Graph Theory in the Information Age

Abbas Mehrabian
University of Waterloo

16 January 2013

Huge networks everywhere!

- The web graph: 30 billion nodes
- Social networks: 7 billion
- Online social networks: 500 million
- Protein interactions
- Human brain: 100 billion
- 10 gr Diamond crystal: $5 * 10^{23}$
- Transistors on chips

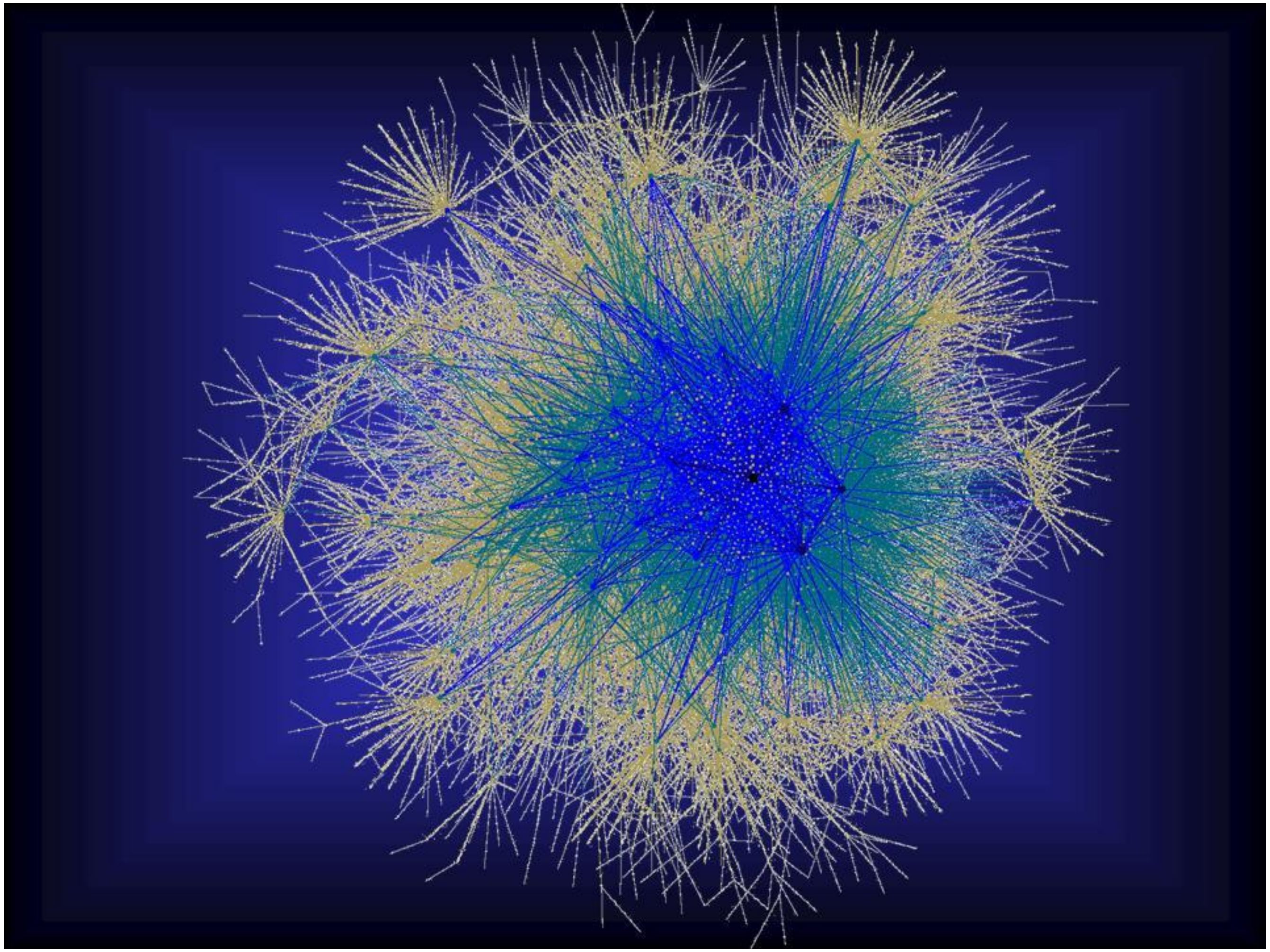


Image from: http://www.math.ucsd.edu/~fan/graphs/gallery/bgpps.jpg

