HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Overlay and P2P Networks

On power law networks II

Prof. Sasu Tarkoma

31.1.2013

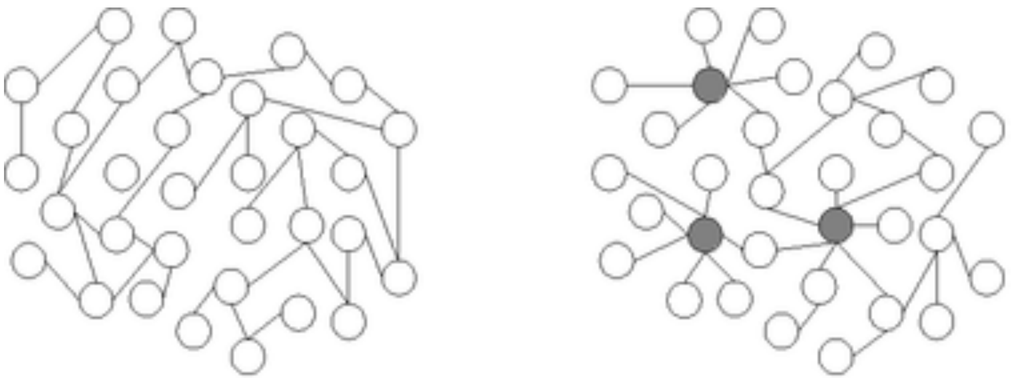
Contents

Scale free and small worlds summary

Search in small worlds

Freenet revisited

Scale-Free Networks



(a) Random network　　(b) Scale-free network

Small Worlds and Scale Free Networks

Small world phenomenon exlains why highly clustered
- graphs can have short average path lengths
- Natural and man-made structures

Watts and Strogatz 1998

It does not explain why this property emerges in real networks

How do power law networks emerge?
- Nodes connect to well-connected nodes, connectivities follows a power law

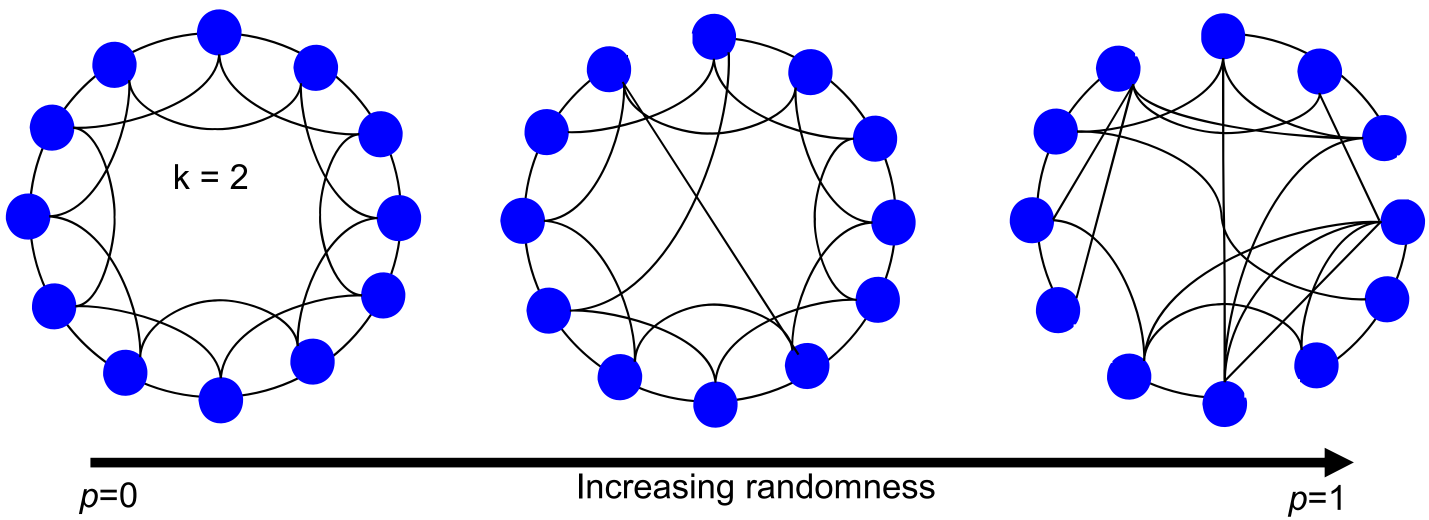
Watts-Strogatz model

D. Watts and S. Strogatz proposed a generative model to explain small world properties

Build a ring of n vertices and connect each vertex with its *k* clockwise neighbors on the ring

