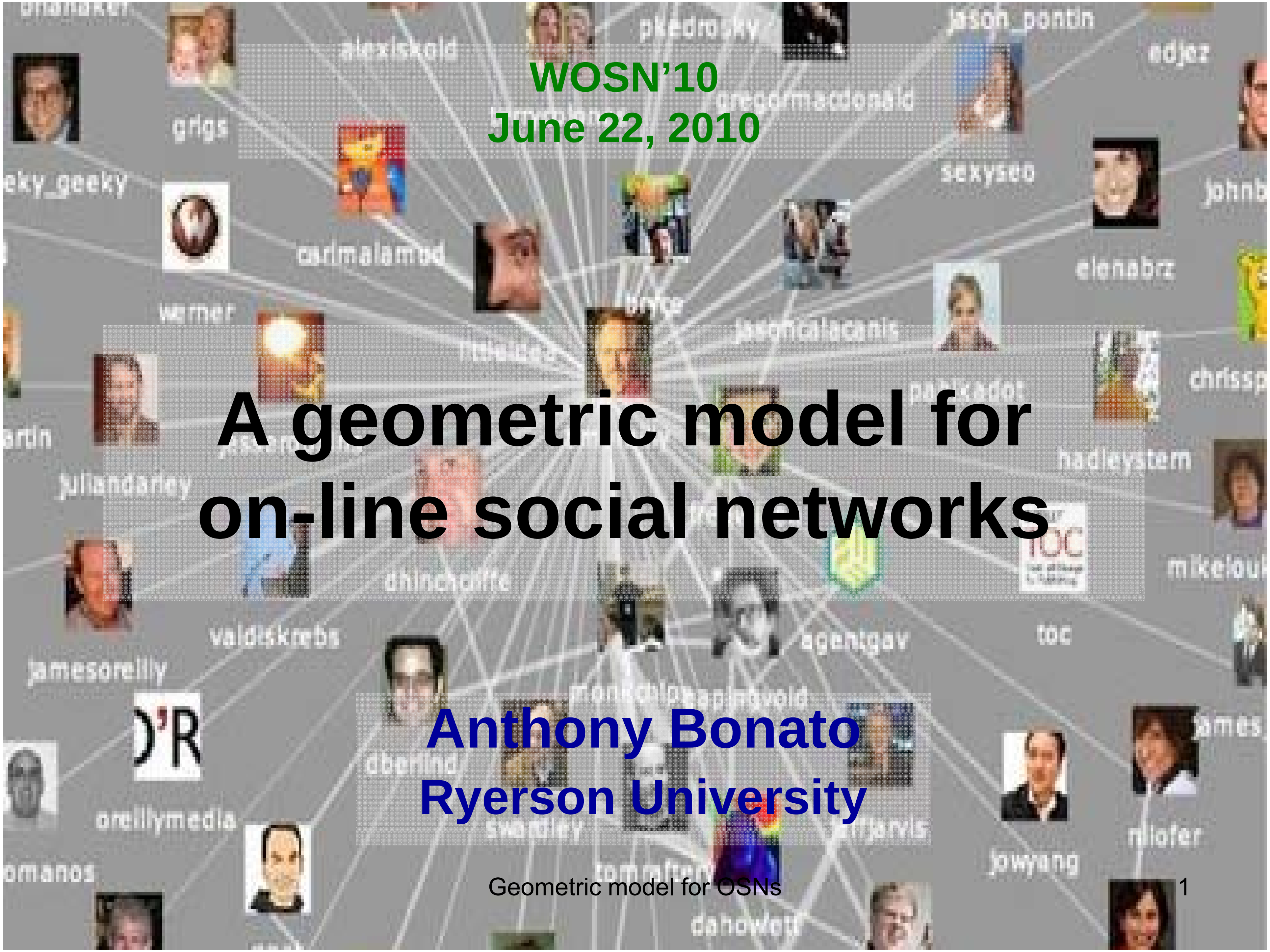


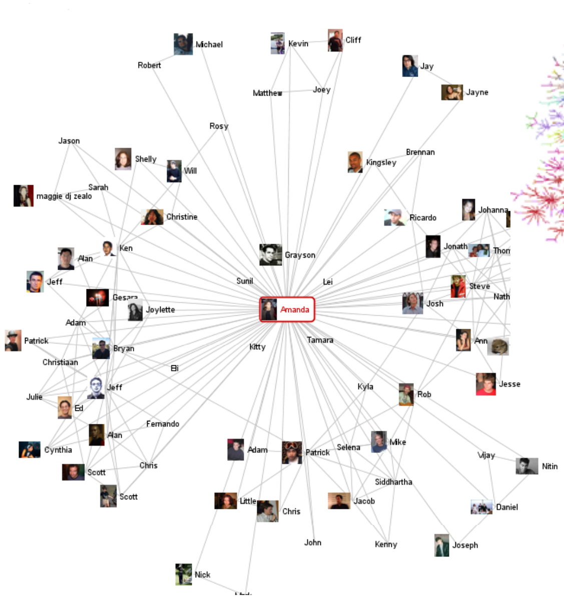
WOSN'10
June 22, 2010

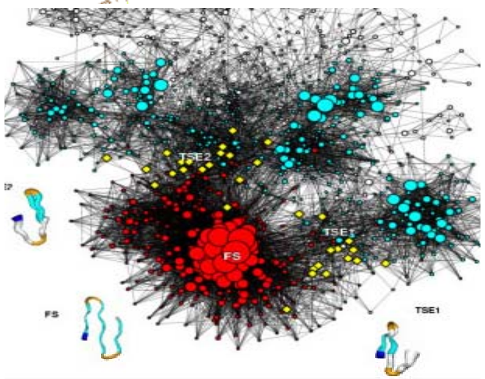
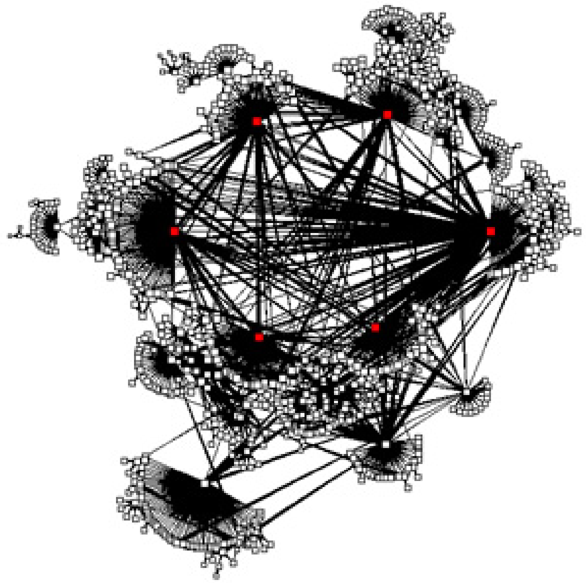
A geometric model for on-line social networks