Challenges

1. Analyze their structure
2. Model them
3. Approximate them
4. Run algorithms on them

PROPERTIES OF THESE NETWORKS

Power-law degree distribution

- n = number of vertices
- n_k = number of vertices with degree k

$$n_k \approx C\, n k^{-\beta}$$

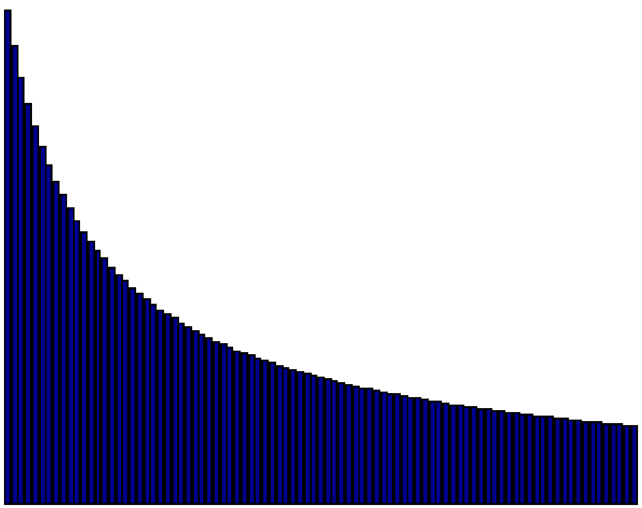


Figure 1: *Power law degree distribution.*

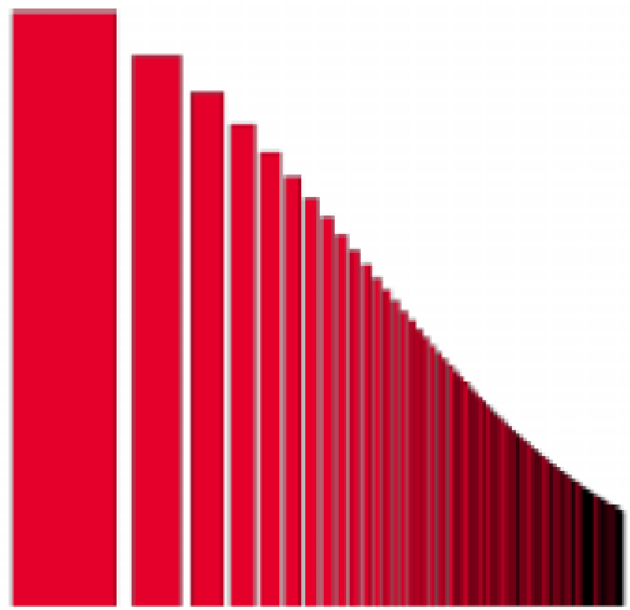
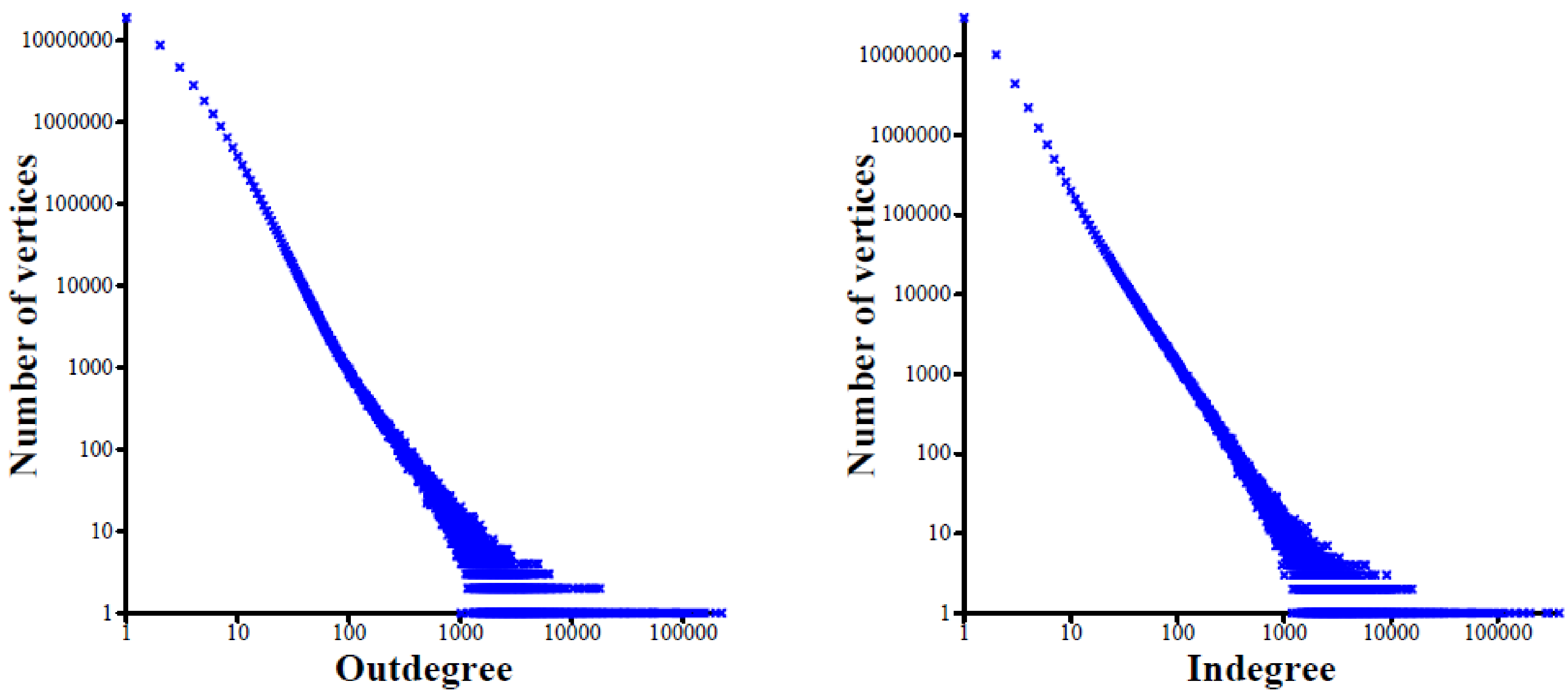