Draw a random number between 0 and 1 for each edge

Rewire each edge with probability *p*: if the edge's random number is smaller than p, keep the source vertex of the edge fixed, and choose a new target vertex uniformly at random from all other vertices



http://www.l3s.de/~balke/lecture-p2p/Vorlesung_7.pdf

The Copying Generative Model

Proposed by R. Kumar, P. Raghavan, et al. in 2000

In each time step randomly copy one of the existing nodes keeping its links

Connect the original node and the copy

Randomly remove edges from both nodes with a very small probability, and give removed edge random new target nodes

The probability of a node getting a new edge is proportional to its degree

More edges increases probability for a neighbour being chosen

Barabási-Albert Model

Scale-free networks with power-law node degree distribution
The network grows in time
No random edge generation

Higher the degree, higher the probability that the new vertex will attach (**preferential attachment**)

$$\Pi(v) = \frac{deg(v)}{\sum_{w \in V} deg(w)}$$

Generative model

1. Start with a small network (a number of nodes and edges at random)
2. At every step, add a new vertex x. Add m edges to x to existing vertices. The target is drawn with the probability given by preferential attachment.

Preferential Attachment

Consider dynamic Web graph

Pages join one at a time

Each page has one link out

Let $X_j(t)$ be the number of pages of degree j at time t.

New page links:

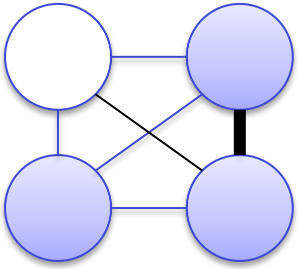
With probability α, link to a random page

With probability $(1- \alpha)$, a link to a page chosen proportionally to indegree

Local Clustering Coefficient

The clustering coefficient C(v) of vertex v in a directed graph is given by:

the number of **links between the vertices within its neighborhood** divided by the number of **links that could possibly exist between them**



C=1/3

Neighbourhood is the immediately connected neighbours
k(k-1) possible links for k vertices

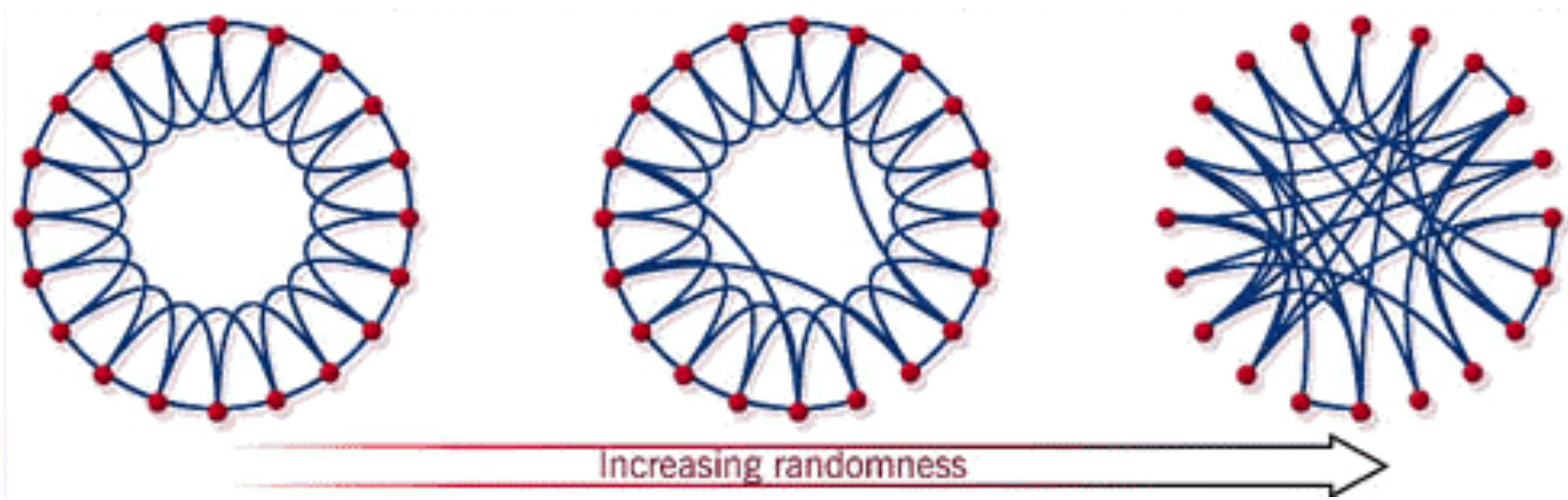
For undirected graph k(k-1)/2 possible links

