Links:** directed (arcs)



- **Examples:**
  - Phone calls
  - Following on Twitter

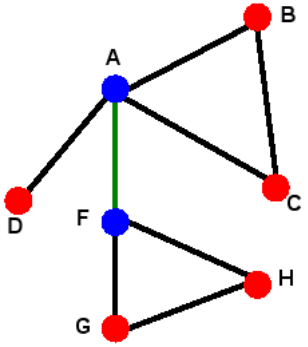
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

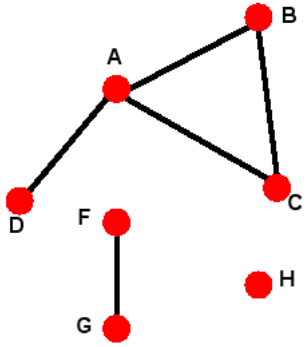
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Connectivity of Graphs

- **Connected (undirected) graph:**
  - Any two vertices can be joined by a path
- A disconnected graph is made up by two or more connected components





Largest Component:
Giant Component

Isolated node (node H)

Bridge edge: If we erase it, the graph becomes disconnected.
Articulation point: If we erase it, the graph becomes disconnected.

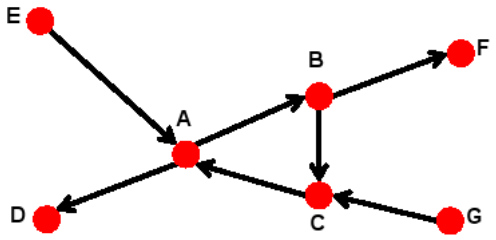
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Connectivity of Directed Graphs

- **Strongly connected directed graph**
  - has a path from each node to every other node and vice versa (e.g., A-B path and B-A path)
- **Weakly connected directed graph**
  - is connected if we disregard the edge directions



Graph on the left is connected but not strongly connected (e.g., there is no way to get from F to G by following the edge directions).

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

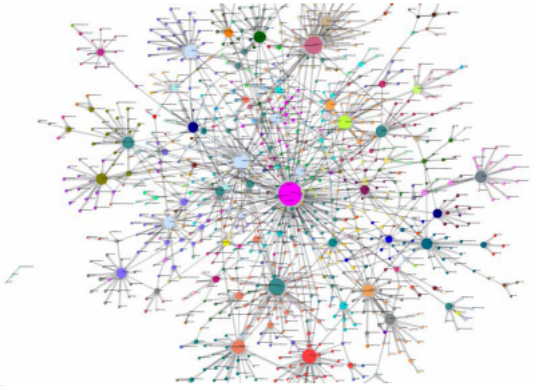
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Web as a Graph

Q: What does the Web “look like” at a global level?

- **Web as a graph:**
  - Nodes = web pages
  - Edges = hyperlinks
  - **Side issue:** What is a node?
    - Dynamic pages created on the fly
    - “dark matter” – inaccessible database generated pages



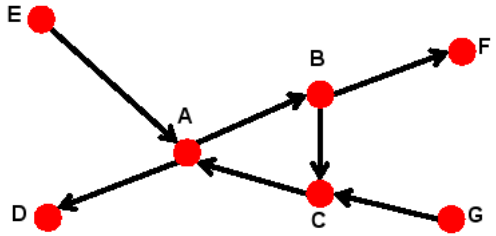
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

What Does the Web Look Like?

- How is the Web linked?
- What is the “map” of the Web?

Web as a directed graph [Broder et al. 2000]:

- Given node v, what can v reach?
- What other nodes can reach v?



$In(v) = \{w \mid w \text{ can reach } v\}$

$Out(v) = \{w \mid v \text{ can reach } w\}$

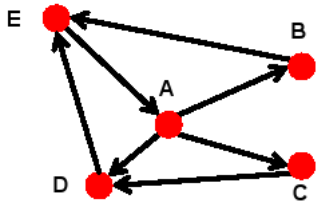
For example:
In(A) = {A,B,C,E,G}
Out(A)={A,B,C,D,F}

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

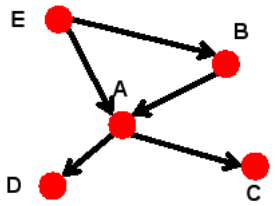
Directed Graphs

- **Two types of directed graphs:**
  - **Strongly connected:**
    - Any node can reach any node via a directed path

      $In(A)=Out(A)=\{A,B,C,D,E\}$



  - **DAG – Directed Acyclic Graph:**
    - Has no cycles: if u can reach v, then v can not reach u



- **Any directed graph can be expressed in terms of these two types!**

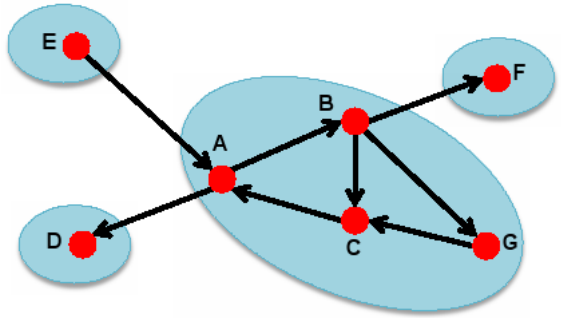
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Strongly Connected Component

- **Strongly connected component (SCC)** is a set of nodes S so that:
  - Every pair of nodes in S can reach each other
  - There is no larger set containing S with this property



Strongly connected components of the graph: {A,B,C,G}, {D}, {E}, {F}

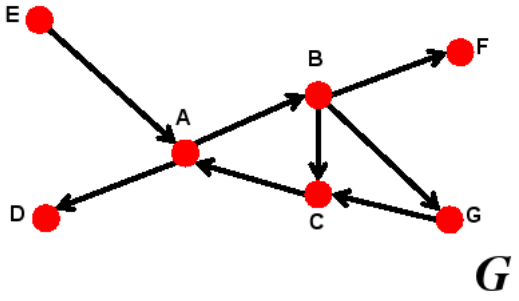
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

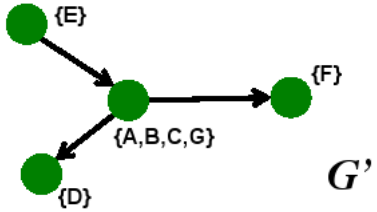
Strongly Connected Component

- **Fact: Every directed graph is a DAG on its SCCs**
  - **(1)** SCCs partitions the nodes of G
    - Each node is in exactly one SCC
  - **(2)** If we build a graph G' whose nodes are SCCs, and with an edge between nodes of G' if there is an edge between corresponding SCCs in G, then G' is a DAG



G

(1) Strongly connected components of graph G: {A,B,C,G}, {D}, {E}, {F}
(2) G' is a DAG:



G'

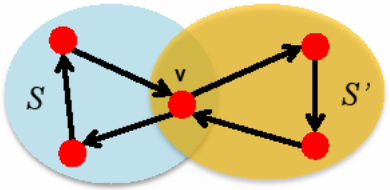
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Proof of (1)

- **Claim: SCCs partitions nodes of G.**
  - This means: Each node is member of exactly 1 SCC
- Proof by contradiction:
  - Suppose there exists a node v which is a member of 2 SCCs S and S'



  - But then $S \cup S'$ is one large SCC!
    - Contradiction!

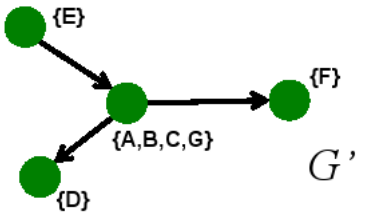
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

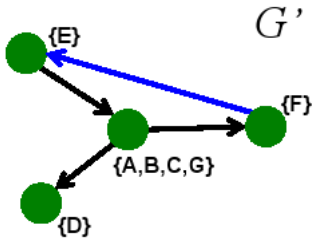
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Proof of (2)

- **Claim: G' (graph of SCCs) is a DAG.**
  - This means: G' has no cycles
- Proof by contradiction:
  - Assume G' is not a DAG
  - Then G' has a directed cycle
  - Now all nodes on the cycle are mutually reachable, and all are part of the same SCC
  - But then G' is not a graph of connections between SCCs (SCCs are defined as maximal sets)
    - Contradiction!





Now {A,B,C,G,E,F} is a SCC!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

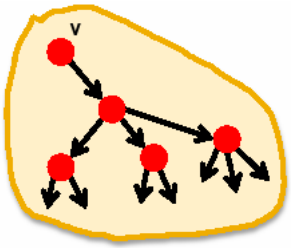
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

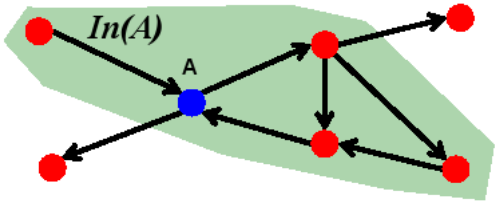
Graph Structure of the Web

- **Goal:** Take a large snapshot of the Web and try to understand how its SCCs “fit together” as a DAG
- **Computational issue:**
  - Want to find a SCC containing node v?
  - **Observation:**
    - ***Out(v)*** … nodes that can be reached from v
    - **SCC containing v is:** $Out(v) \cap In(v)$

      $= Out(v,G) \cap Out(v,\overline{G})$, where $\overline{G}$ is G with all edge directions flipped



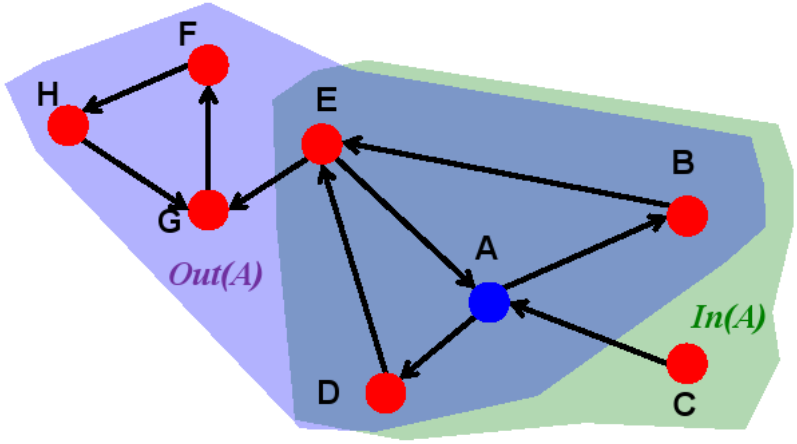
Out(v)



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Out(A) ∩ In(A) = SCC

- **Example:**



- Out(A) = {A, B, D, E, F, G, H}
- In(A) = {A, B, C, D, E}
- So, SCC(A) = Out(A) ∩ In(A) = {A, B, D, E}

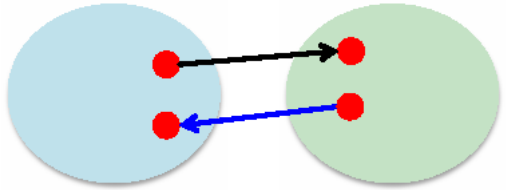
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Graph Structure of the Web

- There is a single giant SCC
- There won't be 2 giant SCCs
- Heuristic argument:
  - It just takes 1 page from one SCC to link to the other SCC
  - If the 2 SCCs have millions of pages the likelihood of this not happening is very very small



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Structure of the Web

- **Broder et al., 2000:**
  - Altavista crawl from October 1999
    - 203 million URLS
    - 1.5 billion links
  - Computer: Server with 12GB of memory
- **Undirected version of the Web graph:**
  - 91% nodes in the largest weakly conn. component
  - Are hubs making the web graph connected?
    - Even if they deleted links to pages with in-degree >10 WCC was still ≈50% of the graph

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Structure of the Web

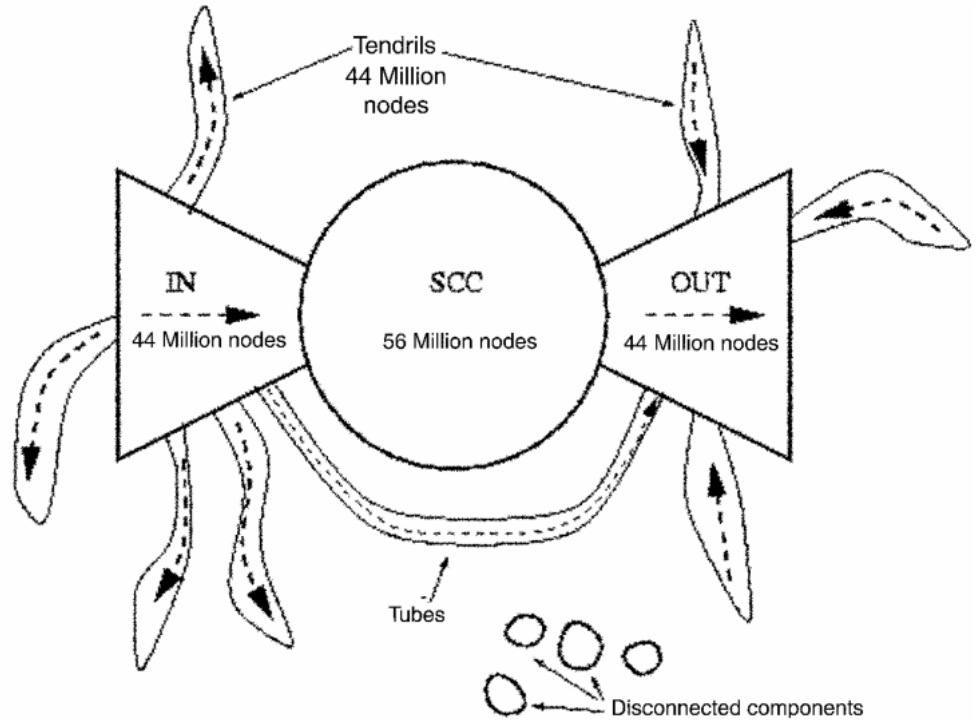
- **Directed version of the Web graph:**
  - **Largest SCC:** 28% of the nodes (56 million)
  - Taking a random node v
    - **Out(v) ≈ 50%** (100 million)
    - **In(v) ≈ 50%** (100 million)
- **What does this tell us about the conceptual picture of the Web graph?**

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Bow-tie Structure of the Web



203 million pages, 1.5 billion links [Broder et al. 2000]

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Basic Network Properties and the Random Graph Model

CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

http://cs224w.stanford.edu

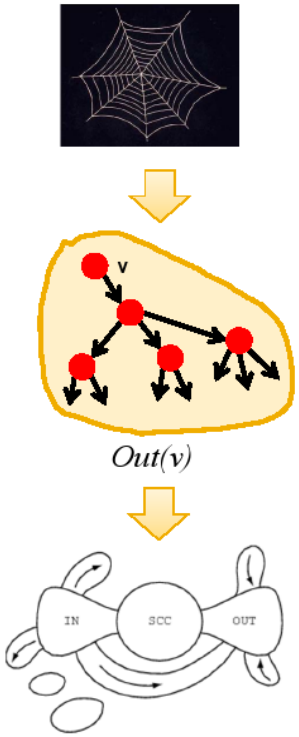
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Structure of Networks

- **For example, last time we talked about Observations and Models for the Web graph:**
  - **1)** We took a real system: **the Web**
  - **2)** We represented it as a **directed graph**
  - **3)** We used the language of graph theory
    - **Strongly Connected Components**
  - **4)** We designed a **computational experiment:**
    - Find In- and Out-components of a given node ***v***
  - **5) We learned something about the structure of the Web: BOWTIE!**

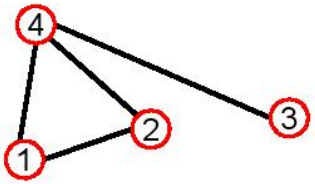


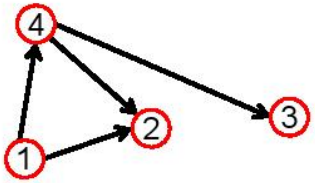
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Adjacency Matrix





$A_{ij} = 1$ if there is a link from node i to node j

$A_{ij} = 0$ otherwise

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \qquad A = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$

Note that for a directed graph (right) the matrix is not symmetric.