Anthony Bonato
Ryerson University

Complex Networks

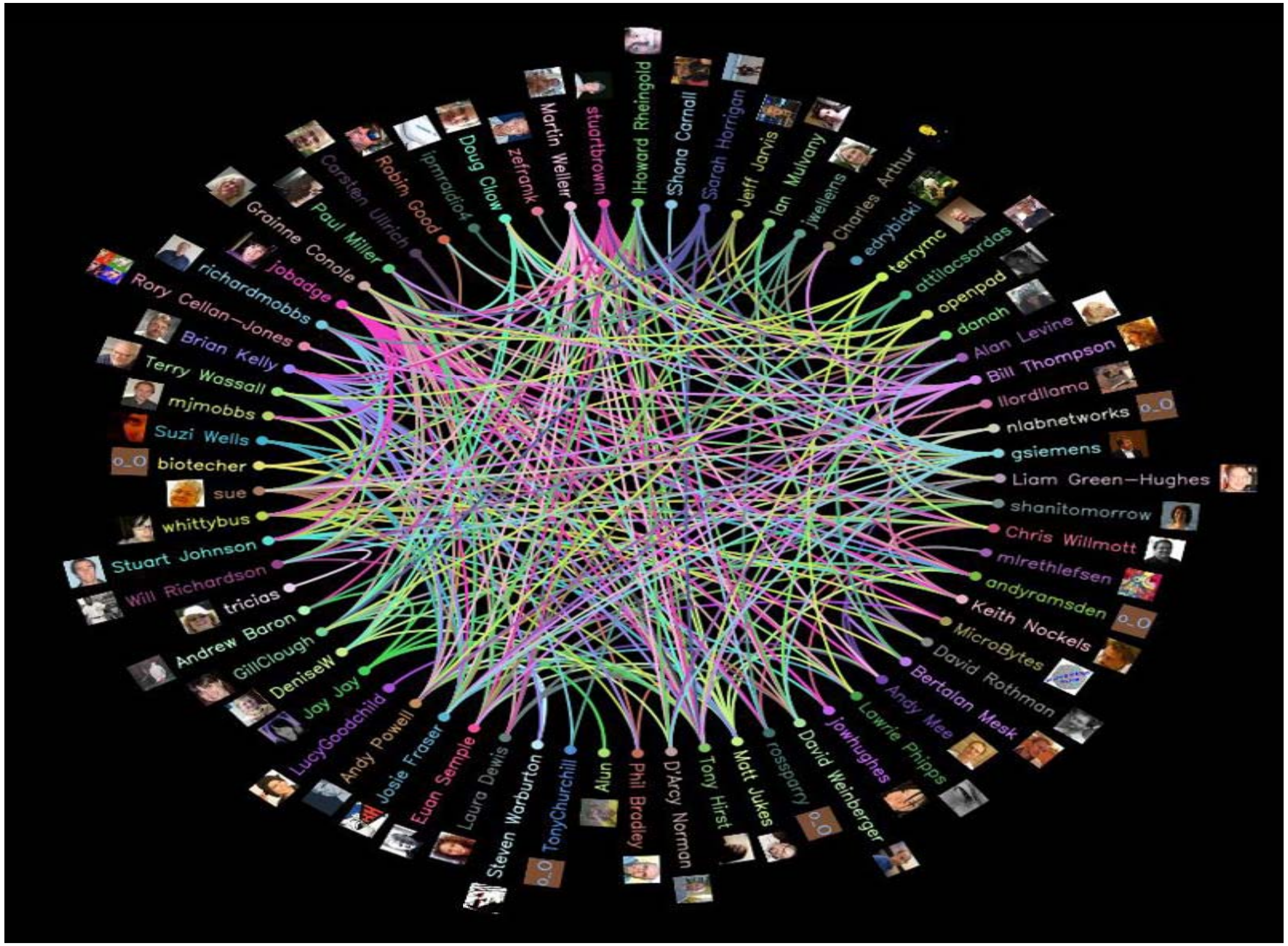
- web graph, social networks, biological networks, internet networks, …



On-line Social Networks (OSNs)

Facebook, Twitter, LinkedIn, MySpace…



Properties of OSNs

- observed properties:
  - power law degree distribution, small world
  - community structure
  - densification power law and shrinking distances

(Kumar et al,06):

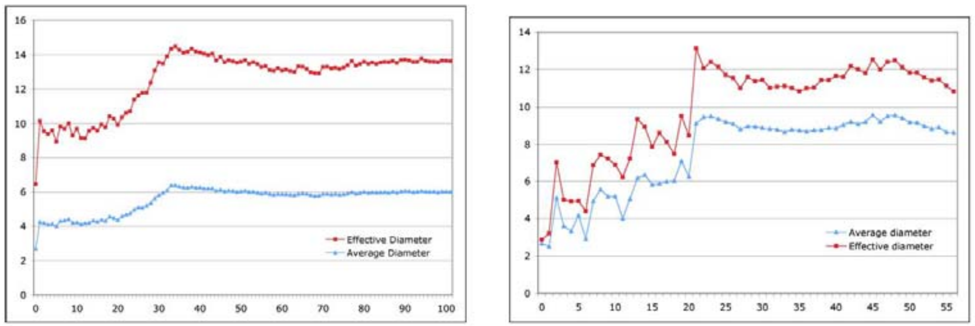
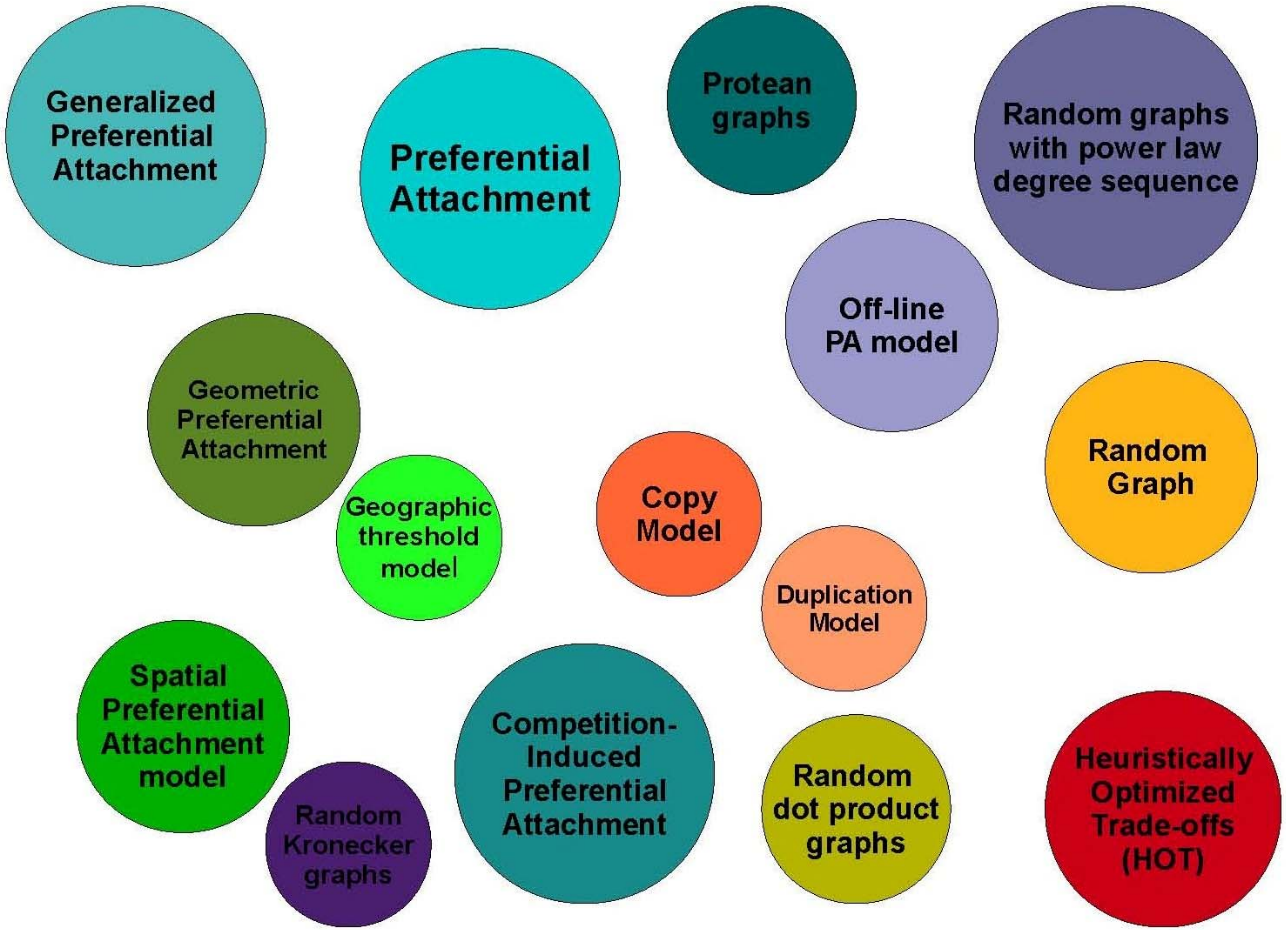