Network average: $$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i.$$

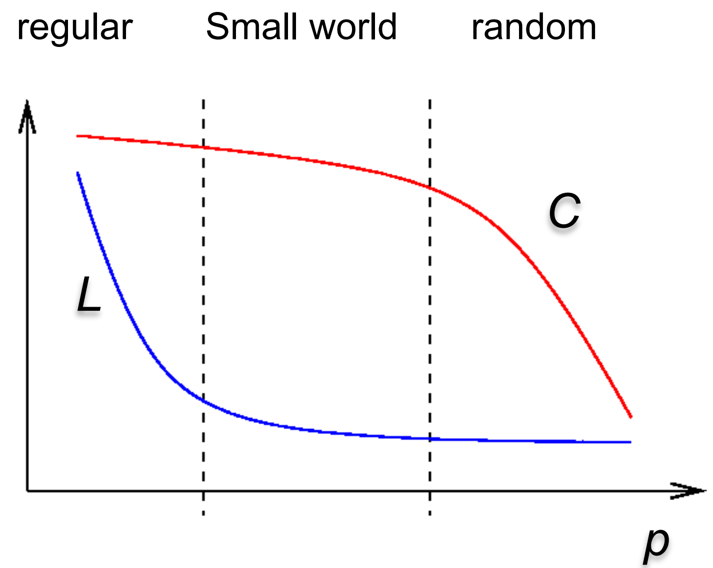
Structured network
- high *clustering*
- large diameter
- regular

Small-world network
- high *clustering*
- small diameter
- almost regular

Random network
- small *clustering*
- small diameter



Reference: Duncan J. Watts & Steven H. Strogatz, Nature 393, 440-442 (1998)



C(p) : clustering coefficient
L(p) : average path length

Reference: Duncan J. Watts & Steven H. Strogatz, Nature 393, 440-442 (1998)

Kleinberg's result

Jon Kleinberg showed that it is possible to do efficient routing on grids with the small world property.

The possibility of efficient routing depends on a balance between the proportion of shortcut edges of different lengths with respect to coordinates in the base grid.

The key idea is to use a frequency of edges of different lengths that decrease **inverse proportionally** to the length.

Kleinberg Small World

Set of points on a n x n grid

Each node only has local information

Routing table creation needs global information

Manhattan distance is used (sum of horizontal and vertical components in a grid)

Kleinberg's result II

Results in an infinite family of small world network models on a **grid** with **power-law distributed random long-range links**

K(n,k,p,q,r)
p – radius of neighbours to which short local links
q – number of random long range links
k - dimension of the mesh
r - clustering exponent of inverse power-law distribution.