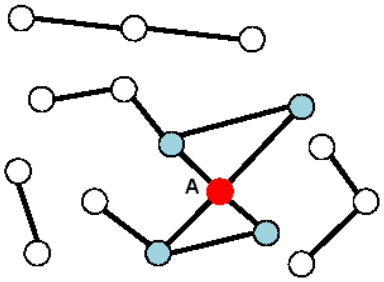
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Node Degrees

Undirected

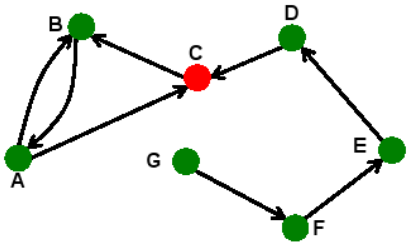


Node degree, k_i: the number of edges adjacent to node i

$$k_A = 4$$

Avg. degree: $\bar{k} = \langle k \rangle = \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2E}{N}$

Directed



In directed networks we define an **in-degree** and **out-degree.**

The (total) degree of a node is the sum of in- and out-degrees.

$$k_C^{in} = 2 \qquad k_C^{out} = 1 \qquad k_C = 3$$

Source: node with $k^{in} = 0$

Sink: node with $k^{out} = 0$

$$\bar{k} = \frac{E}{N} \qquad \overline{k^{in}} = \overline{k^{out}}$$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

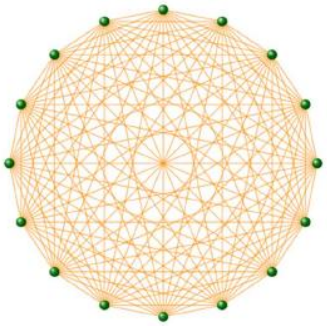
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Complete Graph

The **maximum number of edges** in an undirected graph on N nodes is

$$E_{\max} = \binom{N}{2} = \frac{N(N-1)}{2}$$



A graph with the number of edges $E = E_{max}$ is a **complete graph**, and its average degree is $N-1$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Networks are Sparse Graphs

Most real-world networks are sparse

$$E << E_{max} \quad (\text{or } \bar{k} << N-1)$$

WWW (Stanford-Berkeley):	N=319,717	⟨k⟩=9.65
Social networks (LinkedIn):	N=6,946,668	⟨k⟩=8.87
Communication (MSN IM):	N=242,720,596	⟨k⟩=11.1
Coauthorships (DBLP):	N=317,080	⟨k⟩=6.62
Internet (AS-Skitter):	N=1,719,037	⟨k⟩=14.91
Roads (California):	N=1,957,027	⟨k⟩=2.82
Proteins (S. Cerevisiae):	N=1,870	⟨k⟩=2.39

(Source: *Leskovec et al., Internet Mathematics, 2009*)

Consequence: Adjacency matrix is filled with zeros!

(**Density of the matrix (E/N^2):** WWW=1.51×10^{-5}, MSN IM = 2.27×10^{-8})

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Network Representations

WWW >> directed multigraph with self-edges

Facebook friendships >> undirected, unweighted

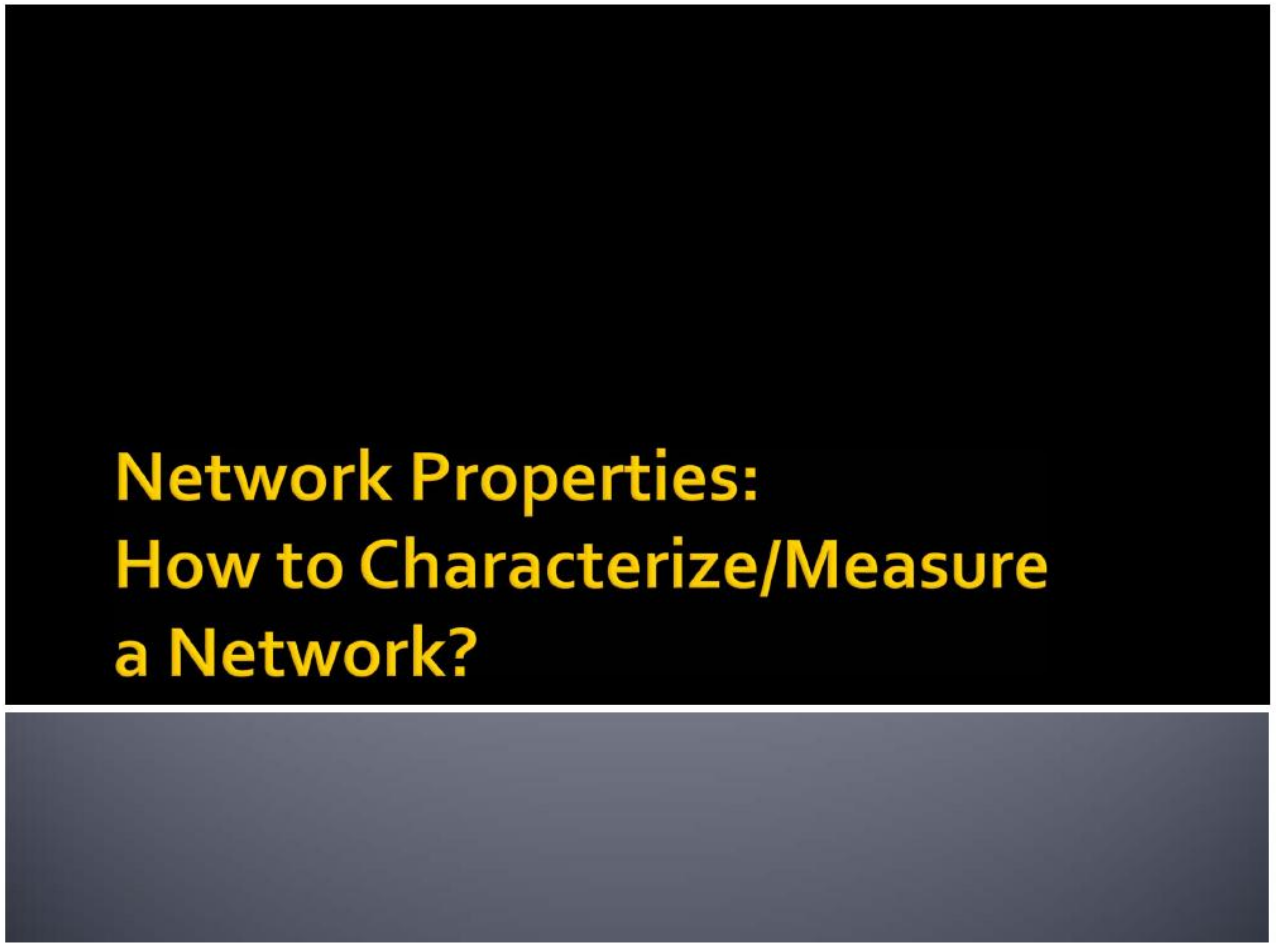
Citation networks >> unweighted, directed, acyclic

Collaboration networks >> undirected multigraph or weighted graph

Mobile phone calls >> directed, (weighted?) multigraph

Protein Interactions >> undirected, unweighted with self-interactions

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

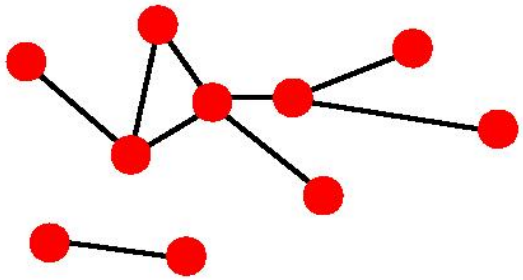
Degree Distribution

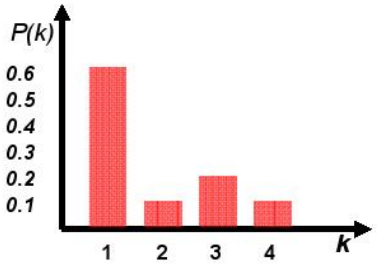
- **Degree distribution $P(k)$:** Probability that a randomly chosen node has degree k

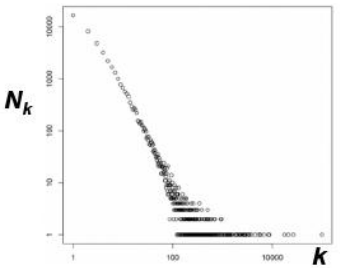
  N_k = # nodes with degree k

- Normalized histogram:

  $P(k) = N_k / N$ → **plot**







Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

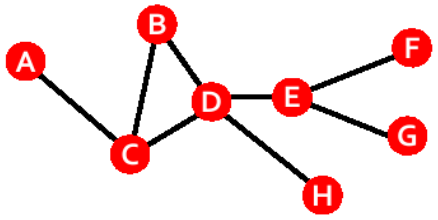
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Paths in a Graph

- A ***path*** is a sequence of nodes in which each node is linked to the next one

$$P_n = \{i_0, i_1, i_2, \ldots, i_n\} \qquad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \ldots, (i_{n-1}, i_n)\}$$

- Path can intersect itself and pass through the same edge multiple times
  - E.g.: ACBDCDEG
  - In a directed graph a path can only follow the direction of the “arrow”



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

EXTRA: Number of Paths

- **Number of paths between nodes u and v :**
  - **Length $h=1$:** If there is a link between u and v, $A_{uv}=1$ else $A_{uv}=0$
  - **Length $h=2$:** If there is a path of length two between u and v then $A_{uk}A_{kv}=1$ else $A_{uk}A_{kv}=0$

$$H_{uv}^{(2)} = \sum_{k=1}^{N} A_{uk} A_{kv} = [A^2]_{uv}$$

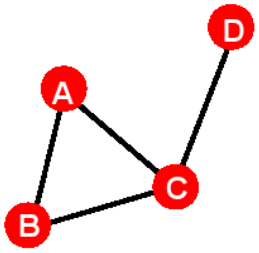
  - **Length h:** If there is a path of length h between u and v then $A_{uk} \ldots A_{kv}=1$ else $A_{uk} \ldots A_{kv}=0$
    So, the no. of paths of length h between u and v is

$$H_{uv}^{(h)} = [A^h]_{uv}$$

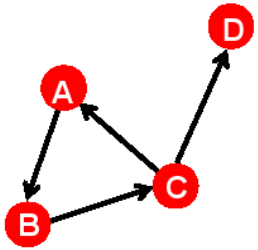
(holds for both directed and undirected graphs)

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Distance in a Graph



$h_{B,D} = 2$



$h_{B,C} = 1$, $h_{C,B} = 2$

- **Distance (shortest path, geodesic)** between a pair of nodes is defined as the number of edges along the shortest path connecting the nodes
  - *If the two nodes are disconnected, the distance is usually defined as infinite
- In **directed graphs** paths need to follow the direction of the arrows
  - Consequence: Distance is **not symmetric**: $h_{A,C} \neq h_{C,A}$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Network Diameter

- **Diameter:** the maximum (shortest path) distance between any pair of nodes in a graph
- **Average path length** for a connected graph (component) or a strongly connected (component of a) directed graph

$$\bar{h} = \frac{1}{2E_{\max}} \sum_{i,j \neq i} h_{ij}$$

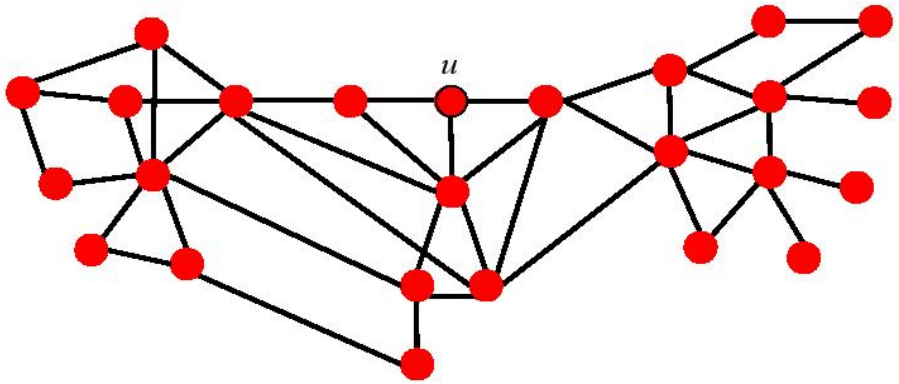
where h_{ij} is the distance from node i to node j

  - Many times we compute the average only over the connected pairs of nodes (we ignore “infinite” length paths)

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Finding Shortest Paths

- **Breath-First Search:**
  - Start with node u, mark it to be at distance $h_u(u)=0$, add u to the queue
  - While the queue not empty:
    - Take node v off the queue, put its unmarked neighbors w into the queue and mark $h_u(w)=h_u(v)+1$