Figure 6: Average and effective diameter of the giant component of Flickr and Yahoo! 360 timegraphs, by week.

Why model complex networks?

- uncover and explain the generative mechanisms underlying complex networks
- predict the future
- nice mathematical challenges
- models can uncover the hidden reality of networks

Many different models



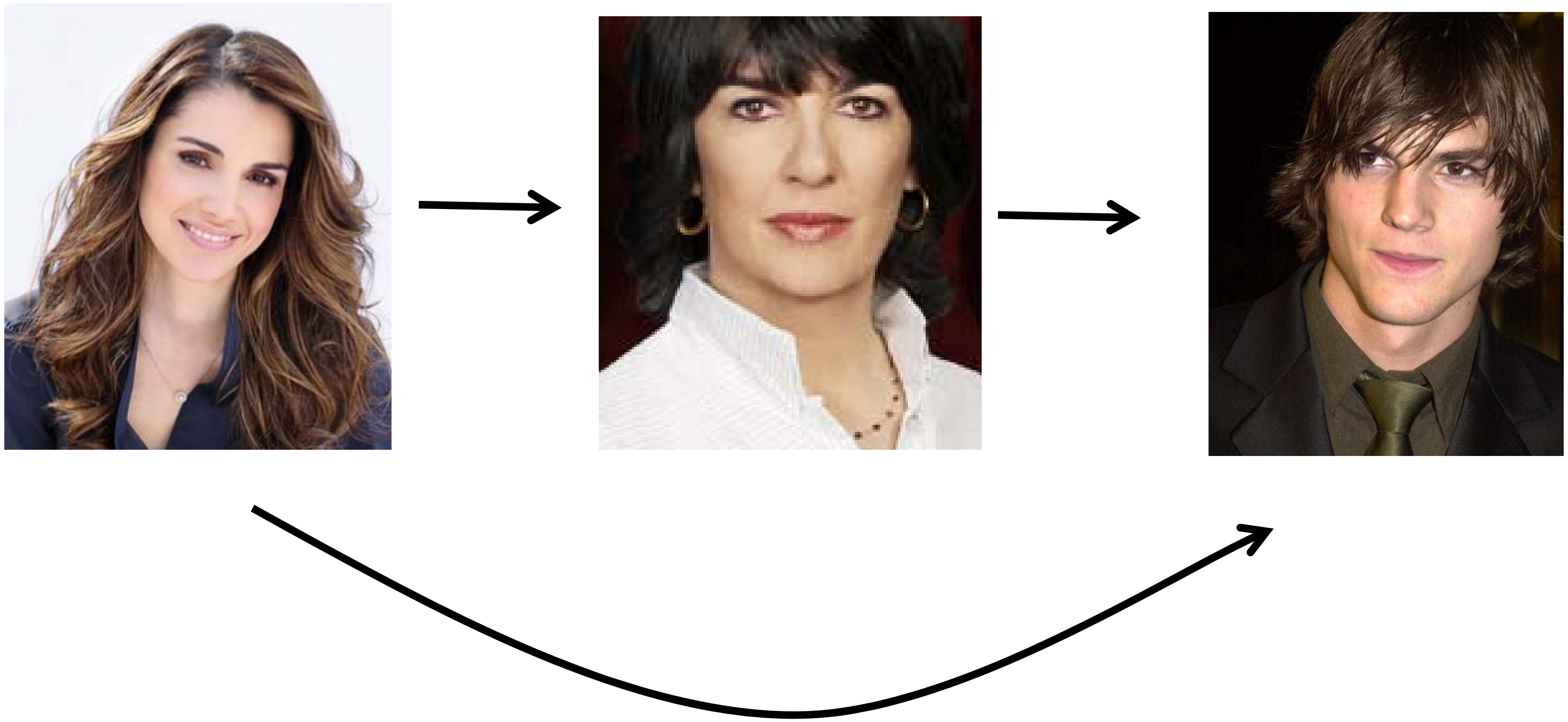
Models of OSNs

- relatively few models for on-line social networks

- goal: find a model which simulates many of the observed properties of OSNs
  - must evolve in a natural way…