$\text{Prob.}[(x,y)] \propto \text{dist}(x,y)^{-r}$

r determines how steeply the probability of links to far away neighbors reduces

Searching in the Small World

Simple greedy routing: use routing table to find the link that takes the message cosest to the target

Assumes that there is a way to associating nodes with points on the grid in order to find ”closest node”

Search processing

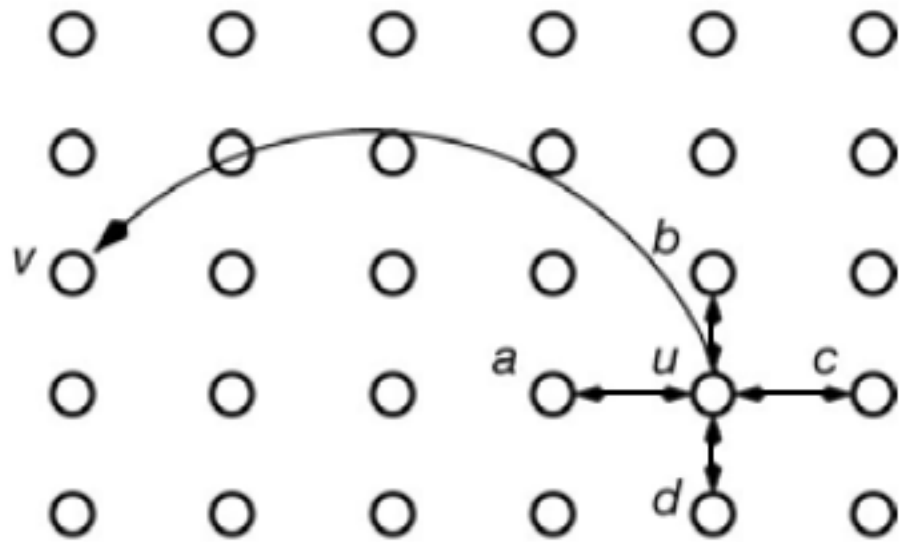
Node makes a greedy routing decision based on

- the coordinates of its local and long-range contacts
- the coordinates of the nodes that the message was previously routed through
- the coordinates of the target node.

- The message that needs to be delivered carries with it the target node coordinates and the coordinates of the nodes already visited nodes.

Example

Node u is connected to all its neighbors (a, b, c , and d) and has a long-range link to some randomly chosen node v with a probablility proportional to dist(u, v)$^{-r}$



Just using the neighbours gives O(n) for destination
If the clustering coefficient is zero, then the long range links are too random
If one then there are too few random links
Two would be the optimal value (links are uniformly distributed over all distances)
Results in logarithmic diameter for the network

Constructing the graph

Every node i is connected to node j within distance d

Connect nodes in higher distance with probability decreasing with growing distance

For every node i, additional q edges are added

Probability that node j is selected is proportional to $d(i,j)^{-r}$

The nodes are selected by generating q random numbers based on the distribution that are distances on the graph, and then choosing a node at that distance.