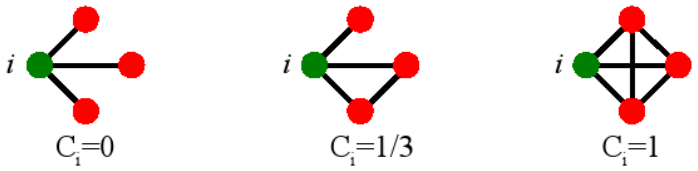
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Clustering Coefficient

- **Clustering coefficient:**
  - What portion of ***i***'s neighbors are connected?
  - Node ***i*** with degree k_i
  - $C_i \in [0,1]$
  - $C_i = \frac{2e_i}{k_i(k_i-1)}$ where e_i is the number of edges between the neighbors of node i

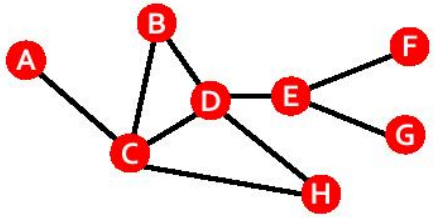


- **Average Clustering Coefficient:** $C = \frac{1}{N}\sum_{i}^{N} C_i$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Clustering Coefficient

- **Clustering coefficient:**
  - What portion of i's neighbors are connected?
  - Node i with degree k_i
  - $C_i = \frac{2e_i}{k_i(k_i-1)}$ where e_i is the number of edges between the neighbors of node i



$k_B=2,\ e_B=1,\ C_B=2/2=1$

$k_D=4,\ e_D=2,\ C_D=4/12=1/3$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

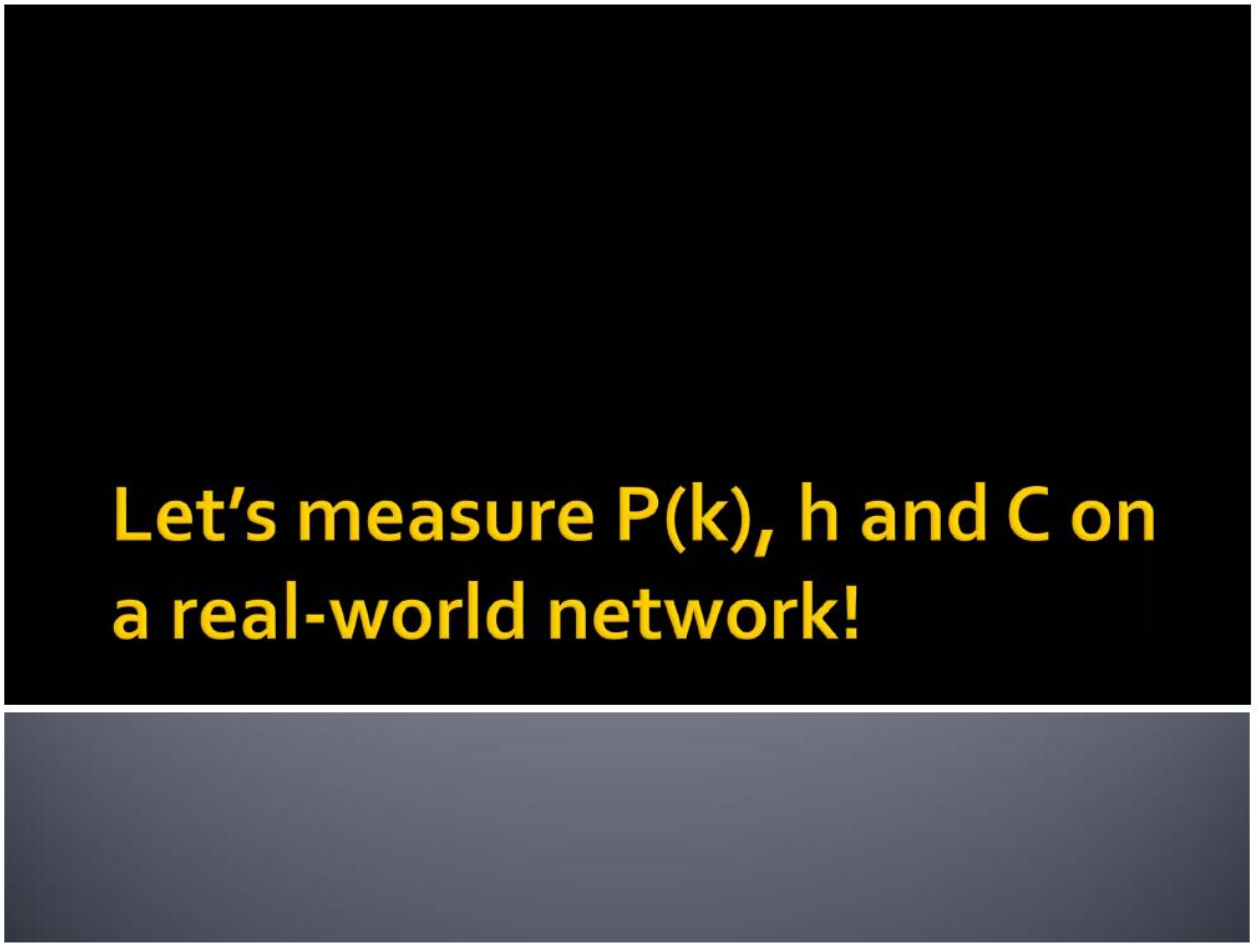
Key Network Properties

Degree distribution: $P(k)$

Path length: h

Clustering coefficient: C

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

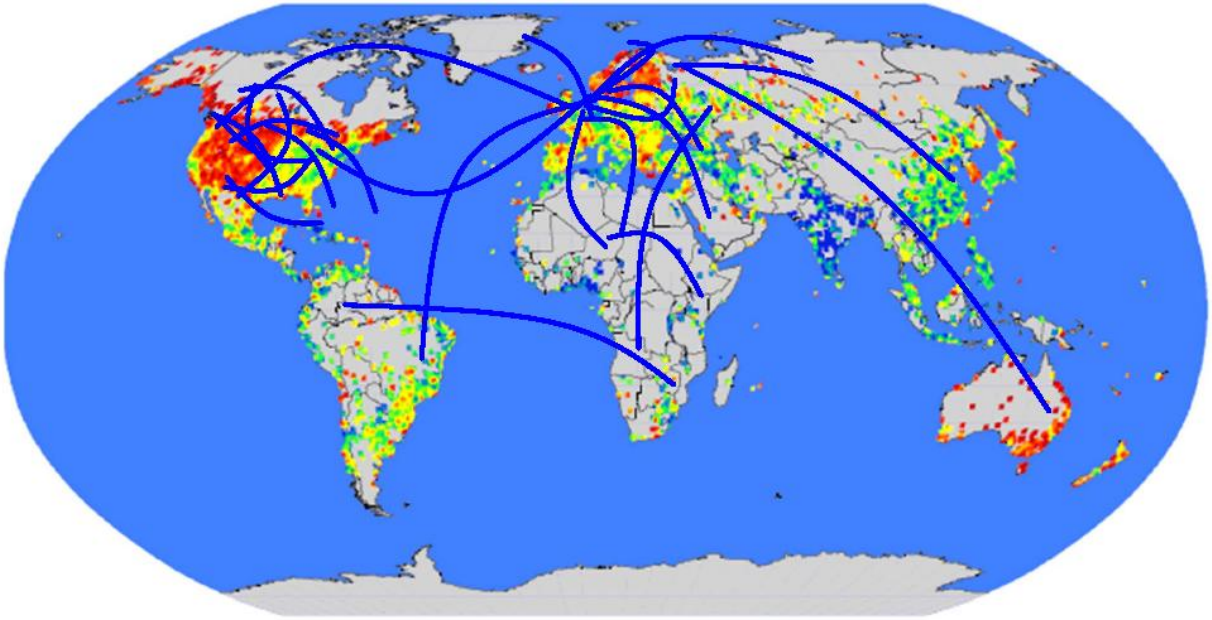
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

The MSN Messenger

- **MSN Messenger activity in June 2006:**
  - 150Gb/day (compressed)
  - 4.5Tb / month
  - 245 million users logged in
  - 180 million users engaged in conversations
  - More than 30 billion conversations
  - More than 255 billion exchanged messages

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Communication network



Network: 180M people, 1.3B edges

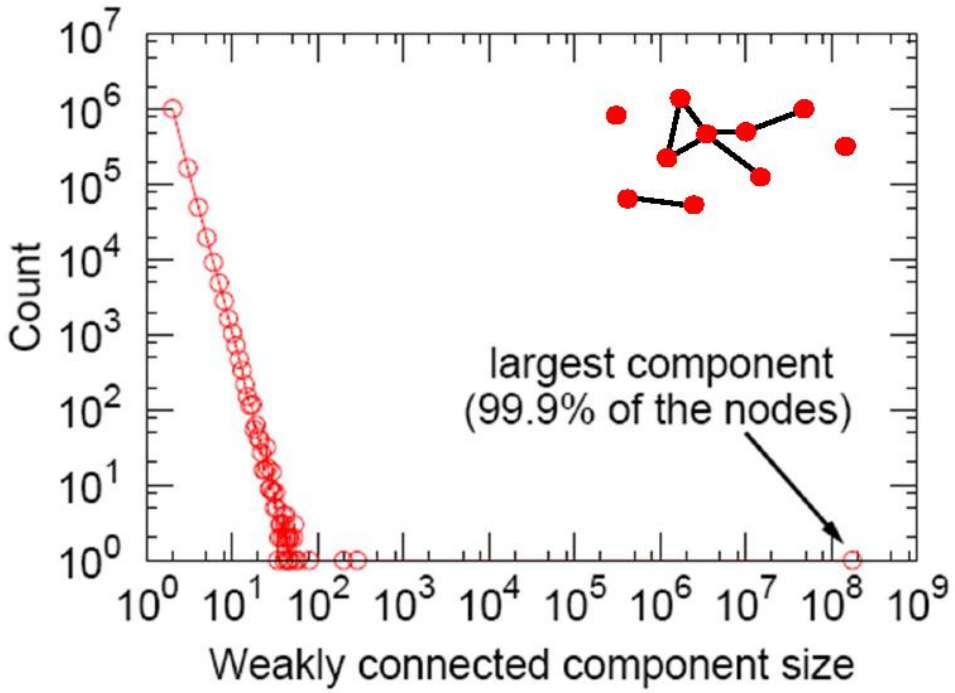
27

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN Network: Connectivity

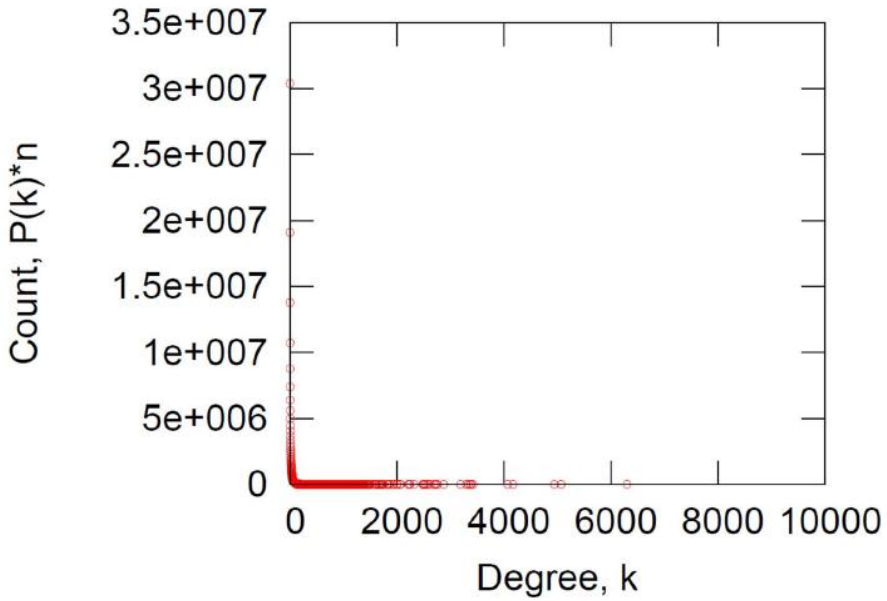


Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN: Degree Distribution

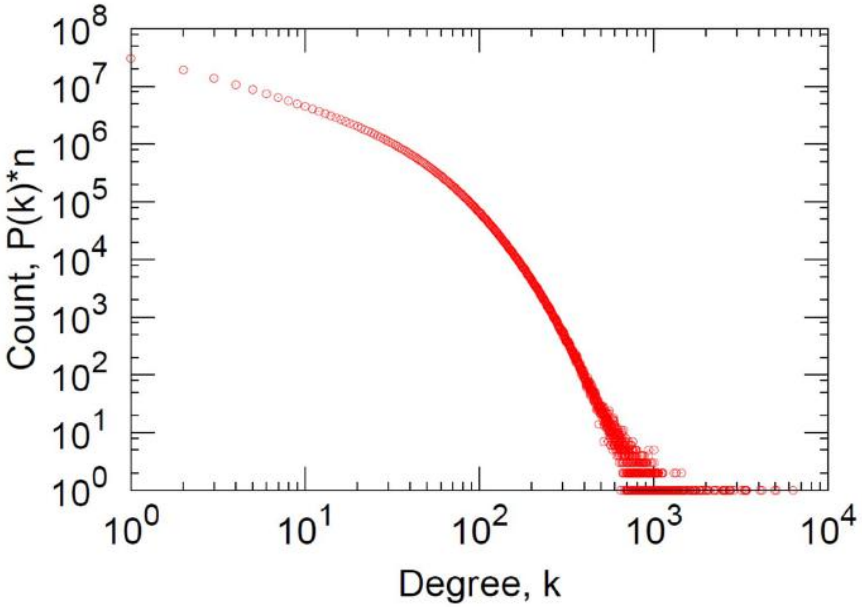


Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN: Log-Log Degree Distribution