"All models are wrong, but some are more useful."
– G.P.E. Box

Transitivity

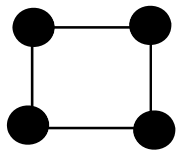


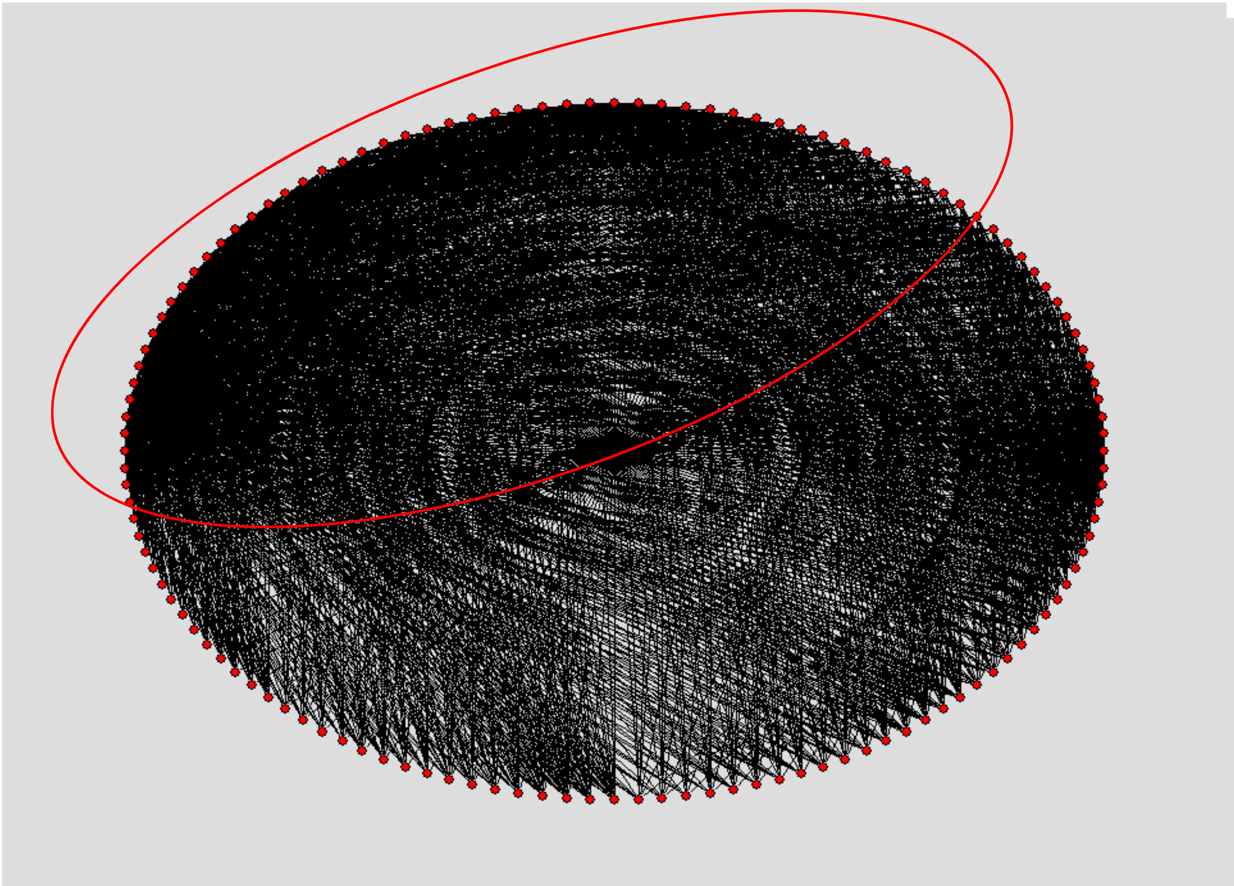
Iterated Local Transitivity (ILT) model (Bonato, Hadi, Horn, Prałat, Wang, 08)

- key paradigm is transitivity: friends of friends are more likely friends

- start with a graph of order n
- to form the graph G_{t+1} for each node x from time t, add a node x', the clone of x, so that xx' is an edge, and x' is joined to each node joined to x

$G_0 = C_4$

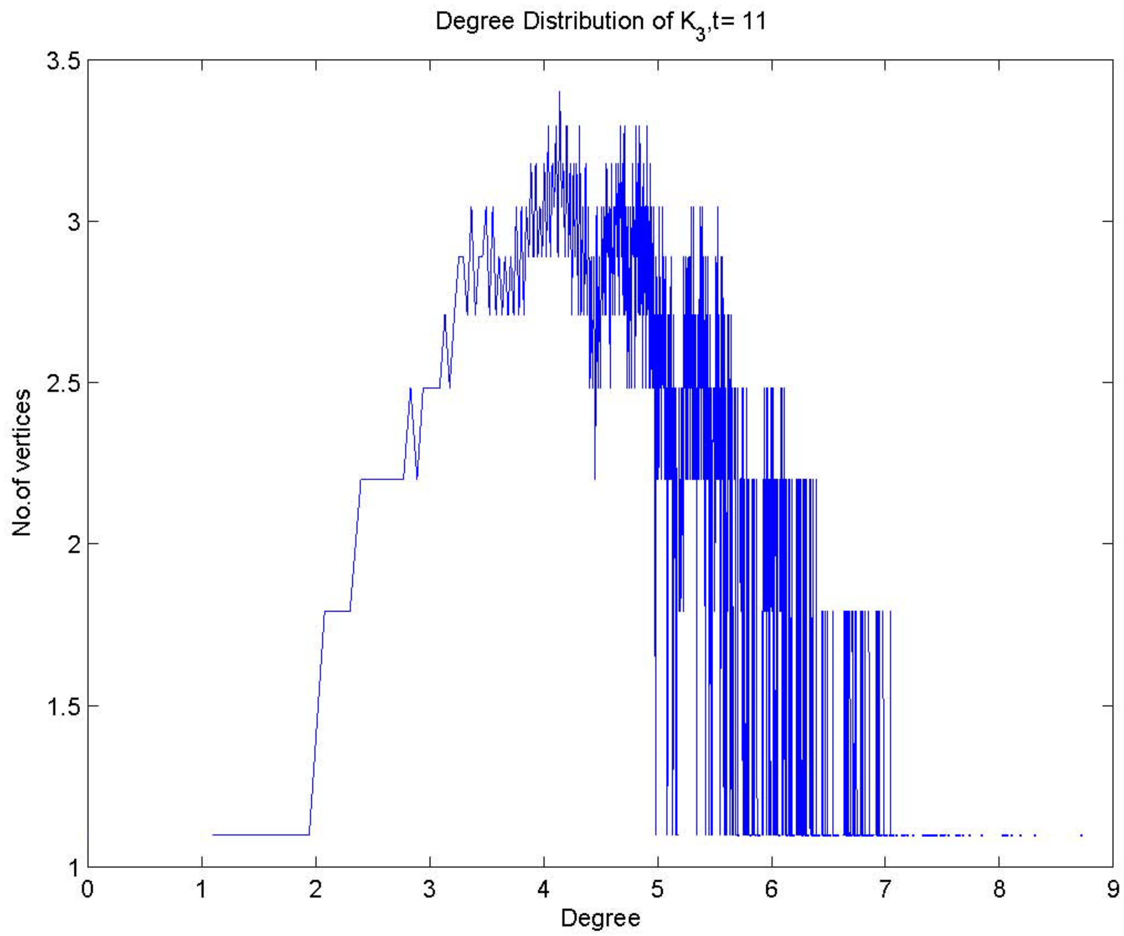




Properties of ILT model

- densification power law
- distances decrease over time
- community structure: bad spectral expansion (Estrada, 06)

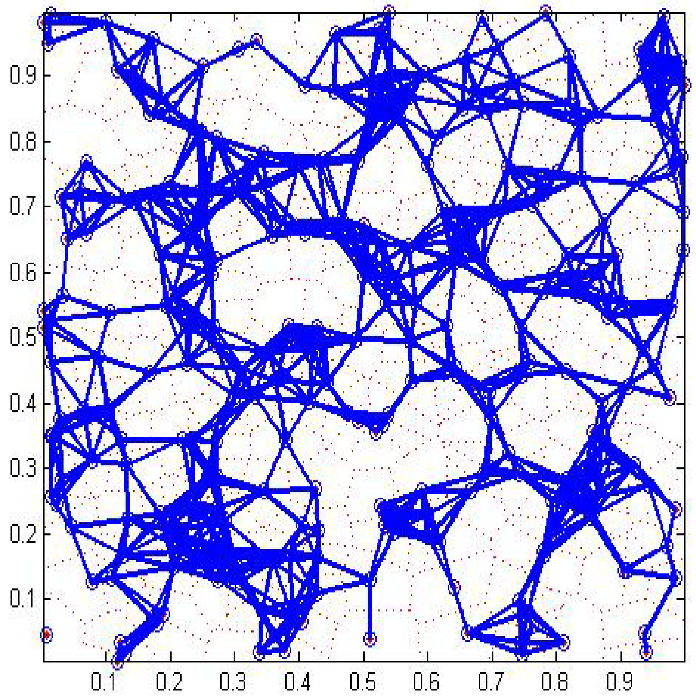
…Degree distribution



Geometry of OSNs?