Figure 2: *Log-scale of Figure 1.*

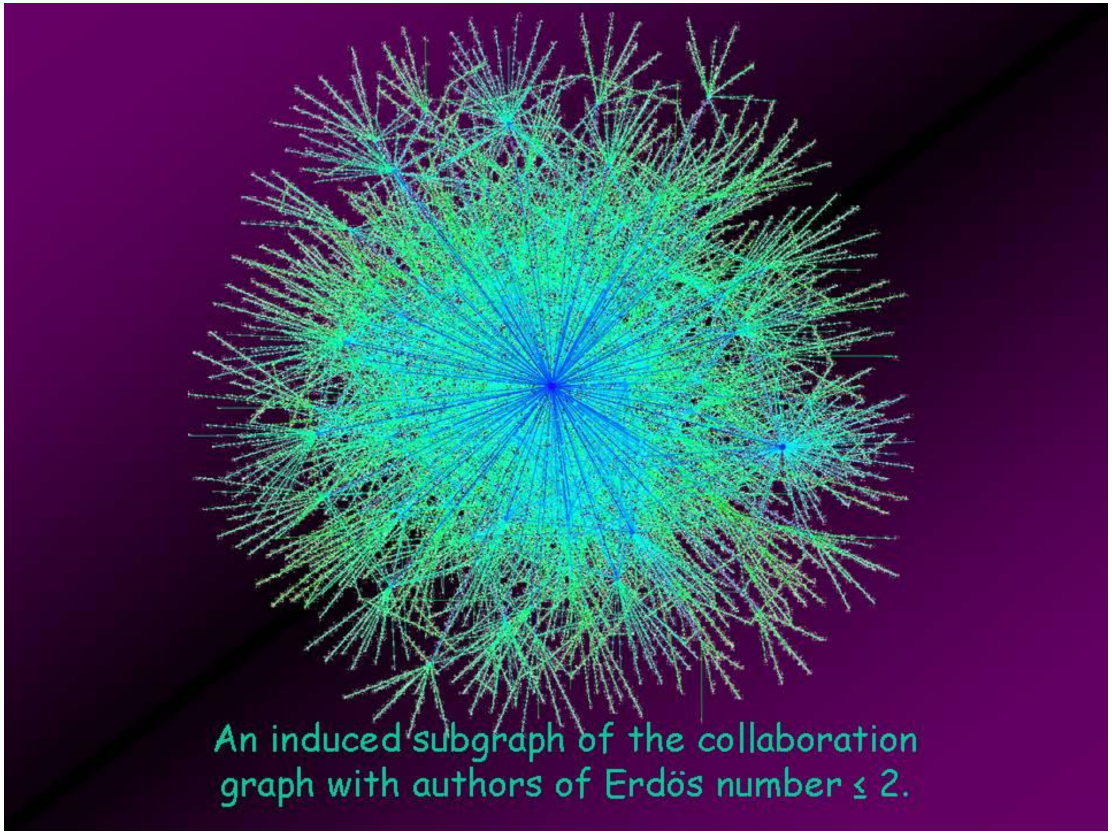
Power-law degree distribution

- Call Graph (AT&T): $\beta = 2.1$



Power-law degree distribution

- Kumar et al. (IBM): Web crawl of 40 million web pages in 1997
- $\beta_{in} = 2.1; \beta_{out} = 2.7$