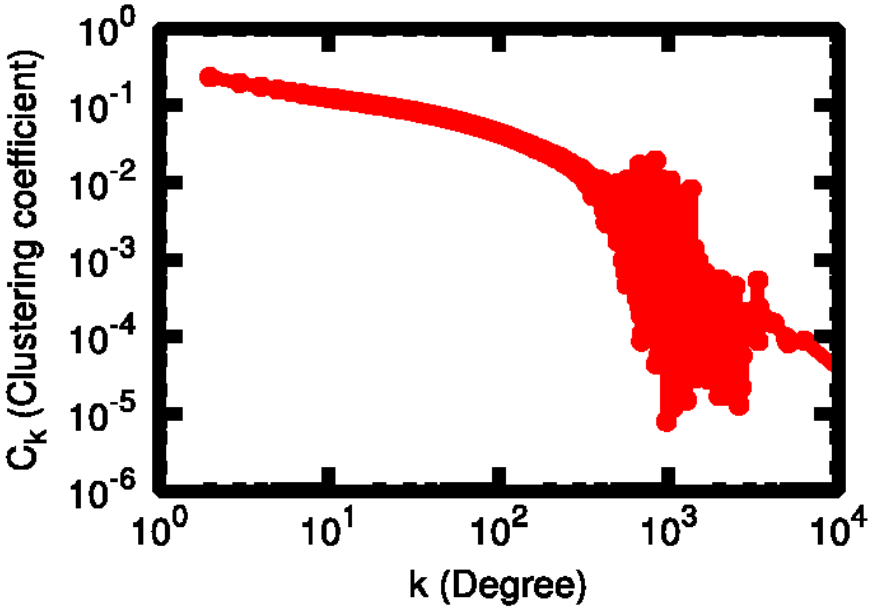
We plot the same data as on the previous slide, just the axes are now logarithmic.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN: Clustering



Avg. clustering of the MSN:
C = 0.1140

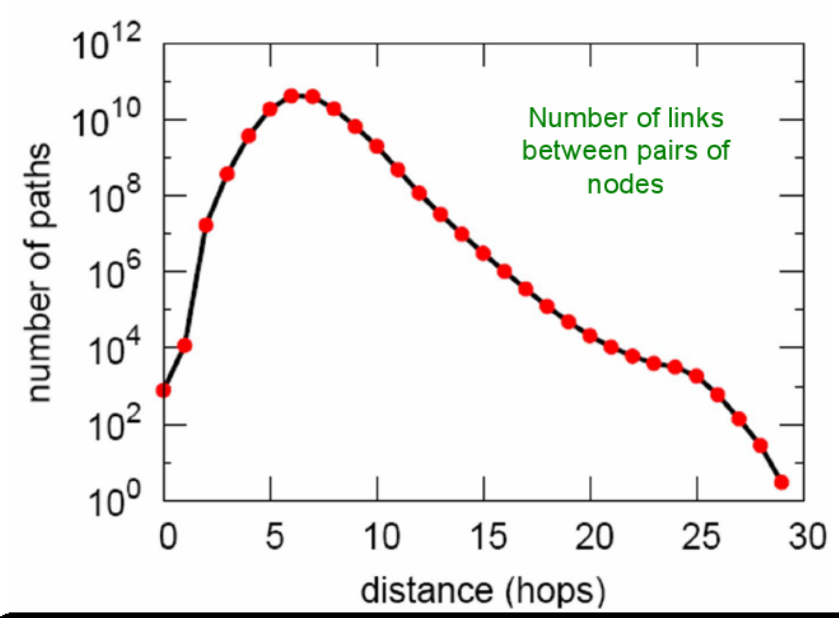
C_k: average C_i of nodes *i* of degree *k*: $C_k = \frac{1}{N_k} \sum_{i:k_i=k} C_i$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN: Diameter



Avg. path length **6.6**
90% of the people can be reached in < 8 hops

nodes as we do BFS out of a random node

Steps	#Nodes
0	1
1	10
2	78
3	3,96
4	8,648
5	3,299,252
6	28,395,849
7	79,059,497
8	52,995,778
9	10,321,008
10	1,955,007
11	518,410
12	149,945
13	44,616
14	13,740
15	4,476
16	1,542
17	536
18	167
19	71
20	29
21	16
22	10
23	3
24	2
25	3

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN: Key Network Properties

Degree distribution: *heavily skewed*

avg. degree= 14.4

Path length: *6.6*

Clustering coefficient: *0.11*

Are these metrics “expected”?
Are they “surprising”?

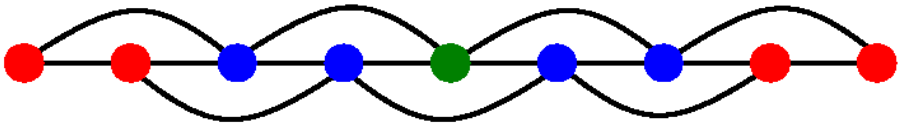
To answer this we need a null-model!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Is MSN Network like a “chain”?



- $P(k) = \delta(k-4)$ $\quad k_i = 4$ for all nodes
- $C = ½$ all as $N \to \infty$
- Path length: $h_{max} = \left\lceil \frac{N-1}{2} \right\rceil = O(N)$
  - The average shortest path-length: $\bar{h} = O(N)$
- So, we have: **Constant degree, Constant avg. clustering coeff. Linear avg. path-length**

Note about calculations:
We are interested in quantities as graphs get large ($N \to \infty$)

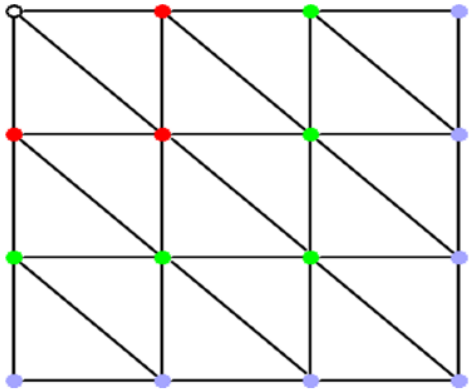
We will use big-O:
$f(x) = O(g(x))$ as $x \to \infty$
if $f(x) < g(x)*c$ for all $x > x_0$
and some constant c.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Is MSN Network like a “grid”?



- $P(k) = \delta(k-6)$
  - $k = 6$ for each inside node
- $C = 6/15$ for inside nodes
- **Path length:**

$$h_{max} = O(\sqrt{N})$$

- **In general, for lattices:**
  - Average path-length is $\bar{h} \approx N^{1/D}$ (D... lattice dimensionality)
  - Constant degree, constant clustering coefficient

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Erdös-Renyi Random Graph Model

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Simplest Model of Graphs

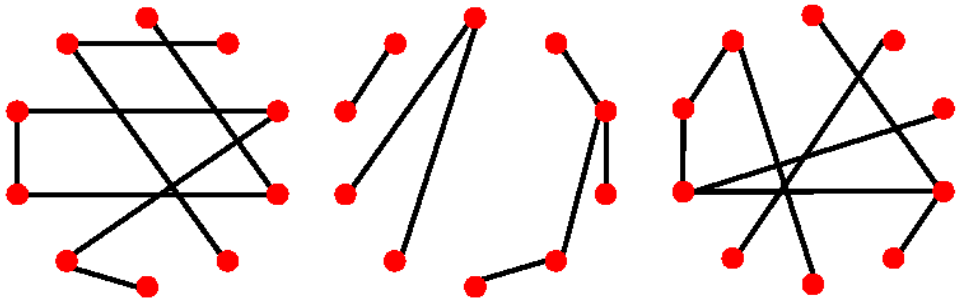
- **Erdös-Renyi Random Graphs** [Erdös-Renyi, ‘60]
- **Two variants:**
  - $G_{n,p}$: undirected graph on n nodes and each edge (u,v) appears i.i.d. with probability p
  - $G_{n,m}$: undirected graph with n nodes, and m uniformly at random picked edges

What kinds of networks does such model produce?

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Random Graph Model

- ***n*** **and** ***p*** **do not uniquely determine the graph!**
  - The graph is a result of a random process
- We can have many different realizations given the same ***n*** and ***p***



n = 10
p= 1/6

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Random Graph Model: Edges

- **How likely is a graph on E edges?**
- ***P(E)***: the probability that a given G_{np} generates a graph on exactly E edges:

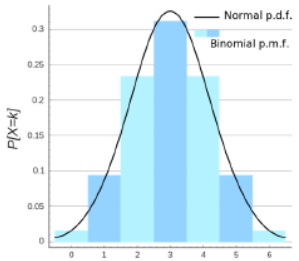
$$P(E) = \binom{E^{\max}}{E} p^{E} (1-p)^{E_{\max} - E}$$

where $E_{max}=n(n-1)/2$ is the maximum possible number of edges in an undirected graph of n nodes

P(E) is exactly the

Binomial distribution >>>

Number of successes in a sequence of **n** independent yes/no experiments



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Node Degrees in a Random Graph

- **What is expected degree of a node?**
  - Let X_v be a rnd. var. measuring the degree of node v
  - **We want to know:** $E[X_v] = \sum_{j=0}^{n-1} j\, P(X_v = j)$
    - **For the calculation we will need: Linearity of expectation**
      - For any random variables $Y_1, Y_2, \ldots, Y_k$
      - If $Y = Y_1 + Y_2 + \ldots Y_k$, then $E[Y] = \sum_i E[Y_i]$
- **An easier way:**
  - Decompose X_v to $X_v = X_{v,1} + X_{v,2} + \ldots + X_{v,n-1}$
    - where $X_{v,u}$ is a $\{0,1\}$-random variable which tells if edge (v,u) exists or not

$$E[X_v] = \sum_{u=1}^{n-1} E[X_{vu}] = (n-1)p$$

How to think about this?
- Prob. of node u linking to node v is p
- u can link (flips a coin) to all other $(n-1)$ nodes
- Thus, the expected degree of node u is: $p(n-1)$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Properties of G_{np}

Degree distribution: $P(k)$

Path length: h

Clustering coefficient: C

What are values of these properties for G_{np}?

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Degree Distribution

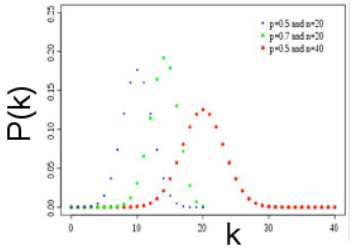
- **Fact: Degree distribution of G_{np} is Binomial.**
- Let ***P(k)*** denote a fraction of nodes with degree ***k***:

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Select *k* nodes out of *n-1*

Probability of having *k* edges

Probability of missing the rest of the *n-1-k* edges



Mean, variance of a binomial distribution

$$\bar{k} = p(n-1)$$

$$\sigma^2 = p(1-p)(n-1)$$

$$\frac{\sigma}{\bar{k}} = \left[\frac{1-p}{p}\frac{1}{(n-1)}\right]^{1/2} \approx \frac{1}{(n-1)^{1/2}}$$

As the network size increases, the distribution becomes increasingly narrow—we are increasingly confident that the degree of a node is in the vicinity of $\bar{k}$.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Clustering Coefficient of G_{np}

- **Remember:** $C_i = \frac{2e_i}{k_i(k_i-1)}$

  Where e_i is the number of edges between i's neighbors

- Edges in G_{np} appear i.i.d with prob. p
- **So:** $e_i = p\frac{k_i(k_i-1)}{2}$

  Each pair is connected with prob. p

  Number of distinct pairs of neighbors of node i of degree k_i

- **Then:** $C = \frac{p \cdot k_i(k_i-1)}{k_i(k_i-1)} = p = \frac{\bar{k}}{N}$

Clustering coefficient of a random graph is small.
For a fixed avg. degree, C decreases with the graph size N.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Network Properties of G_{np}

Degree distribution: $P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$

Clustering coefficient: $C = p = \bar{k}/n$

Path length: *next!*

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Network Properties of G_{np}

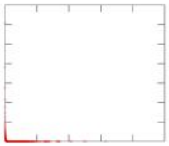
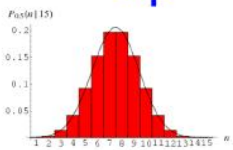
Degree distribution: $P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$

Path length: $O(\log n)$

Clustering coefficient: $C = p = \bar{k} / n$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

MSN vs. G_{np}

	MSN	G_{np}
Degree distribution:		
Path length:	6.6	$O(\log n)$ $h \approx 8.2$
Clustering coefficient:	0.11	$\bar{k}/n$ $C \approx 8 \cdot 10^{-8}$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Real Networks vs. G_{np}

- **Are real networks like random graphs?**
  - Giant connected component: ☺
  - Average path length: ☺
  - Clustering Coefficient: ☹
  - Degree Distribution: ☹
- **Problems with the random network model:**
  - Degreed distribution differs from that of real networks
  - Giant component in most real network does NOT emerge through a phase transition
  - No local structure – clustering coefficient is too low
- **Most important: Are real networks random?**
  - The answer is simply: **NO!**

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Real Networks vs. G_{np}

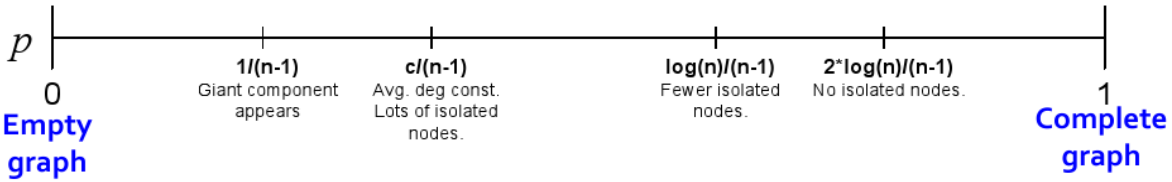
- **If G_{np} is wrong, why did we spend time on it?**
  - It is the reference model for the rest of the class.
  - It will help us calculate many quantities, that can then be compared to the real data
  - It will help us understand to what degree is a particular property the result of some random process

So, while G_{np} is WRONG, it will turn out to be extremly USEFUL!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

“Evolution” of a Random Graph

- **Graph structure of G_{np} as p changes:**



- **Emergence of a Giant Component:**

  avg. degree $k=2E/n$ or $p=k/(n-1)$

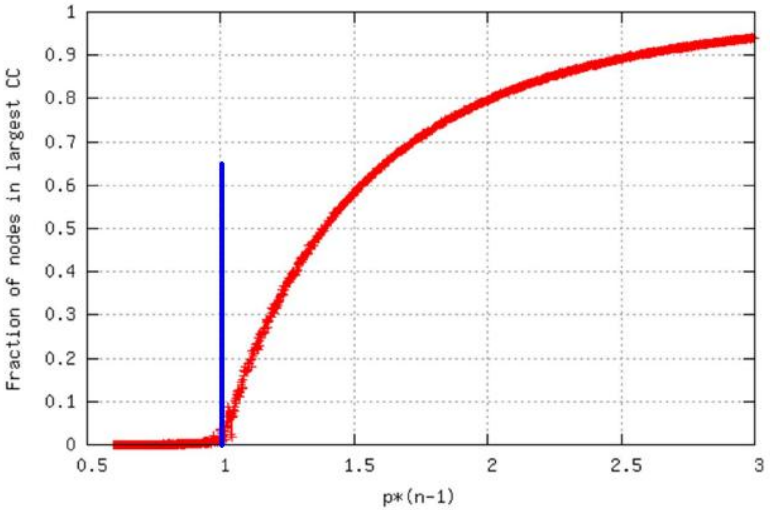
  - $k=1-\varepsilon$: all components are of size $\Omega(\log n)$
  - $k=1+\varepsilon$: 1 component of size $\Omega(n)$, others have size $\Omega(\log n)$