- OSNs live in social space: proximity of nodes depends on common attributes (such as geography, gender, age, etc.)

- IDEA: embed OSN in 2-, 3- or higher dimensional space



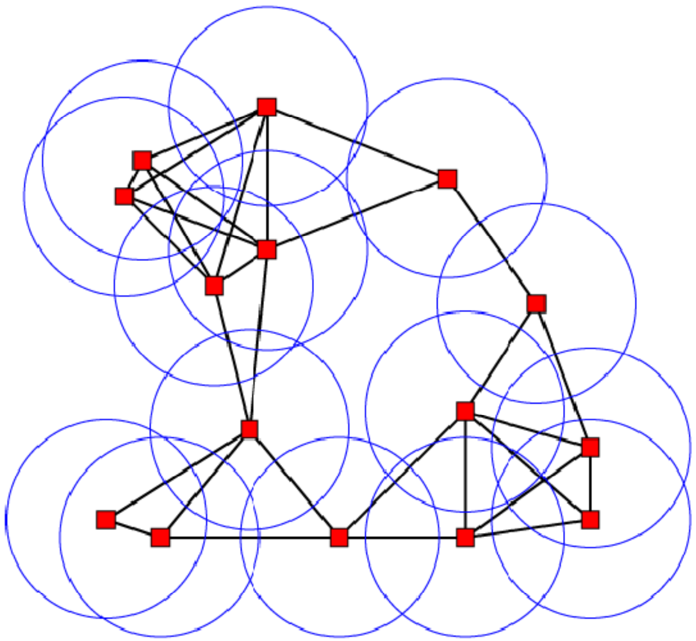
Dimension of an OSN

- dimension of OSN: minimum number of attributes needed to classify or group users

- like game of "20 Questions": each question narrows range of possibilities

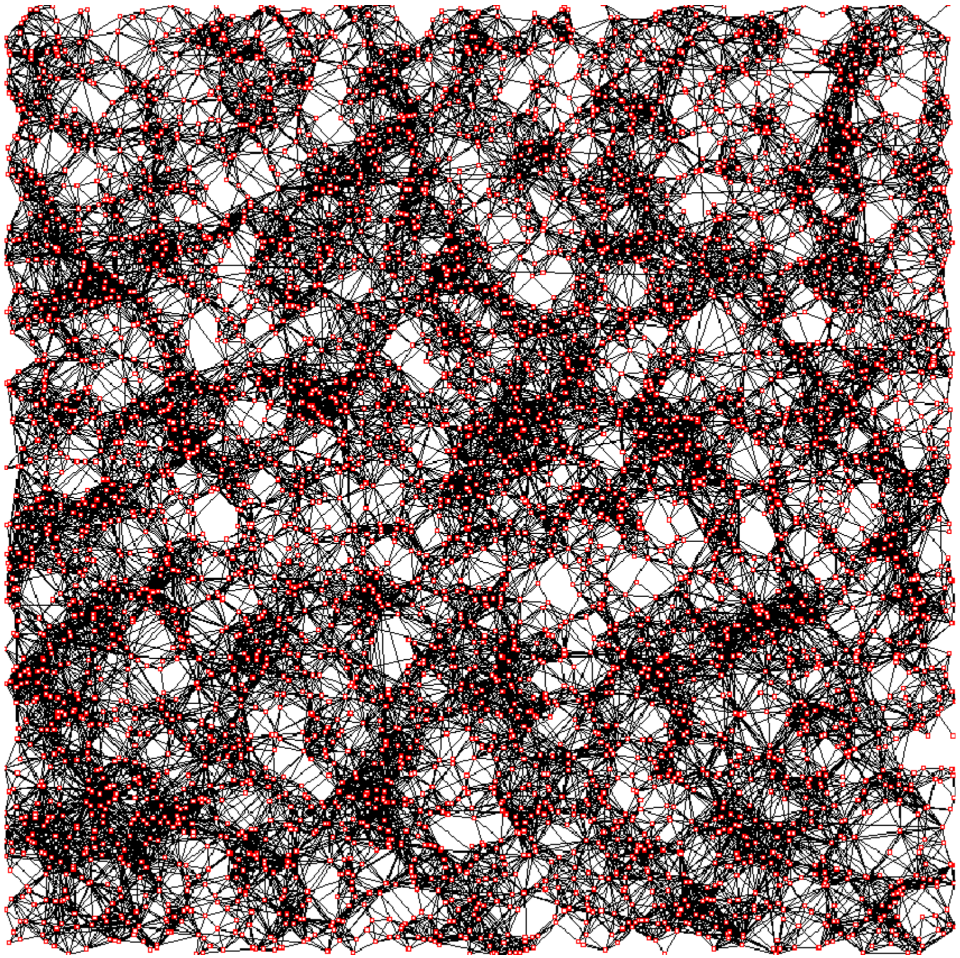
- what is a credible mathematical formula for the dimension of an OSN?

Random geometric graphs



- nodes are randomly placed in space
- each node has a constant sphere of influence
- nodes are joined if their sphere of influence overlap

Simulation with 5000 nodes

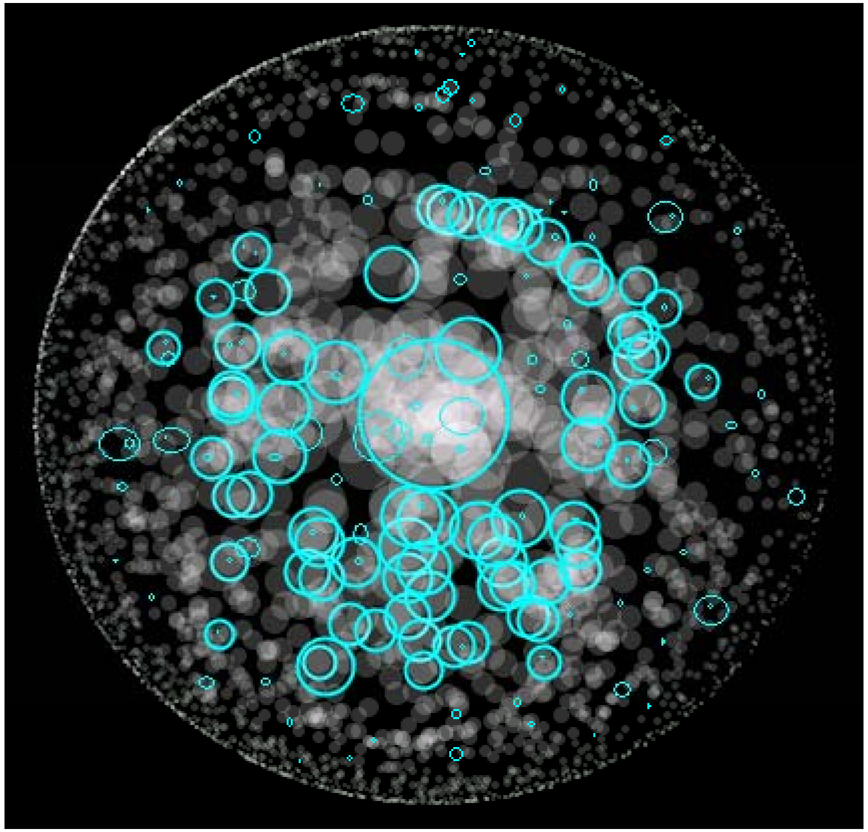


Spatially Preferred Attachment (SPA) model (Aiello, Bonato, Cooper, Janssen, Prałat, 08)