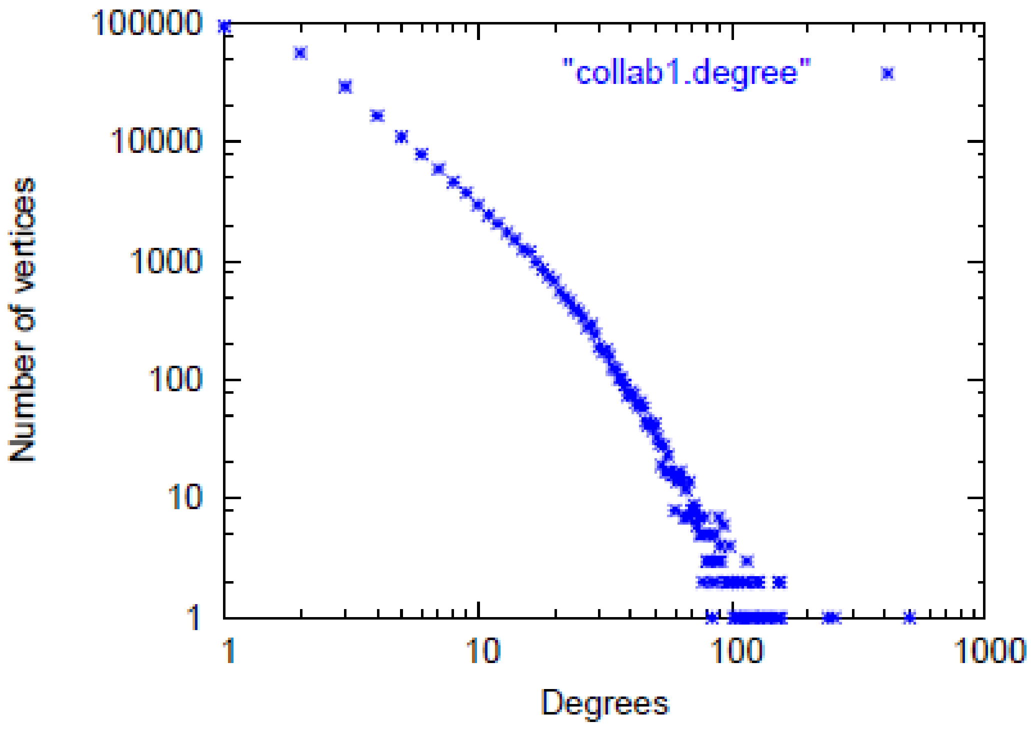


Collaboration Graph

Image from: http://www.math.ucsd.edu/~fan/graphs/gallery/collabb.jpg

Power-law degree distribution

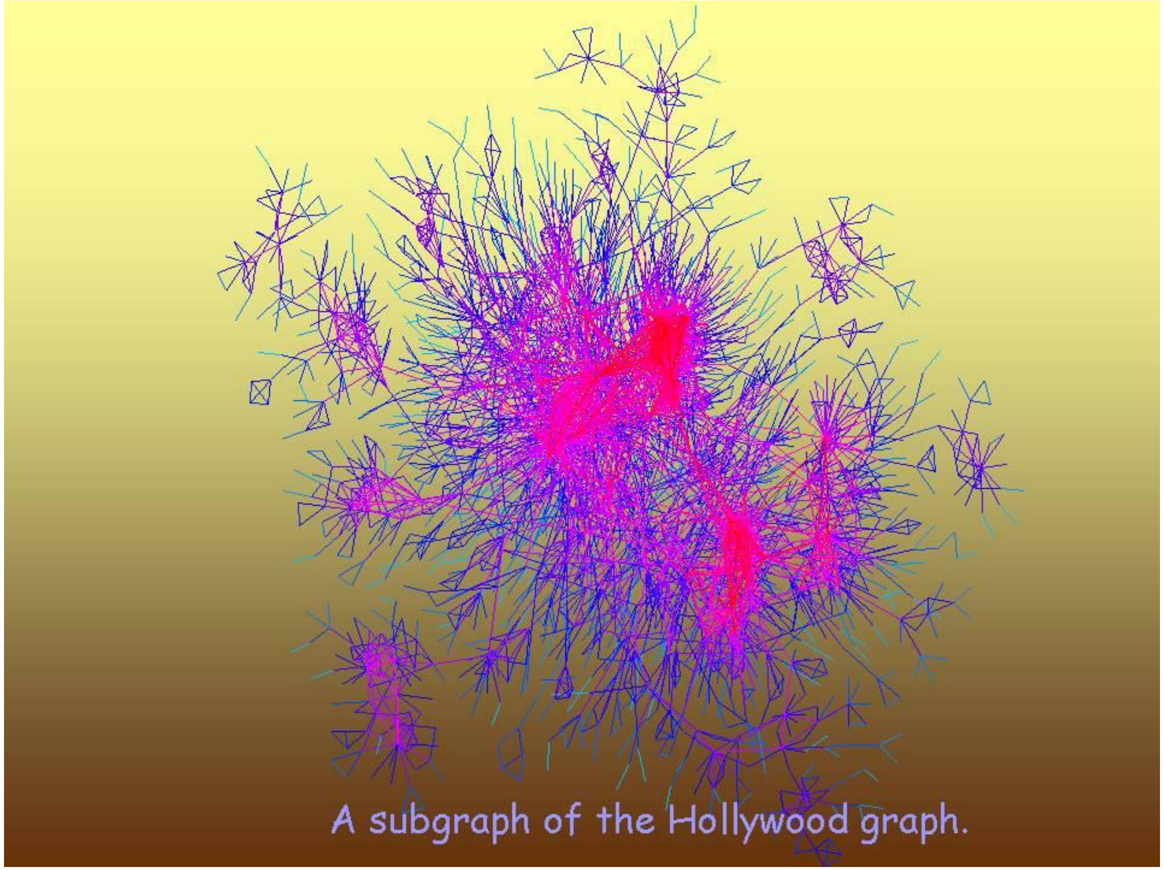
- Collaboration graph (2004) 401000 nodes: $\beta = 2.46$



Power-law degree distribution

- Hollywood graph (225000 nodes) $\beta = 2.3$

Image from: http://www.math.ucsd.edu/~fan/graphs/gallery/holys.jpg



Power-law degree distribution

- Biological networks:
  1. Yeast protein-protein networks: $\beta = 1.6$
  2. Yeast gene expression networks: $\beta \in [1.4, 1.7]$
  3. Gene functional interaction network: $\beta = 1.6$

Notation

$f = O(g)$ if $f \leq cg$ for a constant $c > 0$

$f = \Omega(g)$ if $f \geq cg$ for a constant $c > 0$

$f = \Theta(g)$ if $c_1 g \leq f \leq c_2 g$ for constants $c_1, c_2 > 0$

Power-law degree distribution

$$n_k \approx C\, nk^{-\beta}$$

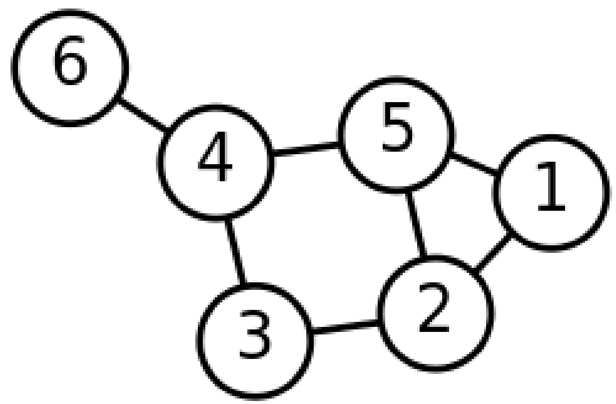
Assume $\beta > 2$.

$$2|E(G)| = \sum deg(v) = \sum kn_k \approx Cn\sum k^{1-\beta}$$
$$= O(n)$$

Many real-world networks are sparse!

Small world phenomenon

Metric on graph vertices



d(1,4) = 2

d(2,6) = 3

Small world phenomenon

S = pair of vertices with finite distance

L(G) = average distance between pairs in S

$$diam(G) = \max\{d(u,v) : \{u,v\} \in S\}$$

Small world phenomenon

1. Average distance and the diameter are small (usually $O(\log n)$)

Examples:

- Six degrees of separation (Milgram’s test)
- Broder et al. (2000): $L(Web\ graph) = 6.8$

Small world phenomenon

2. Two vertices having a common neighbour are more likely to be adjacent.

Local clustering coefficient of v:
probability that two random neighbours of v are adjacent