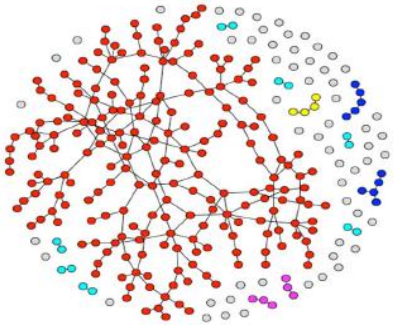
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

G_{np} Simulation Experiment





Fraction of nodes in the largest component

- G_{np}, n=100k, $p(n-1)$ = 0.5 … 3

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

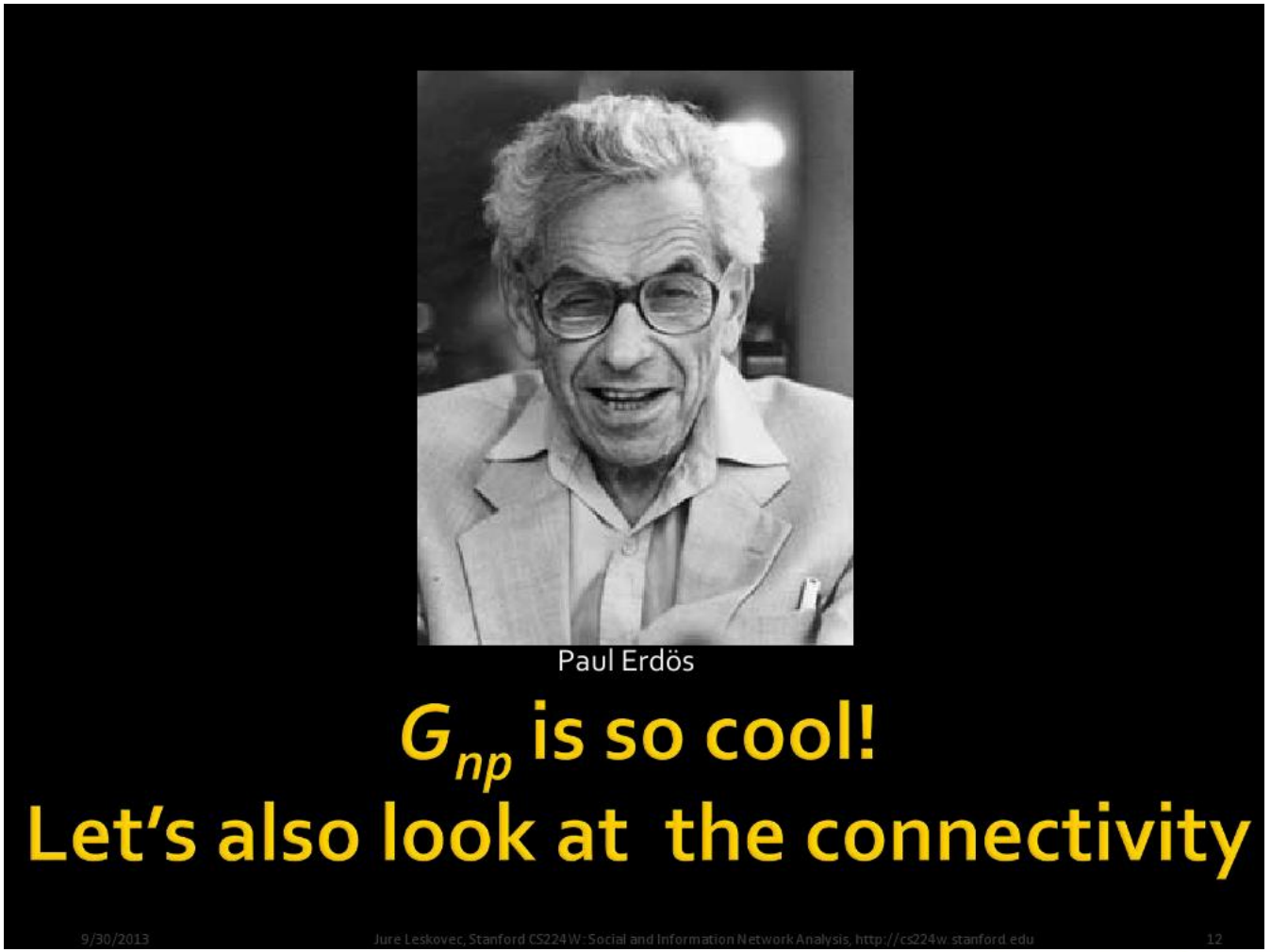
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Diameter of G_{np} and the Small-World Phenomena

CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

http://cs224w.stanford.edu

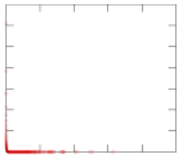
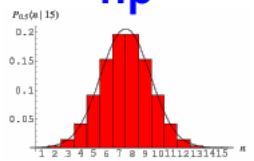


Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Back to MSN vs. G_{np}

	MSN	G_{np}
Degree distribution:		
Path length:	6.6	$O(\log n)$ $h \approx 8.2$
Clustering coefficient:	0.11	$\bar{k}/n$ $C \approx 8 \cdot 10^{-8}$
Connected component:	99%	$\bar{k} \approx 14$ $\log_{10}(180M) \approx 8$ So, GCC should kind of be there.

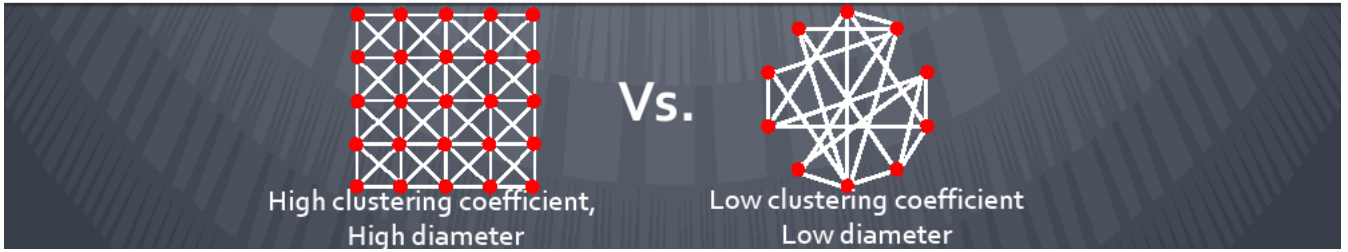
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

The Small-World Model

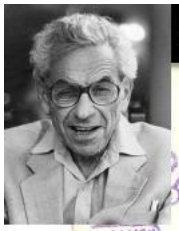
Can we have high clustering while also having short paths?



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

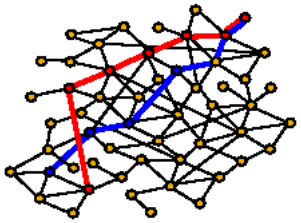


Erdös numbers are small!

Figure 1
To appear in Topics in Graph Theory (F. Harary, ed.) New York Academy of Sciences (1979).

The Small-World Experiment

- **What is the typical shortest path length between any two people?**
  - **Experiment on the global friendship network**
    - Can’t measure, need to probe explicitly
- **Small-world experiment** [Milgram ’67]
  - Picked 300 people in Omaha, Nebraska and Wichita, Kansas
  - Ask them to get a letter to a stock-broker in Boston by passing it through friends
- **How many steps did it take?**

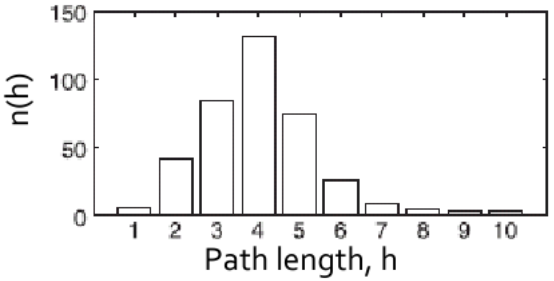


Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Dodds-Muhamad-Watts, '03]

Columbia Small-World Study

- **In 2003 Dodds, Muhamad and Watts performed the experiment using e-mail:**
  - 18 targets of various backgrounds
  - 24,000 first steps (~1,500 per target)
  - 65% dropout per step
  - 384 chains completed (1.5%)



Avg. chain length = 4.01

Problem: People stop participating

Correction factor: $n^*(h) = \frac{n(h)}{\prod_{i=0}^{h-1}(1-r_i)}$

r_i .... drop-out rate at hop i

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

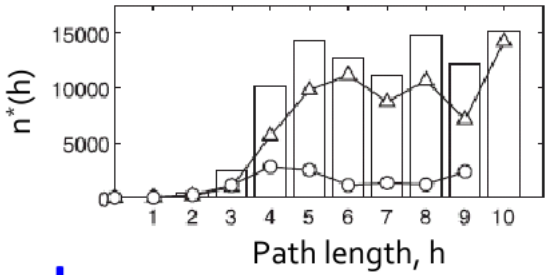
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Dodds-Muhamad-Watts, '03]

Small-World in Email Study

- **After the correction:**
  - **Typical path length h = 7**



- **Some not well understood phenomena in social networks:**
  - **Funneling effect:** Some target's friends are more likely to be the final step
    - Conjecture: High reputation/authority
  - **Effects of target's characteristics:** Structurally why are high-status target easier to find
    - Conjecture: Core-periphery network structure

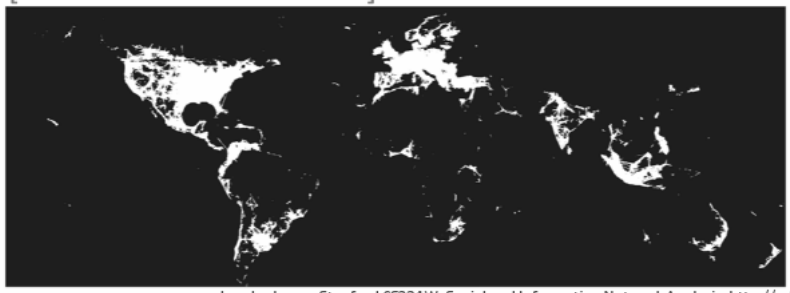
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

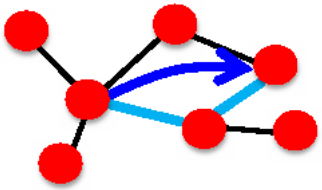
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

6-Degrees: Should We Be Surprised?

- **Assume each human is connected to 100 other people**
  Then:
  - Step 1: reach 100 people
  - Step 2: reach 100*100 = 10,000 people
  - Step 3: reach 100*100*100 = 1,000,000 people
  - Step 4: reach 100*100*100*100 = 100M people
  - **In 5 steps we can reach 10 billion people**
- **What's wrong here?**
  - **92% of new FB friendships are to a friend-of-a-friend**
    [Backstom-Leskovec '11]

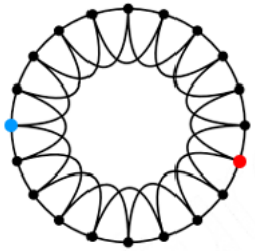




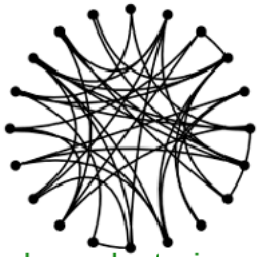
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Small-World: How?

- **Could a network with high clustering be at the same time a small world?**
  - How can we at the same time have **high clustering** and **small diameter?**



High clustering
High diameter



Low clustering
Low diameter

  - Clustering implies edge “locality”
  - Randomness enables “shortcuts”

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

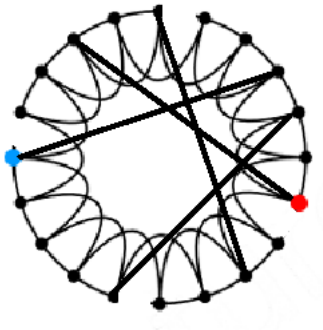
[Watts-Strogatz, '98]

Solution: The Small-World Model

Small-world Model [Watts-Strogatz '98]

2 components to the model:

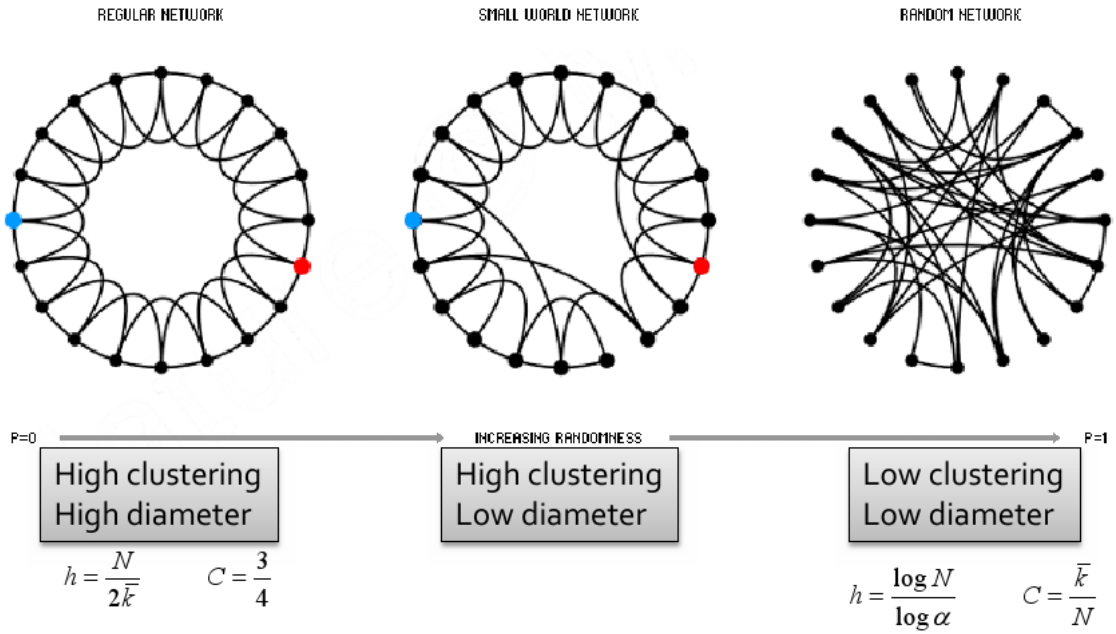
- **(1)** Start with a **low-dimensional regular lattice**
  - Has high clustering coefficient
- Now introduce randomness ("shortucts")
- **(2) Rewire:**
  - Add/remove edges to create shortcuts to join remote parts of the lattice
  - For each edge with prob. p move the other end to a random node