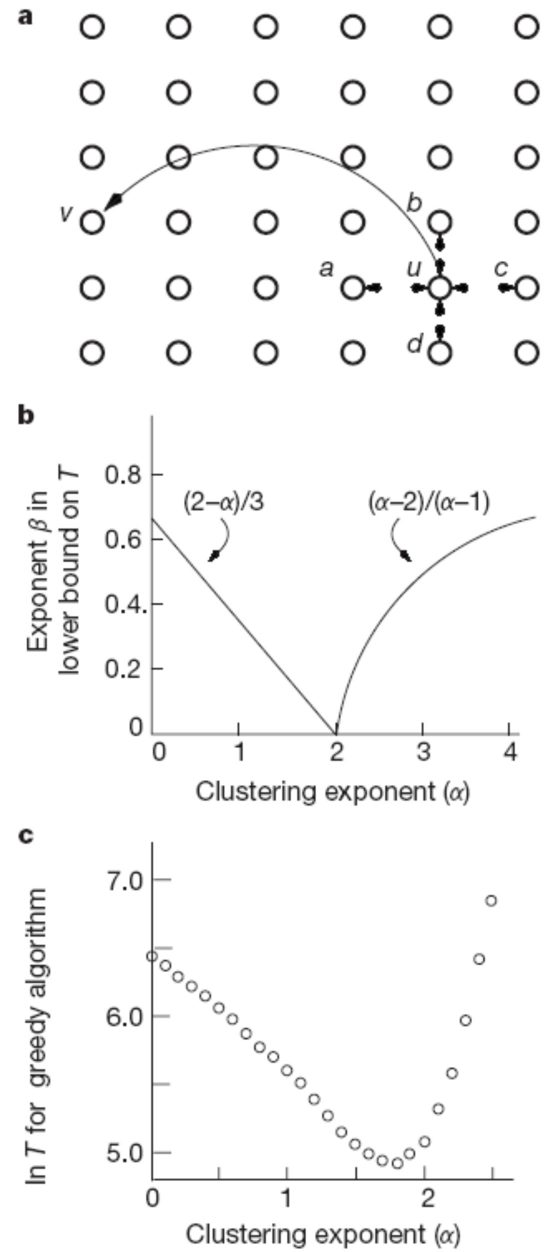
Delivery Time in Lattice Networks

For k=2, dip in time-to-search at r=2
- For low r, random graph; no “geographic” correlation in links
- For high r, not a small world; no short paths to be found.

Searcheability dips at r=2, in simulation
- Corresponds to using greedy heuristic of sending message to the node with the least lattice distance to goal

For d-dimensional lattice, minimum occurs at r=d

Expected Delivery time =
- **O((log n)²), for r = 2 (and the special case k=r).**
- $\Omega(n^{(2-r)/3})$, for $0 \le r < 2$.
- $\Omega(n^{(r-2)/(r-1)})$, for $2 < r$.



Theorem

Theorem: The routing algorithm will find short paths if and only if $k = r$. (k is the dimension, r is clustering exponent)

The idea behind the proof is that for any $r < k$ there are too few random edges to make the paths short.

For $r > k$ there are too many random edges, and thus too many choices to which the message could be sent. Message will make a long random walk through the network.

Kleinberg's result III

Simple greedy routing can find routes in **O(log²(n))** hops, where n is the size of the graph

Decentralized

Decisions based on local information

Later work has investigated other topologies than grids (rings, …) and improving efficiency through topology information, cues, etc.

Implication of result: greedy and local solution for building peer-to-peer overlay networks

Note: mathematical assumptions need to hold! If they do not, efficient decentralized search is not possible.

Freenet Routing Revisited

Every file is has a key (derived via a hash function)

A file is stored at some node with a similar key

At each peer each request is forwarded to the node in its routing table having the closest key to the requested one

If the request is successful, the file is sent back via the routing nodes and each node saves the file and adds the sending node's address to its local routing table (i.e., frequently requested files are replicated)

If the routing table is full, the random entry is evicted

Clustering and caching for achieving the small world network benefits in routing

Freenet

Original Freenet did not take small world property into account in routing, no gurantee of the existence of an efficient decentralized search

Freenet 0.7 featured the new clustering technique based on node locations

How to map to Kleinberg's model?

Freenet Mapping I

Kleinberg's model has a base lattice (and then the random network of long-range contacts)

Sandberg's Freenet routing algorithm assumes that a Freenet graph corresponds to the long-range contacts in some unknown k-dimensional Kleinberg network.

If we can find the underlying base lattice, we can then determine the metric for the discovering short paths

Sandberg has shown that finding the unknown base grid can be cast as a problem in statistical estimation in which the lattice coordinates of the actual Freenet nodes are the parameters to be estimated.

Freenet Mapping II

The aim of Freenet is to update node locations so that their assignment results in a small-world embedding in an imaginary base ring lattice

This is realized by having the nodes examine their location keys periodically for a possible location swap

Location swap does not alter the network topology, just the location identifiers!

But they may swap data items.
Location swap will affect future search requests.

Freenet Idea

Assume that the network exhibit small world properties.

Should be possible to recover an **embedded** Kleinberg small-world graph.

This is accomplished by selecting random pairs of nodes and potentially swapping them based on an objective function.