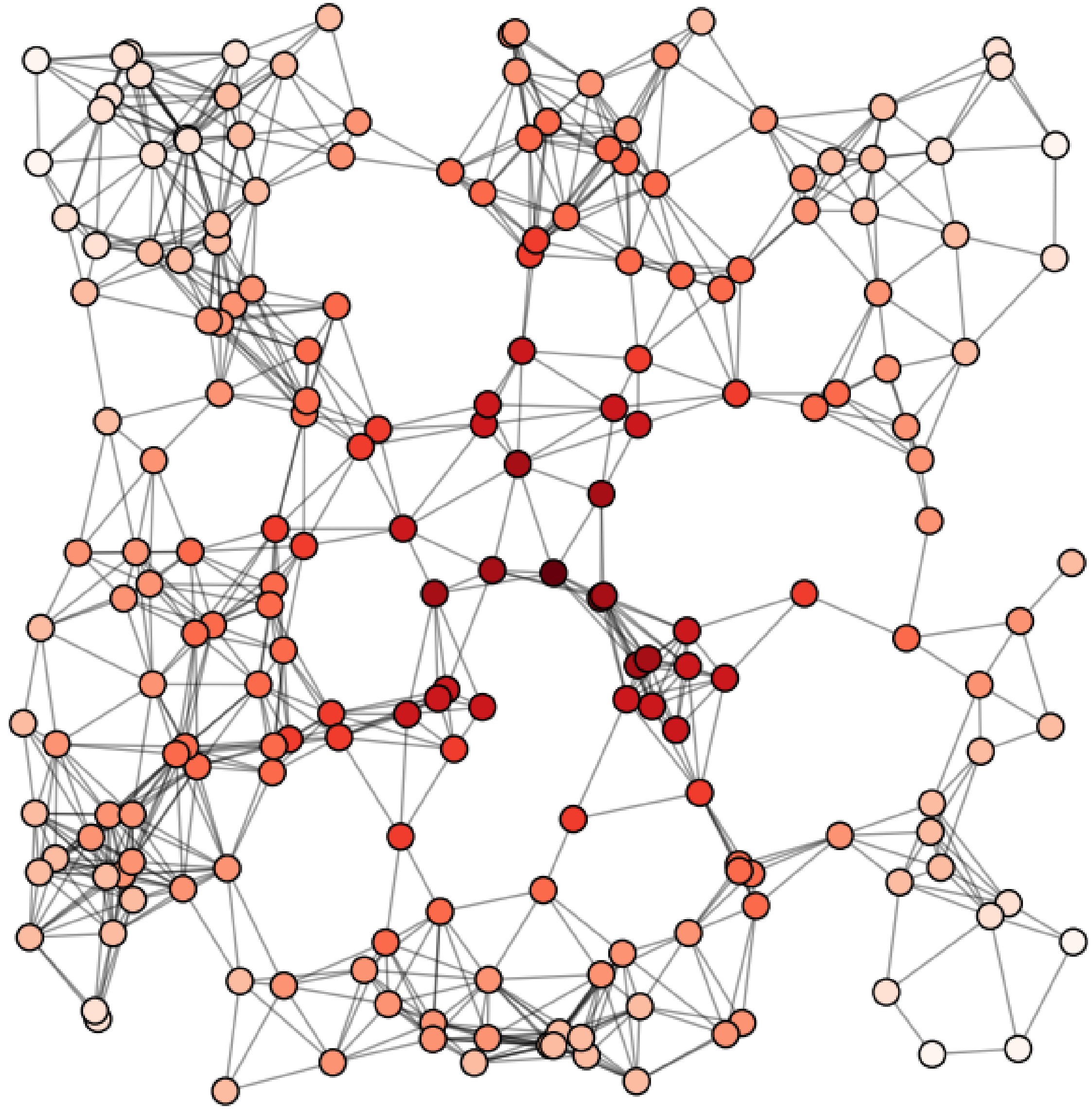


Image from: http://networkx.lanl.gov/archive/networkx-1.4/_images/random_geometric_graph.png

MODELS

Notation

With high probability:

with probability approaching 1 as *n* goes to infinity

Erdös-Renyi Random Graphs

- $\mathrm{G}\ (n, p)$

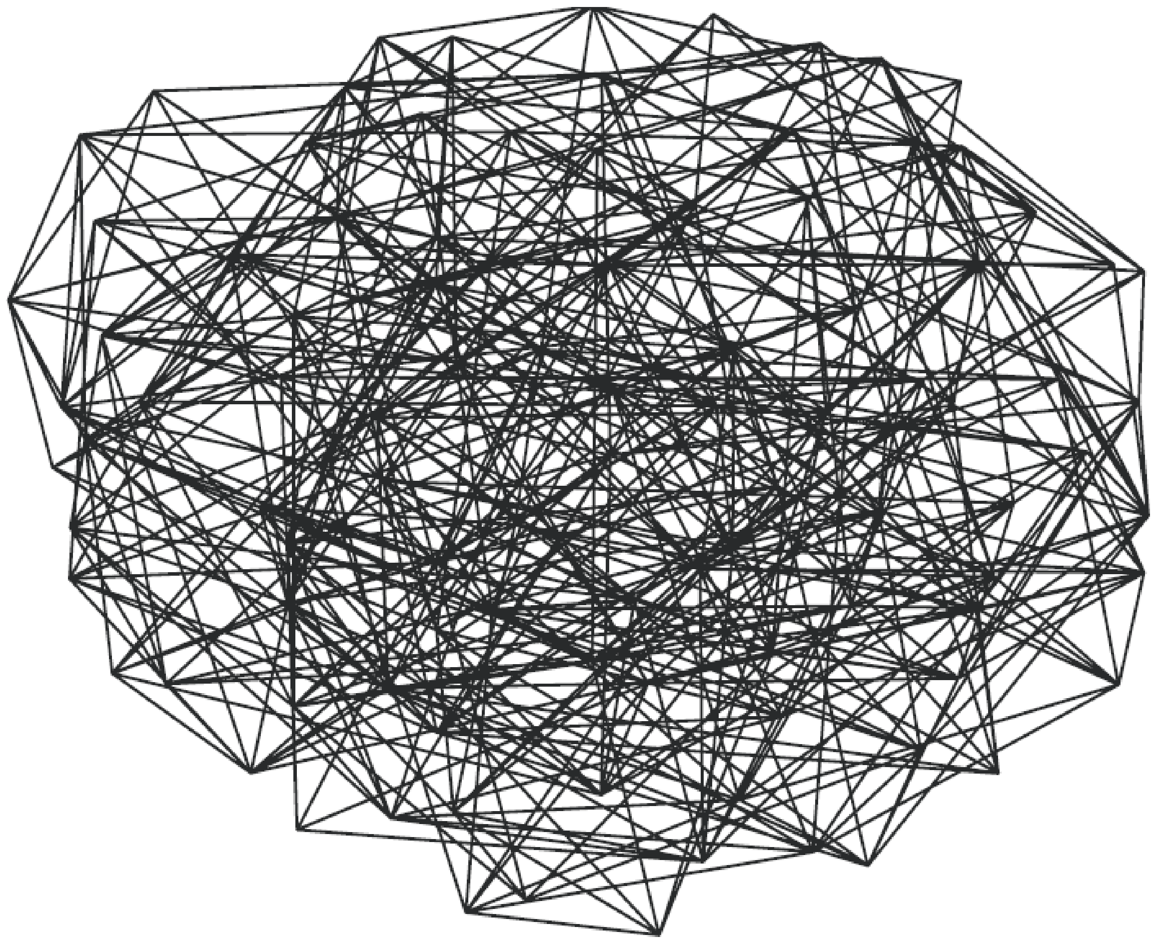


Figure 4.1. A graph with 100 vertices and edges drawn with probability $\frac{1}{2}$.