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

The Small-World Model

[Watts-Strogatz, '98]



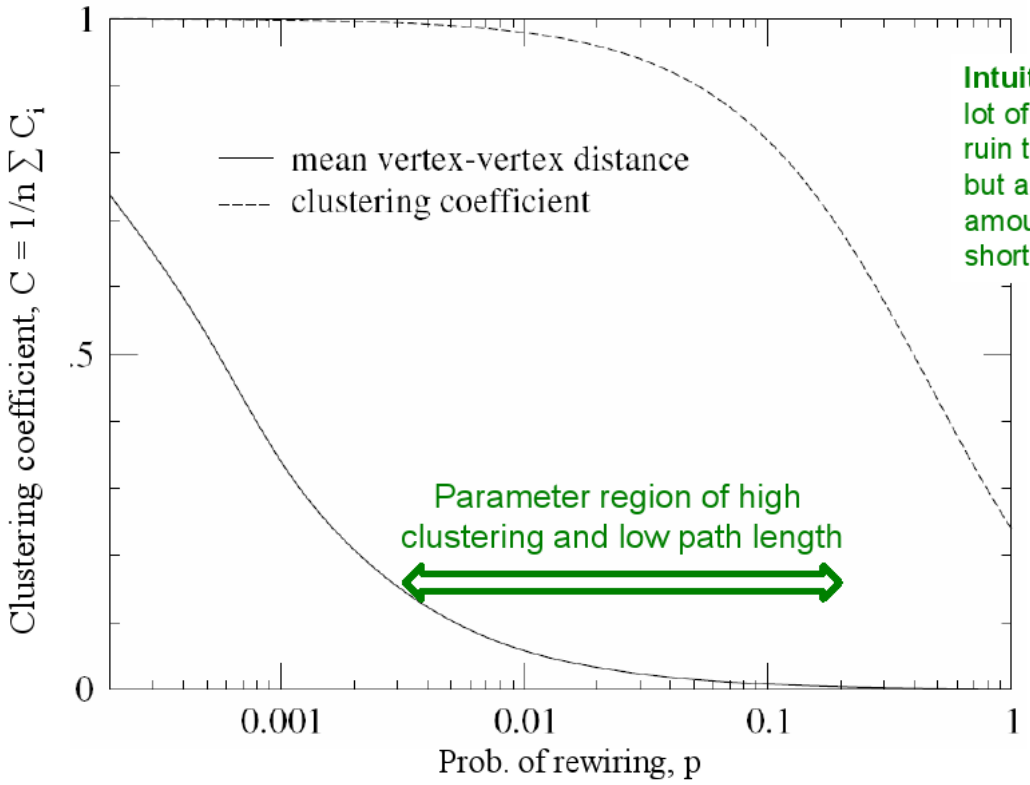
Rewiring allows us to “interpolate” between a regular lattice and a random graph

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

The Small-World Model



Intuition: It takes a lot of randomness to ruin the clustering, but a very small amount to create shortcuts.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Small-World: Summary

- **Could a network with high clustering be at the same time a small world?**
  - Yes! You don’t need more than a few random links
- **The Watts Strogatz Model:**
  - Provides insight on the interplay between clustering and the small-world
  - Captures the structure of many realistic networks
  - Accounts for the high clustering of real networks
  - Does not lead to the correct degree distribution
  - Does not enable **navigation** (next lecture)

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Network Formation Processes: Power-law degrees and Preferential Attachment

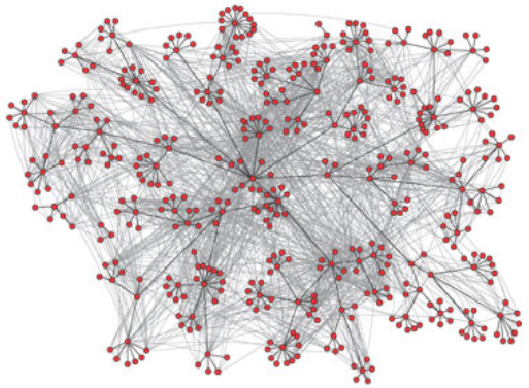
CS224W: Social and Information Network Analysis

Jure Leskovec, Stanford University

http://cs224w.stanford.edu

Network Formation Processes

What do we observe that needs explaining

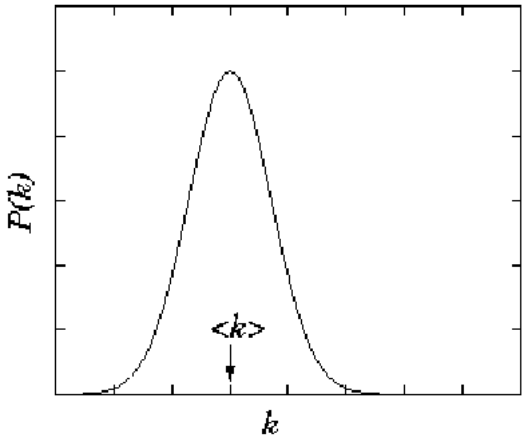


- **Small-world model?**
  - Diameter
  - Clustering coefficient
- **Preferential Attachment:**
  - **Node degree distribution**
    - What fraction of nodes has degree k (as a function of k)?
    - Prediction from simple random graph models: $p(k)$ = **exponential function of** k
    - **Observation: Power-law:** $p(k) = k^{-\alpha}$

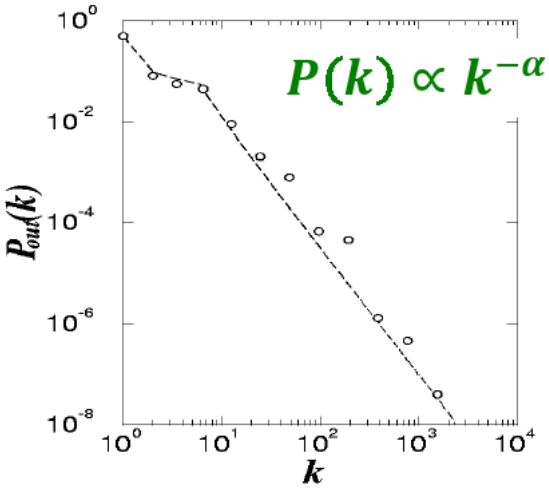
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Degree Distributions

Expected based on G_{np}



Found in data

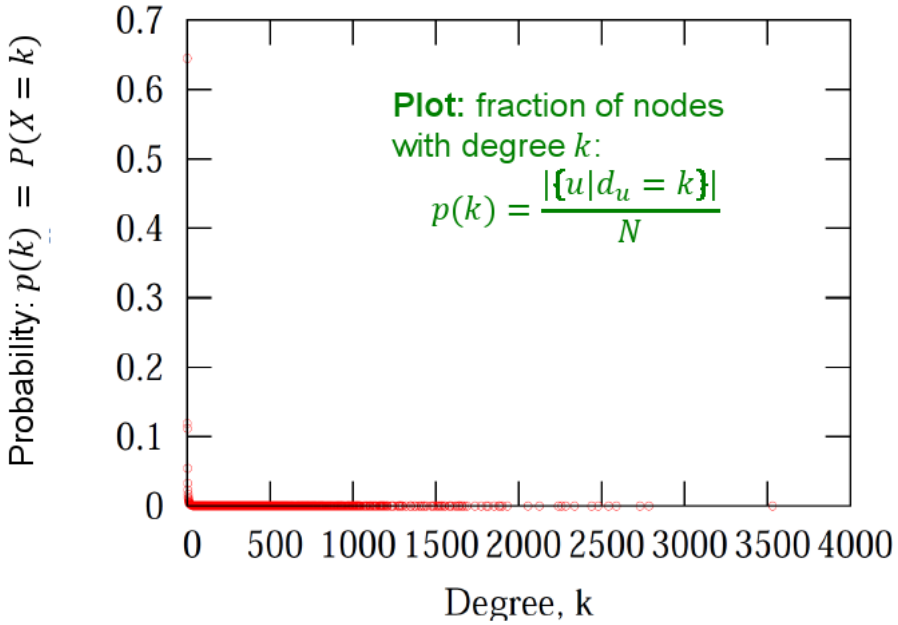


Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Leskovec et al. KDD '08]

Node Degrees in Networks

- **Take a network, plot a histogram of $P(k)$ vs. k**



Flickr social network
n= 584,207,
m=3,555,115

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

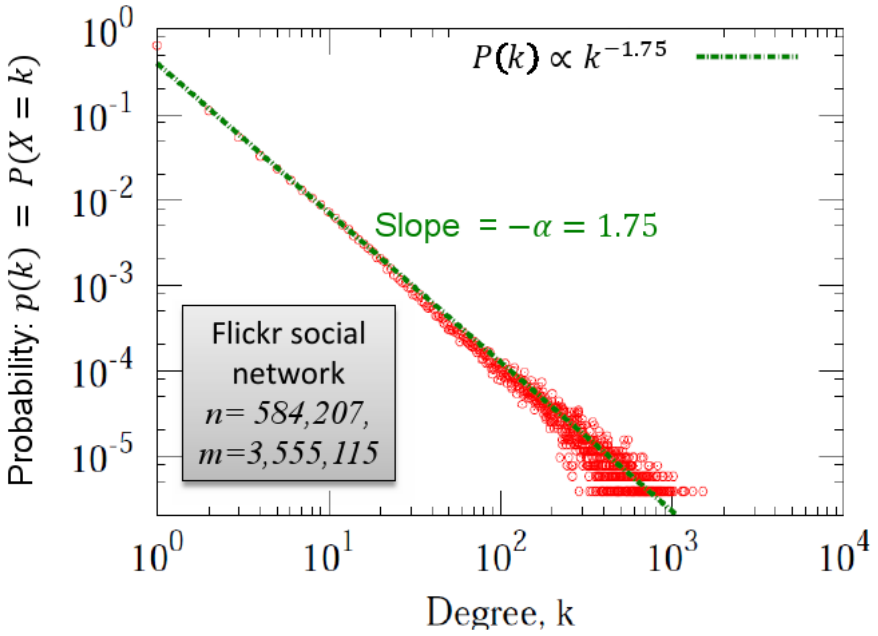
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Leskovec et al. KDD '08]

Node Degrees in Networks

- **Plot the same data on *log-log* scale:**



How to distinguish:
$P(k) \propto \exp(-k)$ **vs.**
$P(k) \propto k^{-\alpha}$ **?**

Take logarithms:
if $y = f(x) = e^{-x}$ then
$\log(y) = -x$
If $y = x^{-\alpha}$ then
$\log(y) = -\alpha \log(x)$
So, on log-log axis power-law looks like a straight line of slope $-\alpha$!

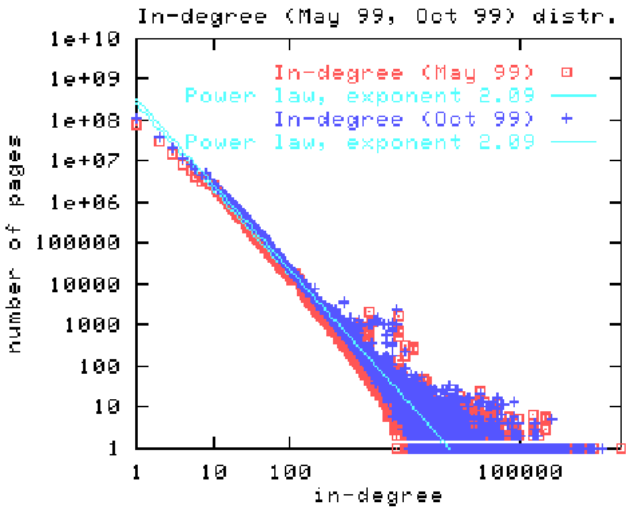
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

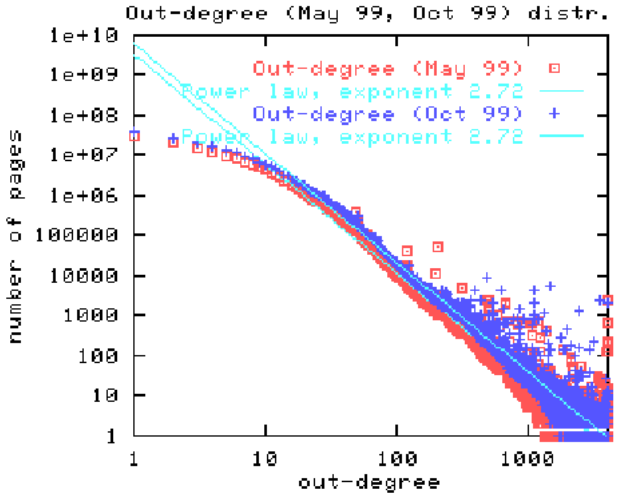
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Node Degrees: Web

- **The World Wide Web** [Broder et al., 2000]