- volume of sphere of influence proportional to in-degree

- nodes are added and spheres of influence shrink over time

- asymptotically almost surely (a.a.s.) leads to power laws graphs



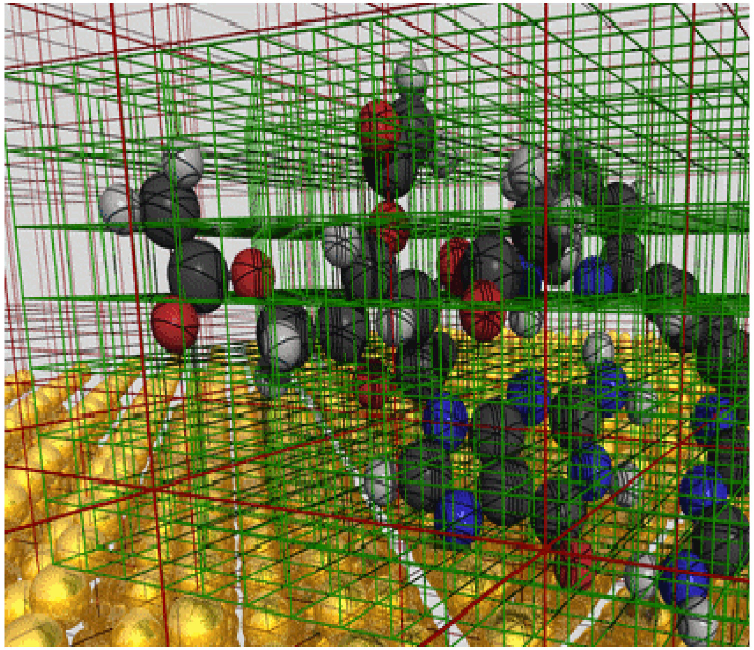
Protean graphs

(Fortunato, Flammini, Menczer,06), (Łuczak, Prałat,06), (Janssen, Prałat,09)

- parameter: α in (0,1)
- each node is ranked 1,2, …, n by some function r
  - 1 is best, n is worst
- at each time-step, one new node v is born, one randomly node chosen dies (and ranking is updated)
- link probability $r^{-\alpha}$
- many ranking schemes a.a.s. lead to power law graphs: random initial ranking, degree, age, etc.

Geometric model for OSNs

- we consider a geometric model of OSNs, where
  - nodes are in m-dimensional hypercube in Euclidean space
  - volume of sphere of influence variable: a function of ranking of nodes



Geometric Protean (GEO-P) Model

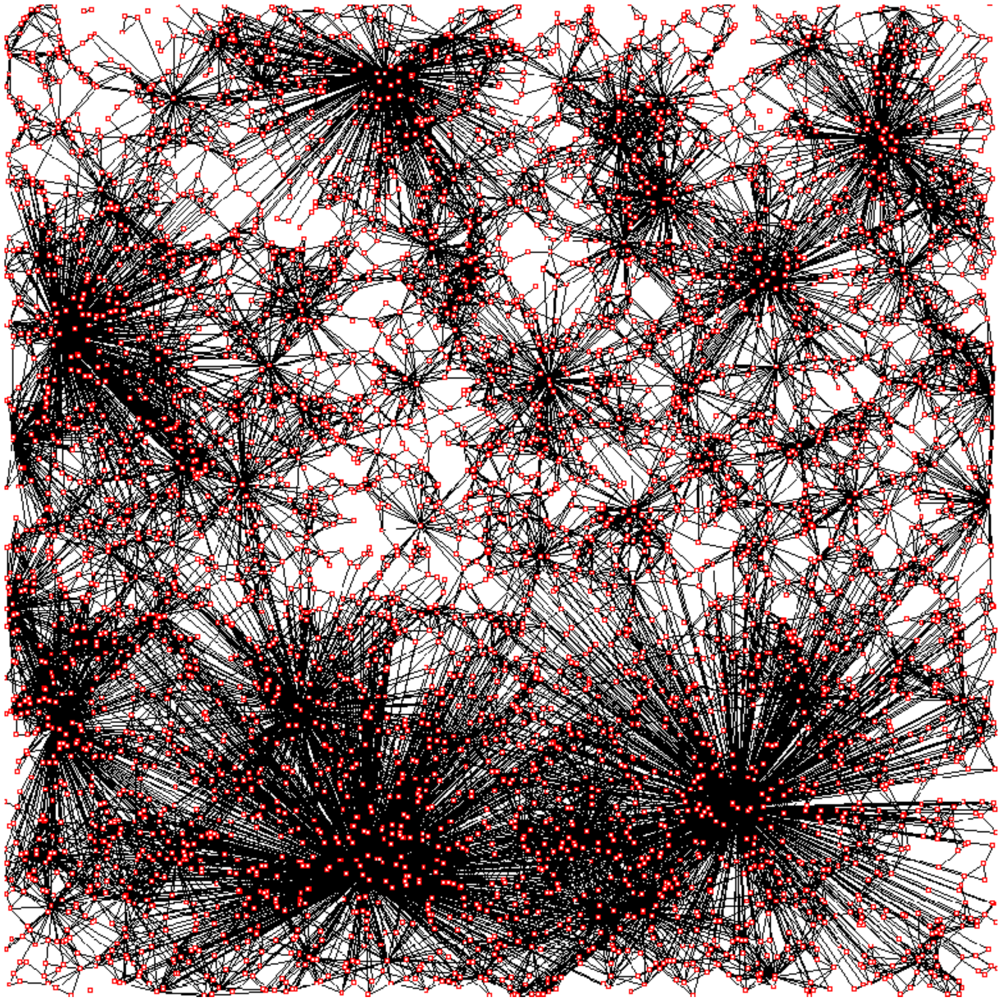
(Bonato, Janssen, Prałat, 10)

- parameters: α, β in $(0,1)$, $\alpha+\beta < 1$; positive integer m
- nodes live in m-dimensional hypercube
- each node is ranked $1,2, \ldots, n$ by some function r
  - we use random initial ranking
- at each time-step, one new node v is born, one randomly node chosen dies (and ranking is updated)
- each existing node u has a sphere of influence with volume $r^{-\alpha}n^{-\beta}$
- add edge uv if v is in the region of influence of u