Erdös-Renyi Random Graphs

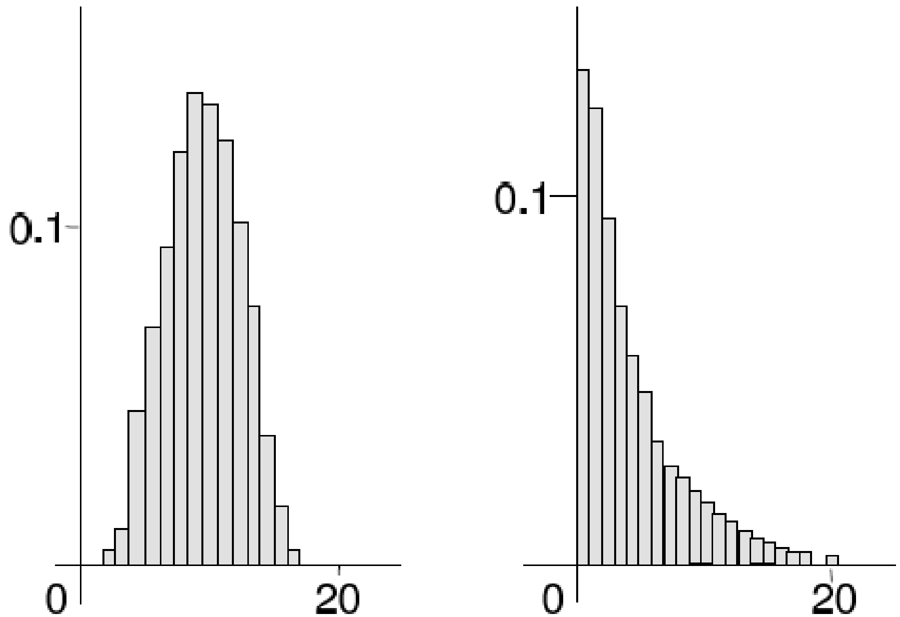
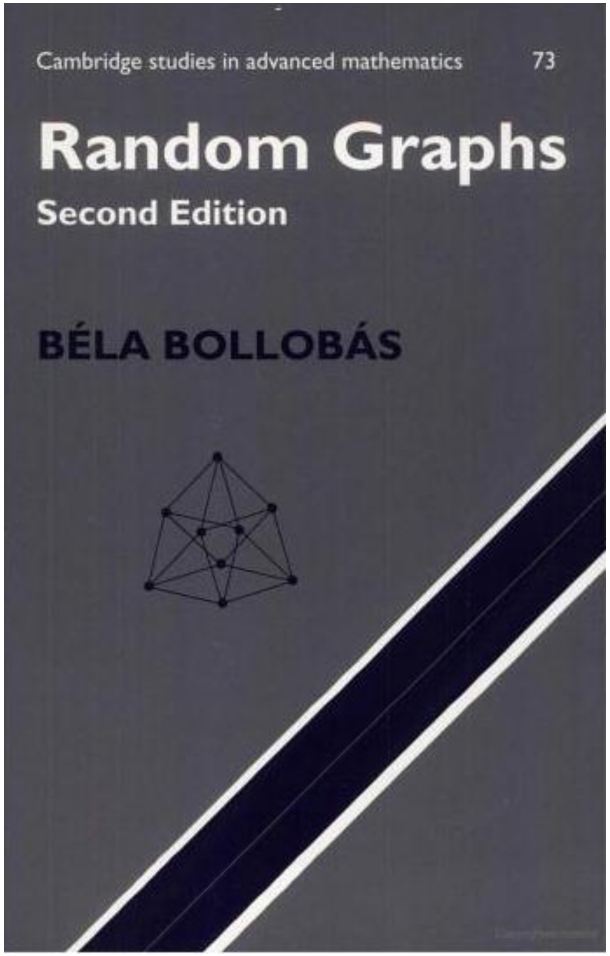


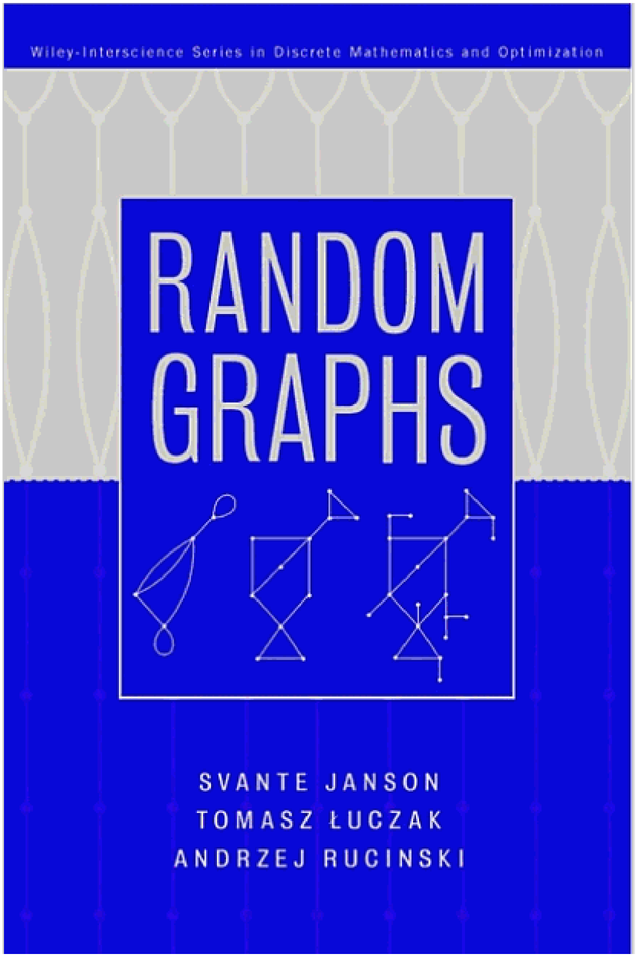
FIGURE 1.2. Degree distributions of an Erdős–Rényi random graph on 100 nodes with edge density .1 (left) and of a real life graph with similar parameters (right). The main feature to observe about the latter is not that the largest frequency is 1, but that it is much more stretched out.