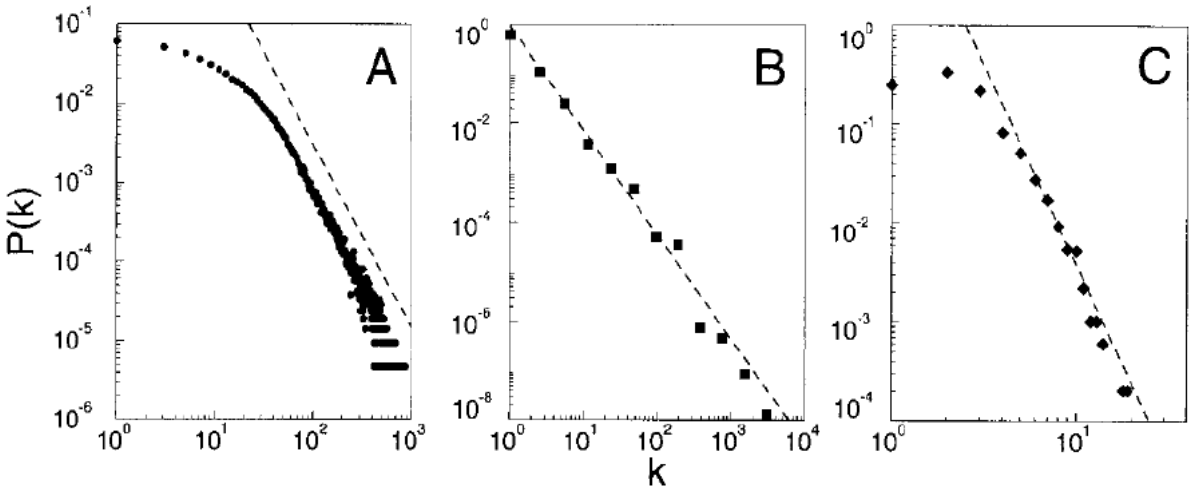
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Node Degrees: Barabasi&Albert

- **Other Networks** [Barabasi-Albert, 1999]



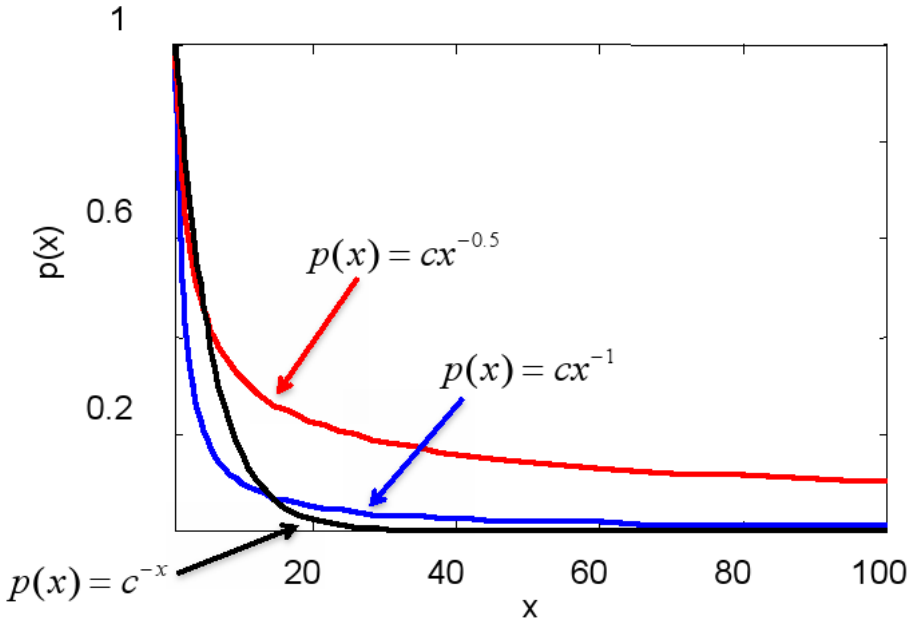
Actor collaborations | Web graph | Power-grid

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Exponential vs. Power-Law



- Above a certain x value, the power law is always higher than the exponential!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

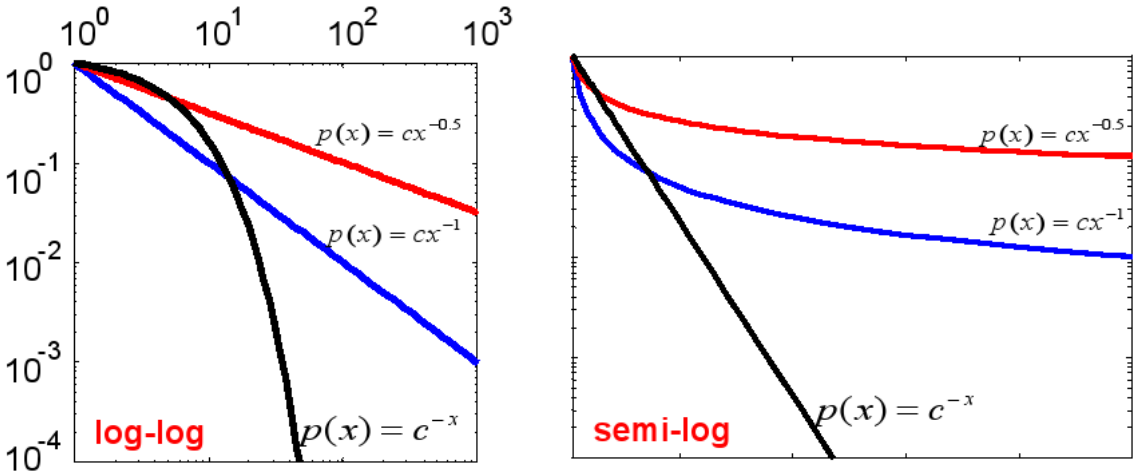
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Clauset-Shalizi-Newman 2007]

Exponential vs. Power-Law

- **Power-law vs. Exponential on log-log and log-lin scales**



x … logarithmic axis
y … logarithmic axis

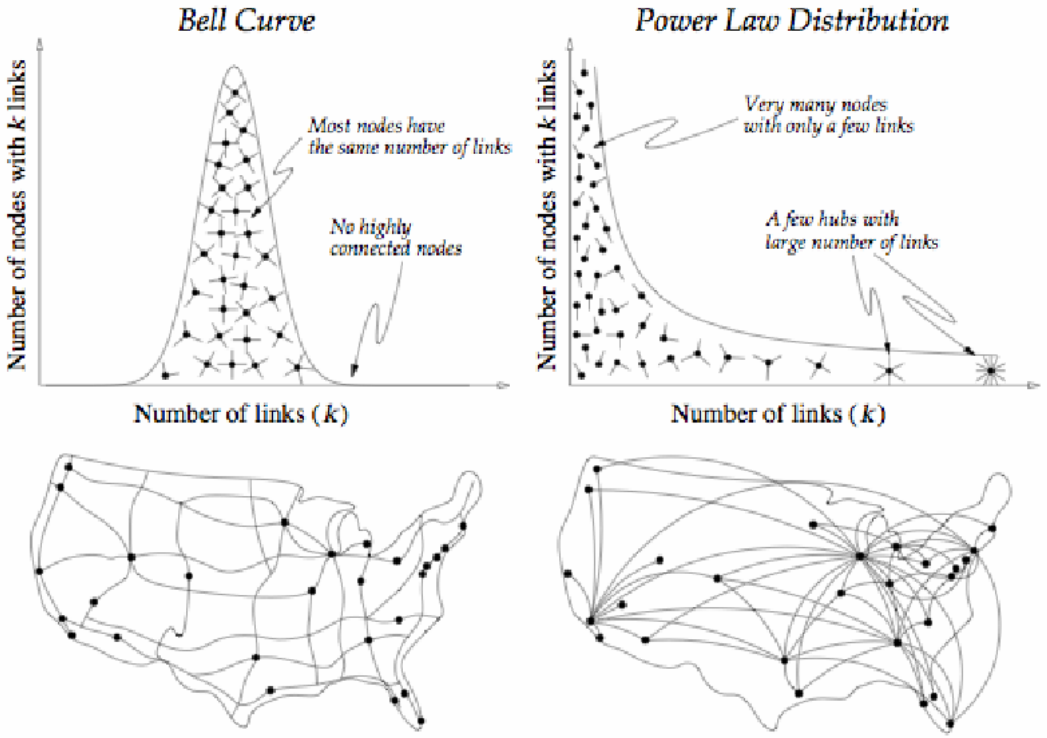
x … linear
y … logarithmic

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Exponential vs. Power-Law



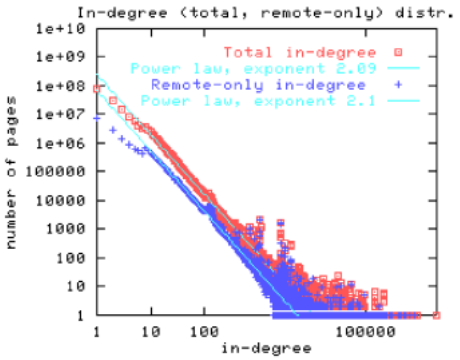
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

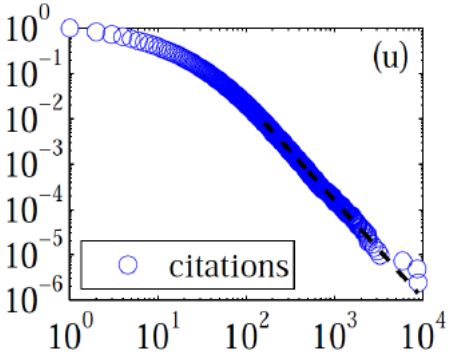
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Power-Law Degree Exponents

- **Power-law degree exponent is typically $2 < \alpha < 3$**
  - **Web graph:**
    - $\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$ [Broder et al. 00]
  - **Autonomous systems:**
    - $\alpha = 2.4$ [Faloutsos^3, 99]
  - **Actor-collaborations:**
    - $\alpha = 2.3$ [Barabasi-Albert 00]
  - **Citations to papers:**
    - $\alpha \approx 3$ [Redner 98]
  - **Online social networks:**
    - $\alpha \approx 2$ [Leskovec et al. 07]





Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Scale-Free Networks

- **Definition:**
  Networks with a power law tail in their degree distribution are called “scale-free networks”

- **Where does the name come from?**
  - **Scale invariance**: There is no characteristic scale
  - **Scale-free function:** $f(ax) = a^{\lambda} f(x)$
    - Power-law function: $f(ax) = a^{\lambda} x^{\lambda} = a^{\lambda} f(x)$

Log() or Exp() are not scale free!

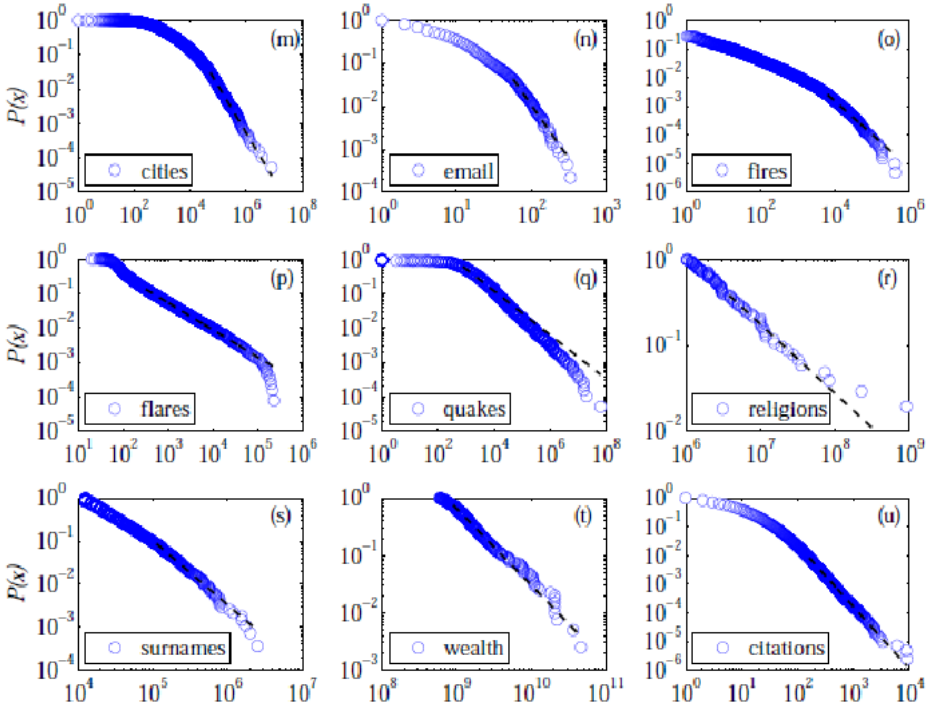
$f(ax) = \log(ax) = \log(a) + \log(x) = \log(a) + f(x)$

$f(ax) = \exp(ax) = \exp(x)^a = f(x)^a$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Clauset-Shalizi-Newman 2007]

Power-Laws are Everywhere

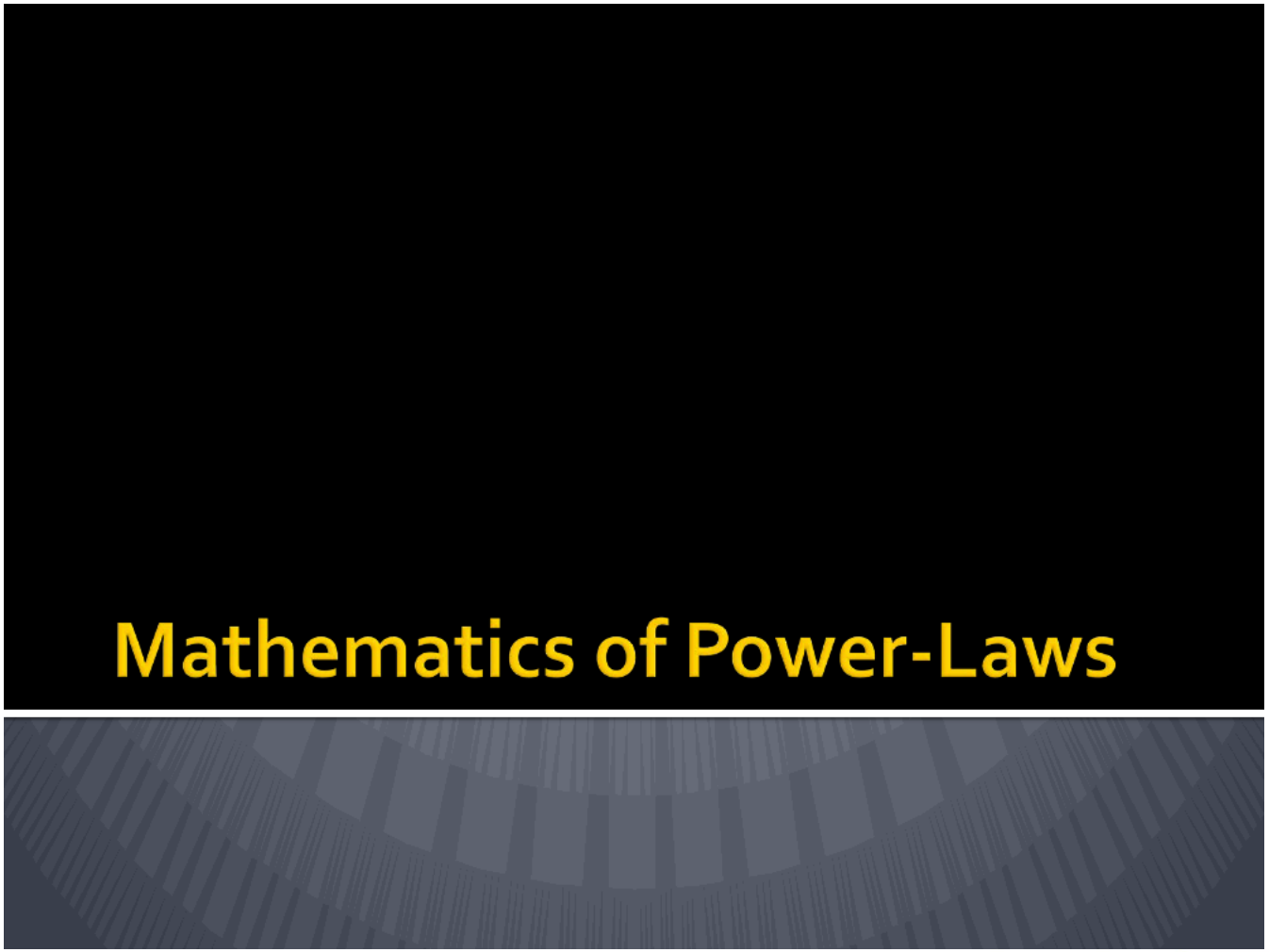


Many other quantities follow heavy-tailed distributions

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

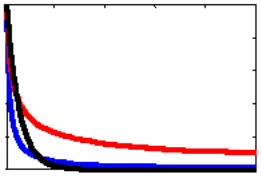
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Heavy Tailed Distributions