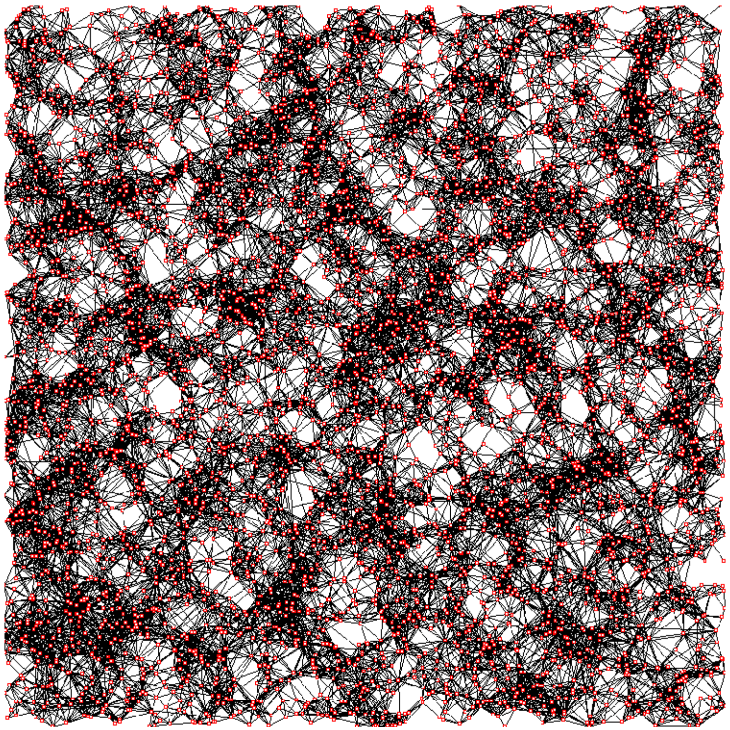
Notes on GEO-P model

- models uses both geometry and ranking
- number of nodes is static: fixed at n
  - order of OSNs at most number of people (roughly…)
- top ranked nodes have larger regions of influence

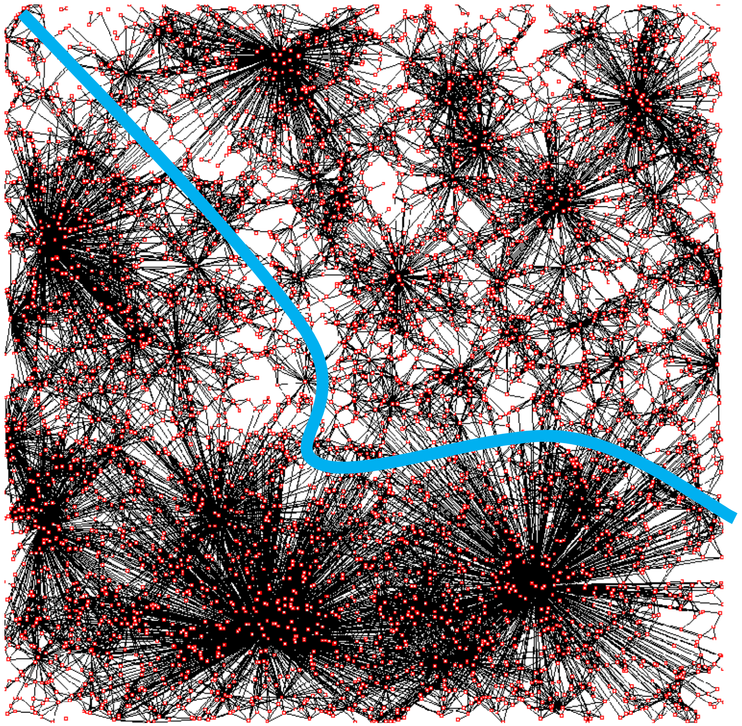
Simulation with 5000 nodes



Simulation with 5000 nodes



random geometric



GEO-P

Properties of the GEO-P model

(Bonato, Janssen, Prałat, 2010)

- a.a.s. the GEO-P model generates graphs with the following properties:
  - power law degree distribution with exponent

    $$b = 1+1/\alpha$$

  - average degree $d = (1+o(1))n^{(1-\alpha-\beta)}/2^{1-\alpha}$
    - densification
  - diameter $D = O(n^{\beta/(1-\alpha)m} \log^{2\alpha/(1-\alpha)m} n)$
    - small world: constant order if $m = C\log n$

Degree Distribution

- for $m < k < M$, a.a.s. the number of nodes of degree at least k equals

$$(1+O(\log^{-1/3} n))\left(\frac{\alpha}{\alpha+1}\right)n^{(1-\beta)/\alpha}k^{-1/\alpha}$$

- $m = n^{1 - \alpha - \beta} \log^{1/2} n$
  - m should be much larger than the minimum degree

- $M = n^{1 - \alpha/2 - \beta} \log^{-2\alpha-1} n$
  - for $k > M$, the expected number of nodes of degree k is too small to guarantee concentration

Density

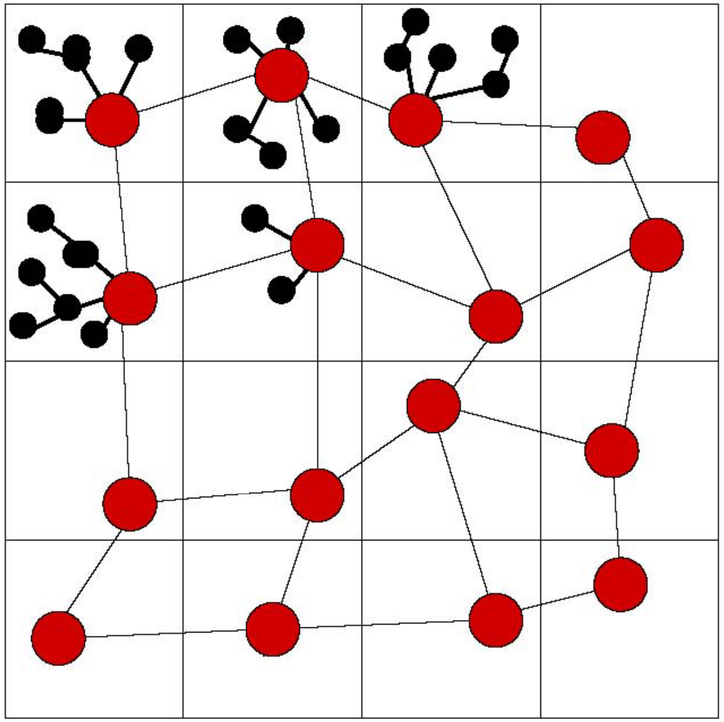
- $i^{-\alpha}n^{-\beta}$ = probability that new node links to node of rank i
- average number of edges added at each time-step

$$\sum_{i=1}^{n} i^{-\alpha} n^{-\beta} \approx \frac{1}{1-\alpha} n^{1-\alpha-\beta}$$

- parameter β controls density
- if $\beta < 1 - \alpha$, then density grows with n (as in real OSNs)

Diameter

- eminent node:
  - old: at least n/2 nodes are younger
  - highly ranked: initial ranking greater than some fixed R
- partition hypercube into small hypercubes
- choose size of hypercubes and R so that
  - a.a.s. each hypercube contains at least $\log^2 n$ eminent nodes
  - sphere of influence of each eminent node covers each hypercube and all neighbouring hypercubes
- choose eminent node in each hypercube: backbone
- show a.a.s. all nodes in hypercube distance at most 2 from backbone



Spectral properties

- the spectral gap λ of G is defined by the difference between the two largest eigenvalues of the adjacency matrix of G
- for G(n,p) random graphs, λ tends to 0 as order grows
- in the GEO-P model, λ is close to 1
- bad expansion/big spectral gaps in the GEO-P model found in social networks but not in the web graph (Estrada, 06)
  - in social networks, there are a higher number of intra- rather than inter-community links