Erdös-Renyi Random Graphs

- For $np > 5 \log n$, have logarithmic diameter.
- Clustering coefficient is small

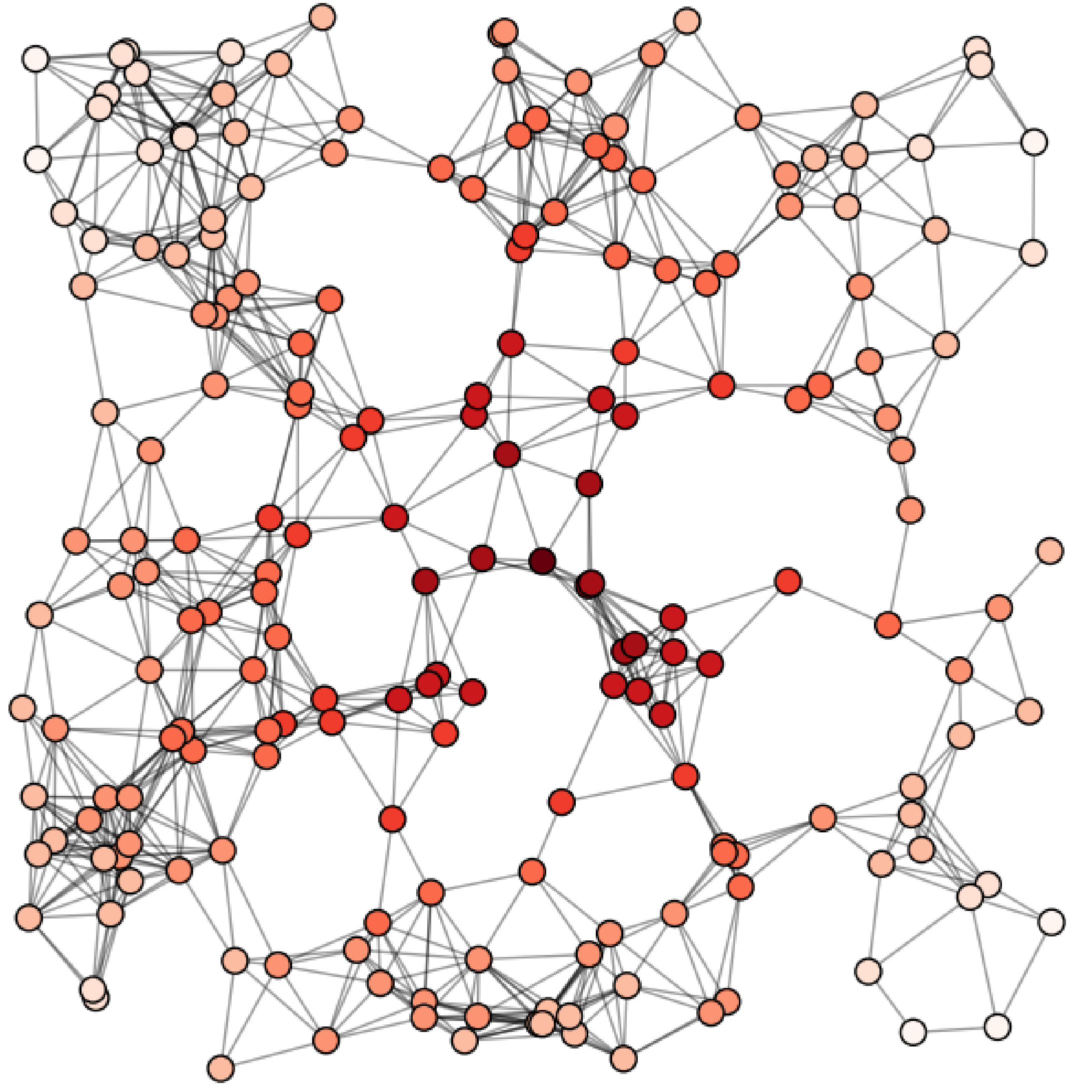
Erdös-Renyi Random Graphs: further reading



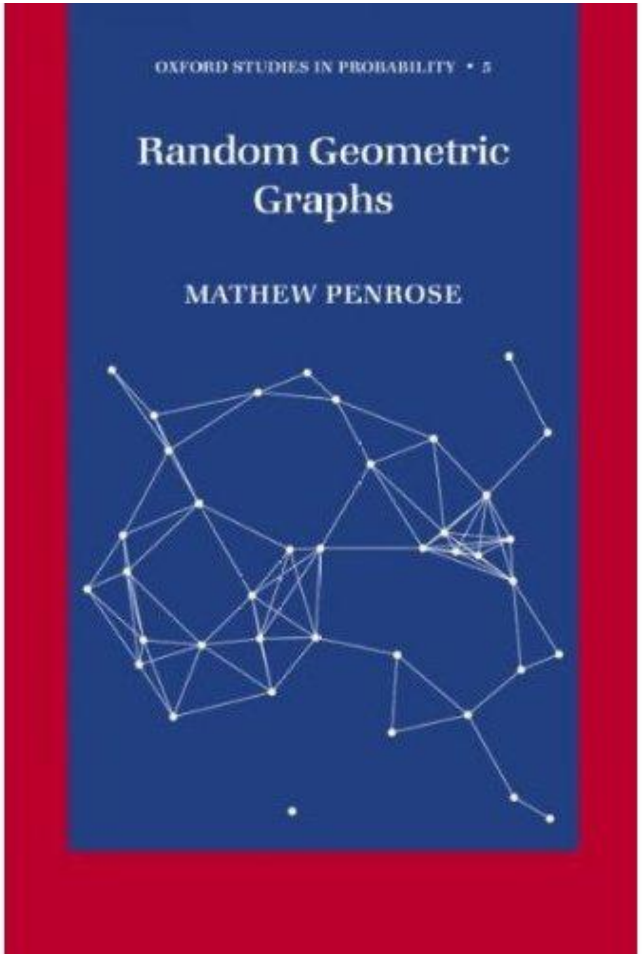


Random Geometric Graphs

- $G(n, r)$



Random Geometric Graphs: further reading



Random Graphs with given expected degree sequence

- Let $w = (w_1, w_2, \dots, w_n)$ be given
- Build a random graph on n vertices with

$$p(i,j) = \frac{w_i w_j}{\sum w_k}$$

- The average degree of vertex j is (almost) w_j

Random Graphs with given expected degree sequence

- Let $d = \frac{\sum w_i^2}{\sum w_i}$
- Let $(w_1, w_2, \dots, w_n)$ be power-law with exponent β.
- Chung and Lu (2007):
  $diam = \Theta(\log n)\ for\ \beta > 2$