- **Degrees are heavily skewed:**

  Distribution $P(X > x)$ is **heavy tailed if:**

$$\lim_{x \to \infty} \frac{P(X > x)}{e^{-\lambda x}} = \infty$$



- **Note:**
  - **Normal PDF:** $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
  - **Exponential PDF:** $p(x) = \lambda e^{-\lambda x}$
    - **then** $P(X > x) = 1 - P(X \leq x) = e^{-\lambda x}$

  are not heavy tailed!

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Clauset-Shalizi-Newman 2007]

Heavy Tailed Distributions

- **Various names, kinds and forms:**
  - Long tail, Heavy tail, Zipf's law, Pareto's law
- **Heavy tailed distributions:**
  - **P(x) is proportional to:**

power law	$P(x) \propto x^{-\alpha}$
power law with cutoff	$x^{-\alpha} e^{-\lambda x}$
stretched exponential	$x^{\beta-1} e^{-\lambda x^{\beta}}$
log-normal	$\frac{1}{x} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right]$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

[Clauset-Shalizi-Newman 2007]

Mathematics of Power-laws

- **What's the expectation of a power-law random variable *x*?**
- $E[x] = \int_{x_m}^{\infty} x\, p(x) dx = z \int_{x_m}^{\infty} x^{-\alpha+1} dx$
- $= -\frac{z}{2-\alpha}\left[x^{2-\alpha}\right]_{x_m}^{\infty} = -\frac{(\alpha-1)x_m^{\alpha-1}}{2-\alpha}\left[\infty^{2-\alpha} - x_m^{2-\alpha}\right]$

Need: $\alpha > 2$!

$$\Rightarrow \boldsymbol{E[x] = \frac{\alpha-1}{\alpha-2} x_m}$$

Power-law density:

$$p(x) = \frac{\alpha-1}{x_m}\left(\frac{x}{x_m}\right)^{-\alpha}$$

$$Z = \frac{\alpha-1}{x_m^{1-\alpha}}$$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Mathematics of Power-Laws

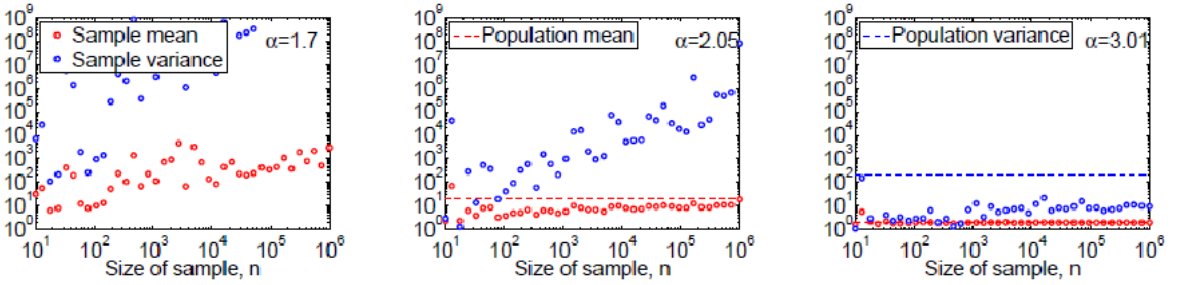
- **Power-laws have infinite moments!**

$$E[x] = \frac{\alpha - 1}{\alpha - 2} x_m$$

  - If $\alpha \leq 2 : E[x] = \infty$
  - If $\alpha \leq 3 : Var[x] = \infty$
    - **Average is meaningless, as the variance is too high!**

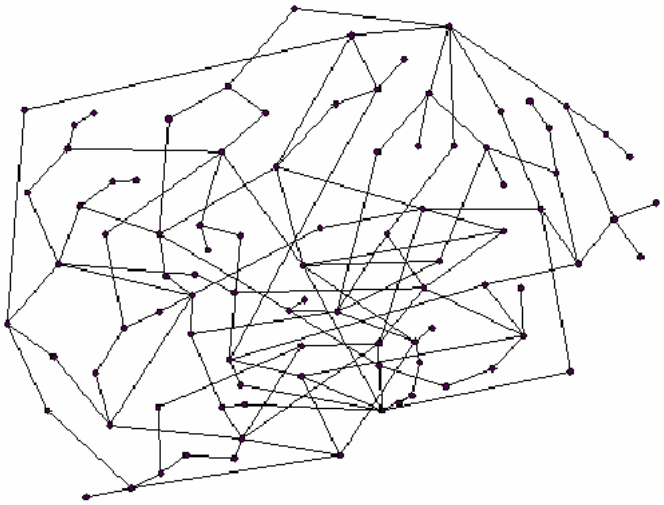
In real networks
2 < α < 3 so:
E[*x*] = const
Var[*x*] = ∞

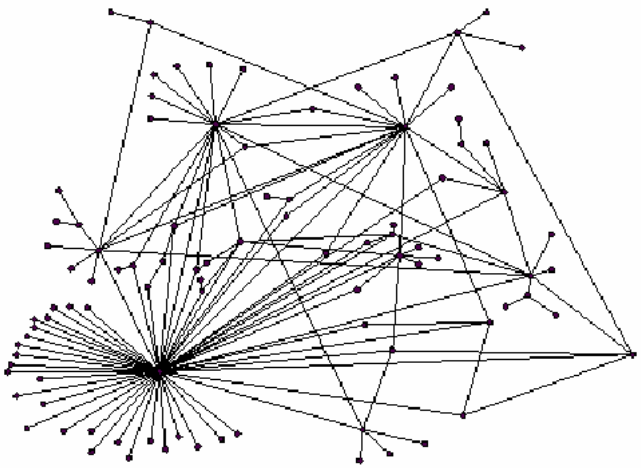
- **Sample average of *n* samples from a power-law with exponent α:**



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

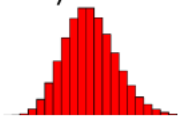
Random vs. Scale-free network





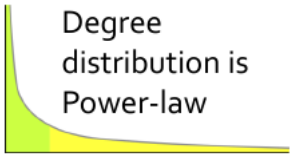
Random network

(Erdos-Renyi random graph)



Degree distribution is Binomial

Scale-free (power-law) network



Degree distribution is Power-law

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Model: Preferential Attachment

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

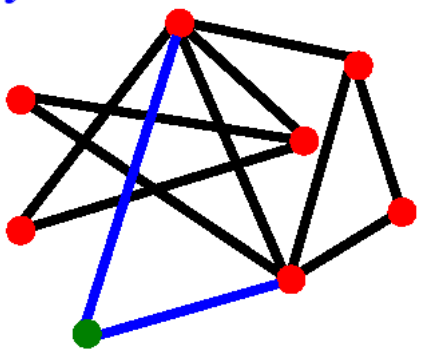
MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Model: Preferential attachment

- **Preferential attachment**
  [Price ‘65, Albert-Barabasi ’99, Mitzenmacher ‘03]
  - Nodes arrive in order **1,2,…,n**
  - At step j, let d_i be the degree of node $i < j$
  - A new node j arrives and creates m out-links
  - Prob. of j linking to a previous node i is **proportional to degree d_i of node i**

$$P(j \to i) = \frac{d_i}{\sum_k d_k}$$



Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

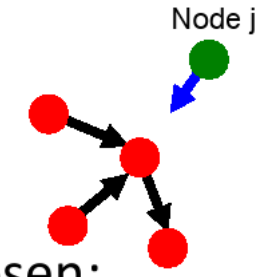
Rich Get Richer

- **New nodes are more likely to link to nodes that already have high degree**
- **Herbert Simon's result:**
  - Power-laws arise from "**Rich get richer**" (**cumulative advantage**)
- **Examples** [Price 65]:
  - **Citations:** New citations to a paper are proportional to the number it already has

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

The Exact Model

[Mitzenmacher, '03]



We will analyze the following model:

- Nodes arrive in order $1,2,3,\ldots,n$
- When node ***j*** is created it makes a **single out-link** to an earlier node $\boldsymbol{i}$ chosen:
  - **1)** With prob. $\boldsymbol{p}$, $\boldsymbol{j}$ links to $\boldsymbol{i}$ chosen **uniformly at random** (from among all earlier nodes)
  - **2)** With prob. $\boldsymbol{1-p}$, node $\boldsymbol{j}$ chooses node $\boldsymbol{i}$ uniformly at random and links **to a node *i* points to**
    - **This is same as saying:** With prob. $\boldsymbol{1-p}$, node $\boldsymbol{j}$ links to node $\boldsymbol{i}$ with prob. proportional to $\boldsymbol{d_i}$ (the in-degree of $\boldsymbol{i}$)
  - **Our graph is directed:** Every node has out-degree **1**

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

The Model Givens Power-Laws

- **Claim:** The described model generates networks where the fraction of nodes with in-degree k scales as:

$$P(d_i = k) \propto k^{-(1+\frac{1}{q})}$$

where q=1-p

So we get power-law degree distribution with exponent:

$$\alpha = 1 + \frac{1}{1-p}$$

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Preferential attachment: Good news

- **Preferential attachment gives power-law degrees**
- Intuitively reasonable process
- Can tune p to get the observed exponent
  - On the web, $P[\text{node has degree } d] \sim d^{-2.1}$
  - $2.1 = 1+1/(1-p) \rightarrow p \sim 0.1$

There are also other network formation mechanisms that generate scale-free networks:

- Random surfer model [Blum-Mugizi]
- Copying model [Kleinberg et al.]
- Forest Fire model [Leskovec et al.]

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Many models lead to Power-Laws

- **Copying mechanism** (directed network)
  - Select a node and an edge of this node
  - Attach to the endpoint of this edge
- **Walking on a network** (directed network)
  - The new node connects to a node, then to every
  - first, second, … neighbor of this node
- **Attaching to edges**
  - Select an edge and attach to both endpoints of this edge
- **Node duplication**
  - Duplicate a node with all its edges
  - Randomly prune edges of new node

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638
MAGYARORSZÁG MEGÚJUL
A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Distances in Preferential Attachment

Extra!

$$\bar{h} = \begin{cases} const & \alpha = 2 \\ \dfrac{\log\log n}{\log(\alpha - 1)} & 2 < \alpha < 3 \\ \dfrac{\log n}{\log\log n} & \alpha = 3 \\ \log n & \alpha > 3 \end{cases}$$

Ultra small world

Small world

Avg. path length

Degree exponent

Size of the biggest hub is of order O(N). Most nodes can be connected within two steps, thus the average path length will be independent of the network size.

The average path length increases slower than logarithmically. In G_{np} all nodes have comparable degree, thus most paths will have comparable length. In a scale-free network vast majority of the path go through the few high degree hubs, reducing the distances between nodes.

Some models produce $\alpha = 3$. This was first derived by Bollobas et al. for the network diameter in the context of a dynamical model, but it holds for the average path length as well.

The second moment of the distribution is finite, thus in many ways the network behaves as a random network. Hence the average path length follows the result that we derived for the random network model earlier.

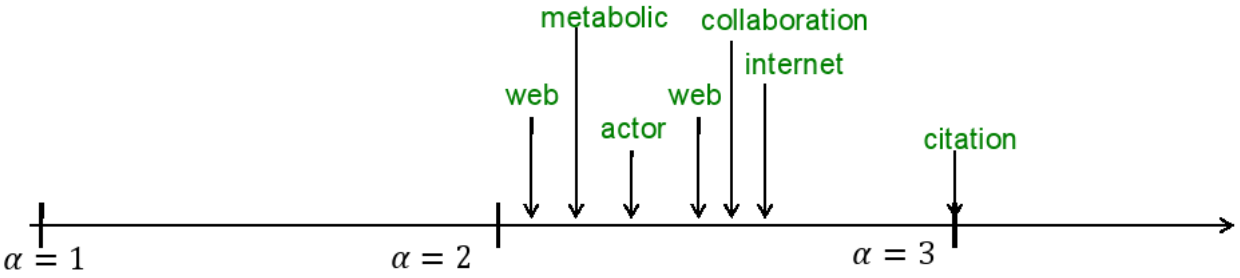
Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Summary: Scale-Free Networks

Extra!



Second moment $\langle k^2 \rangle$ diverges	$\langle k^2 \rangle$ finite
Average $\langle k \rangle$ diverges	$\langle k \rangle$ finite
Ultra small world behavior	Small world
Regime full of anomalies…	The scale-free behavior is relevant
	Behaves like a random network

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Nemzeti Fejlesztési Ügynökség
www.ujszechenyiterv.gov.hu
06 40 638 638

MAGYARORSZÁG MEGÚJUL

A projekt az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.