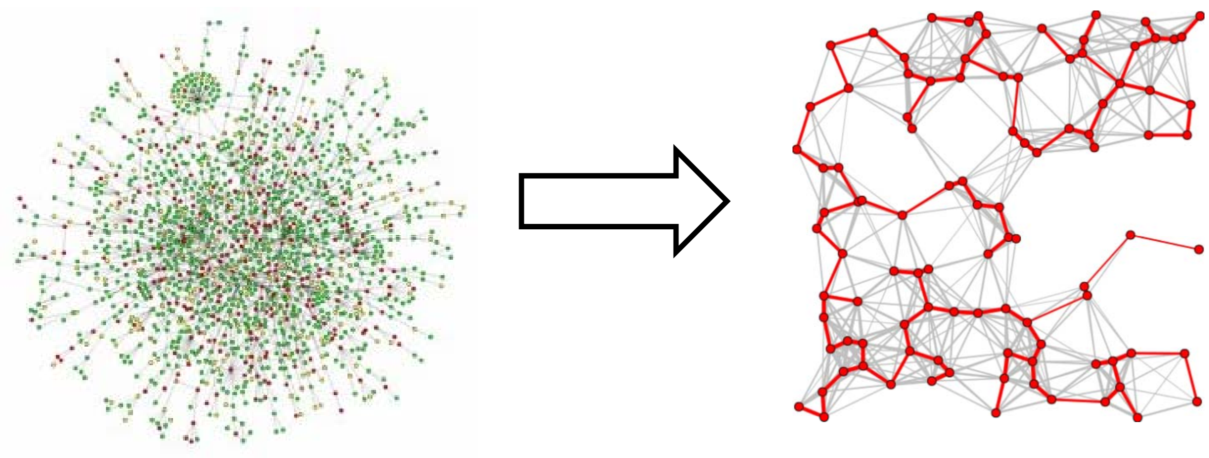
Dimension of OSNs

- given the order of the network n, power law exponent b, average degree d, and diameter D, we can calculate m
- gives formula for dimension of OSN:

$$m = \frac{\log\left(\frac{n}{2d^{\frac{b-1}{b-2}}}\right)}{\log D}$$

Uncovering the hidden reality

- reverse engineering approach
  - given network data (n, b, d, D), dimension of an OSN gives smallest number of attributes needed to identify users
- that is, given the graph structure, we can (theoretically) recover the social space



6 Dimensions of Separation

OSN	Dimension
YouTube	6
Twitter	4
Flickr	4
Cyworld	7

Research directions

- fitting GEO-P model to data
  - is theoretical estimate of log n dimension accurate?
  - find similarity measures (see PPI literature)
- community detection
  - first map network in social space?
- spread of influence
  - SIS, SIR models
  - Graph theory: firefighting, Cops and Robbers

- *preprints, reprints, contact*:

search: "Anthony Bonato"

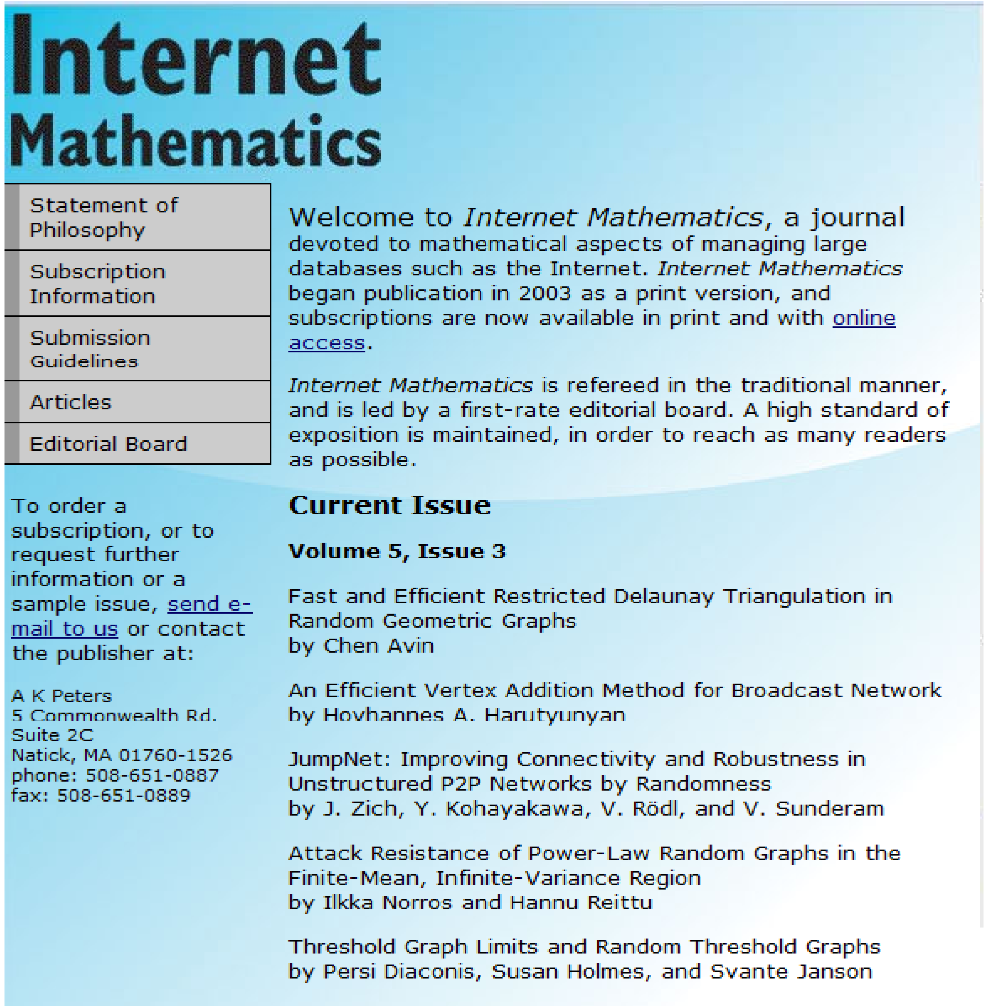
MoMiNIS

MITACS

Communications Security Establishment
Centre de la sécurité des télécommunications

NSERC CRSNG

RYERSON UNIVERSITY


Internet Mathematics

Statement of Philosophy
Subscription Information
Submission Guidelines
Articles
Editorial Board

Welcome to *Internet Mathematics*, a journal devoted to mathematical aspects of managing large databases such as the Internet. *Internet Mathematics* began publication in 2003 as a print version, and subscriptions are now available in print and with online access.

Internet Mathematics is refereed in the traditional manner, and is led by a first-rate editorial board. A high standard of exposition is maintained, in order to reach as many readers as possible.

To order a subscription, or to request further information or a sample issue, send e-mail to us or contact the publisher at:

A K Peters
5 Commonwealth Rd.
Suite 2C
Natick, MA 01760-1526
phone: 508-651-0887
fax: 508-651-0889

Current Issue

Volume 5, Issue 3

Fast and Efficient Restricted Delaunay Triangulation in Random Geometric Graphs
by Chen Avin

An Efficient Vertex Addition Method for Broadcast Network
by Hovhannes A. Harutyunyan

JumpNet: Improving Connectivity and Robustness in Unstructured P2P Networks by Randomness
by J. Zich, Y. Kohayakawa, V. Rödl, and V. Sunderam

Attack Resistance of Power-Law Random Graphs in the Finite-Mean, Infinite-Variance Region
by Ilkka Norros and Hannu Reittu

Threshold Graph Limits and Random Threshold Graphs
by Persi Diaconis, Susan Holmes, and Svante Janson

- journal relaunch
- new editors
- accepting theoretical and empirical papers on complex networks, OSNs, biological networks