Function minimizes the product of all the distances between any given node and its neighbors.

Location swapping details

1. A node A randomly chooses a node B in its proximity and initiates a swap request. Both nodes share the locations of their respective neighbors and calculate $D_1(A,B)$. $D_1(A,B)$ is the product of the existing distances between A and each of A's neighbors $|L(a) - L(n)|$ multiplied by the product of the existing distances between B and each of B's neighbors.

$$D_1(A,B) = \prod_{(A,n)\in E} |L(A) - L(n)| \cdot \prod_{(B,n)\in E} |L(B) - L(n)| \quad (1)$$

2. The nodes also compute $D_2(A,B)$, the product of the products of the differences between their locations and their neighbors' locations *after* a potential swap:

$$D_2(A,B) = \prod_{(A,n)\in E} |L(B) - L(n)| \cdot \prod_{(B,n)\in E} |L(A) - L(n)| \quad (2)$$

3. If the nodes find that $D_2(A,B) \leq D_1(A,B)$, they swap locations, otherwise they swap locations with probability $\frac{D_1(A,B)}{D_2(A,B)}$. The deterministic swap always decreases the average distances of nodes with their neighbors. The probabilistic swap is used to escape local minima.

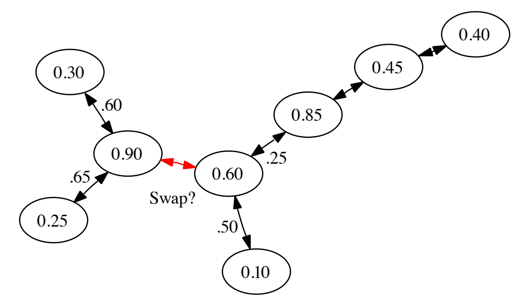


Figure 2. This figure shows an example network with two nodes considering a swap. The result of the swap equation is D_1 = .60 * .65 * .25 * .50 = .04875 and D_2 = .30 * .35 * .05 * .80 = .0042. Since $D_1 > D_2$, they swap.

N. Evans et al. Routing in the Dark: Pitch Black

Is Freenet a small world?

There must be a scale-free power-law distribution of links within the network.

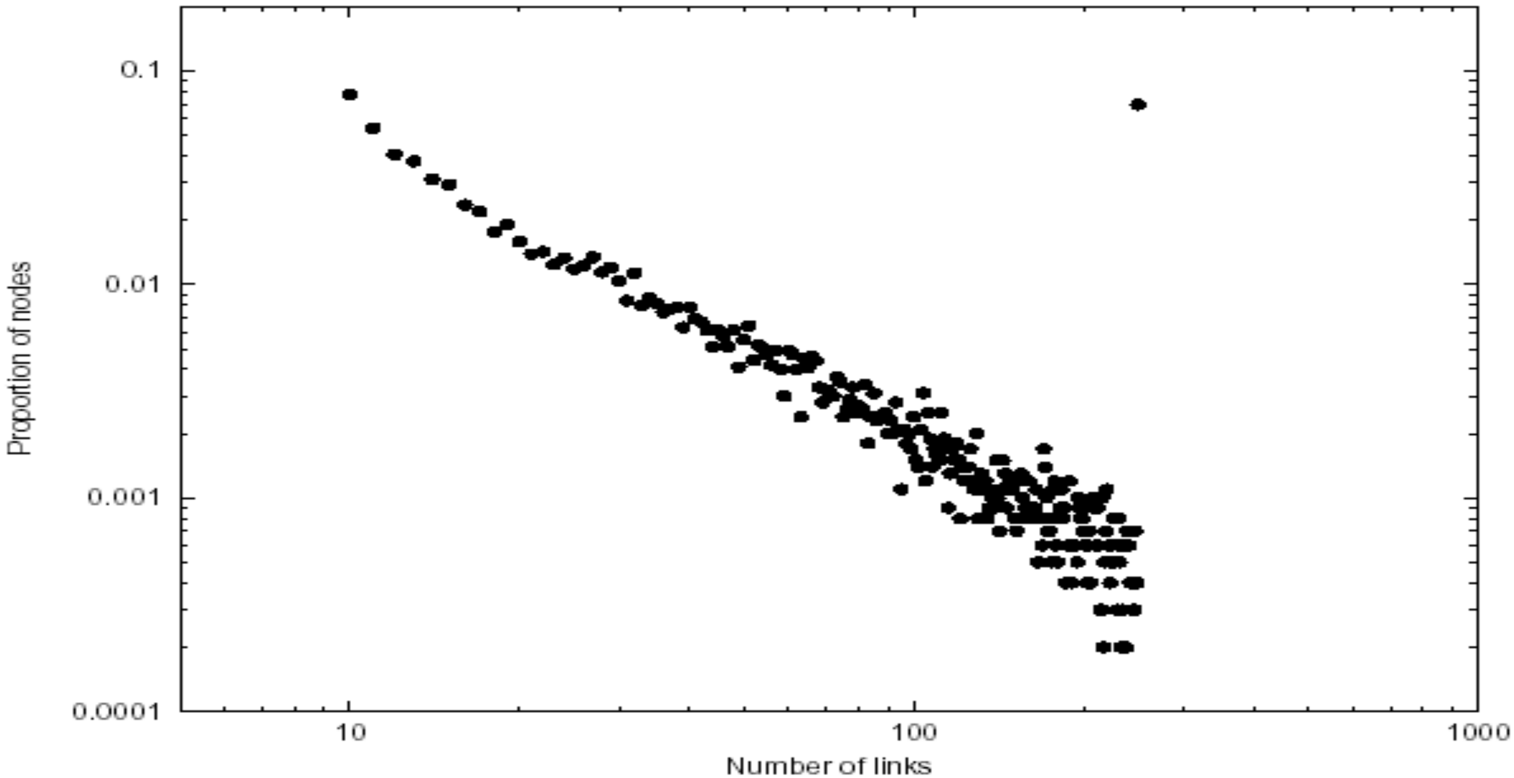


Fig. 5. Distribution of link number among Freenet nodes.

Source: www.ics.forth.gr/dcs/Activities/Projects/p2p/ploumid-freenet.ppt

Applications of Small World Networks

Many applications in peer-to-peer networks

The Gnutella network has been observed to exhibit the clustering and short path lengths of a small world network. Its overlay dynamics lead to a biased connectivity among peers where each peer is more likely connected to peers with higher uptime

The Freenet routing algorithm is built on the small world assumption

Other applications in distributed hashing (DHTs) such as Symphony that uses long-range contacts drawn randomly from a family of harmonic distributions

Stages of power law network research (M. Mitzenmacher, 2003)

There are 5 stages of power law network research.

1) Observe: Gather data to demonstrate power law behavior in a system.
2) Interpret: Explain the importance of this observation in the system context.
3) Model: Propose an underlying model for the observed behavior of the system.
4) Validate: Find data to validate (and if necessary specialize or modify) the model.
5) Control: Design ways to control and modify the underlying behavior of the system based on the model.

References

Barabási, Albert-László, Linked: The New Science of Networks, 2002. ISBN 0-452-28439-2

M. Mitzenmacher. A brief history of generative models for power law and lognormal distributions. Internet Mathematics, 2003. http://www.uvm.edu/~pdodds/research/papers/others/2003/mitzenmacher2003a.pdf

Watts, D. J. and S. H. Strogatz. Collective dynamics of 'small-world' networks. Nature 393:440-42, 1998.

D.J. Watts. Networks, Dynamics and Small-World Phenomenon, American Journal of Sociology, Vol. 105, Number 2, 493-527, 1999.

M. E. J. Newman, Random graphs as models of networks, in Handbook of Graphs and Networks, S. Bornholdt and H. G. Schuster (eds.), Wiley-VCH, Berlin (2003).

J. Kleinberg. The small-world phenomenon: An algorithmic perspective. Proc. 32nd ACM Symposium on Theory of Computing, 2000.