  1. $if\ \beta > 3\ then\ L(G) \sim \frac{\log n}{\log d}$;
  2. $2 < \beta < 3\ then\ L(G) = O(\log \log n)$;

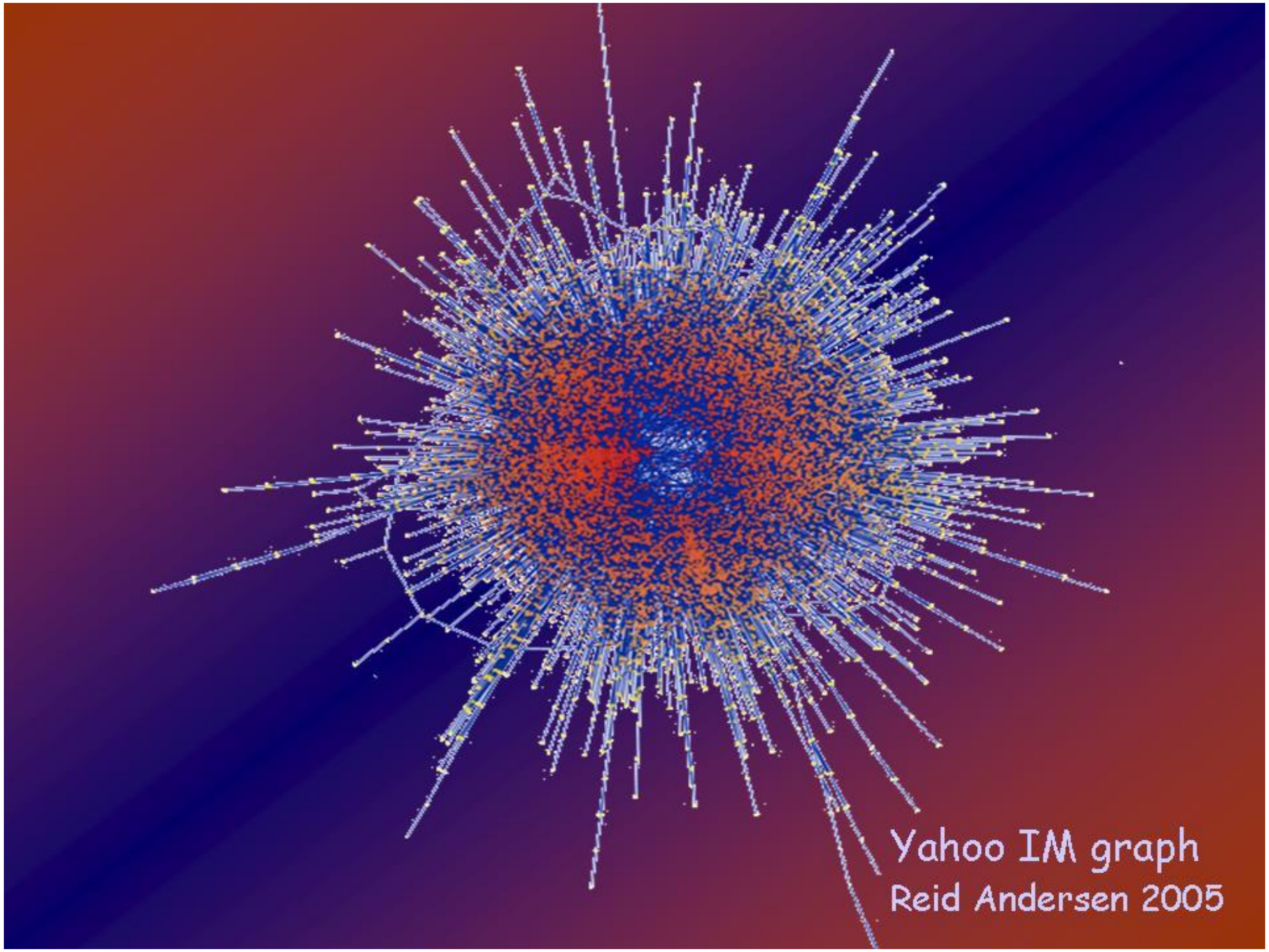


Image from: http://www.math.ucsd.edu/~fan/graphs/gallery/ims.jpg

Linearized Chord Diagram

- Bollobas, Riordan, Spencer, Tusnady (2001)
- An “evolving random graph”
- First “preferential attachment model” analyzed rigorously

Linearized Chord Diagram

- G(m,t)
- Start with a vertex
- In step i, add one new vertex and join it to exactly one old vertex with probability proportional to their degree
- After t steps, a tree with t vertices is obtained.
- Merge every m consecutive vertices
- Obtain a graph with t/m vertices

Linearized Chord Diagram

- Bollobas, Riordan, Spencer, Tusnady (2001).

$.For\ fixed\ m, \varepsilon, with\ probability\ tending\ to\ 1\ as\ t \to \infty, for\ all\ 0 \leq k \leq \frac{t}{m},$

$$1 - \varepsilon < \frac{n_k}{Ck^3} < 1 + \varepsilon$$

- Bollobas and Riordan (2004).
- $With\ probability\ tending\ to\ 1\ as\ t \to \infty,$

$$diameter \sim \frac{\log t}{\log log\ t}$$

Linearized chord diagram

- More preferential attachment models:
  1. Aiello, Chung, Lu (2002): with any $\beta \in (2, \infty)$
  2. Cooper, Frieze (2003): many parameters
  3. Buckley, Osthus (2004).
  4. Bollobas, Borgs, Chayes, Riordan (2004)

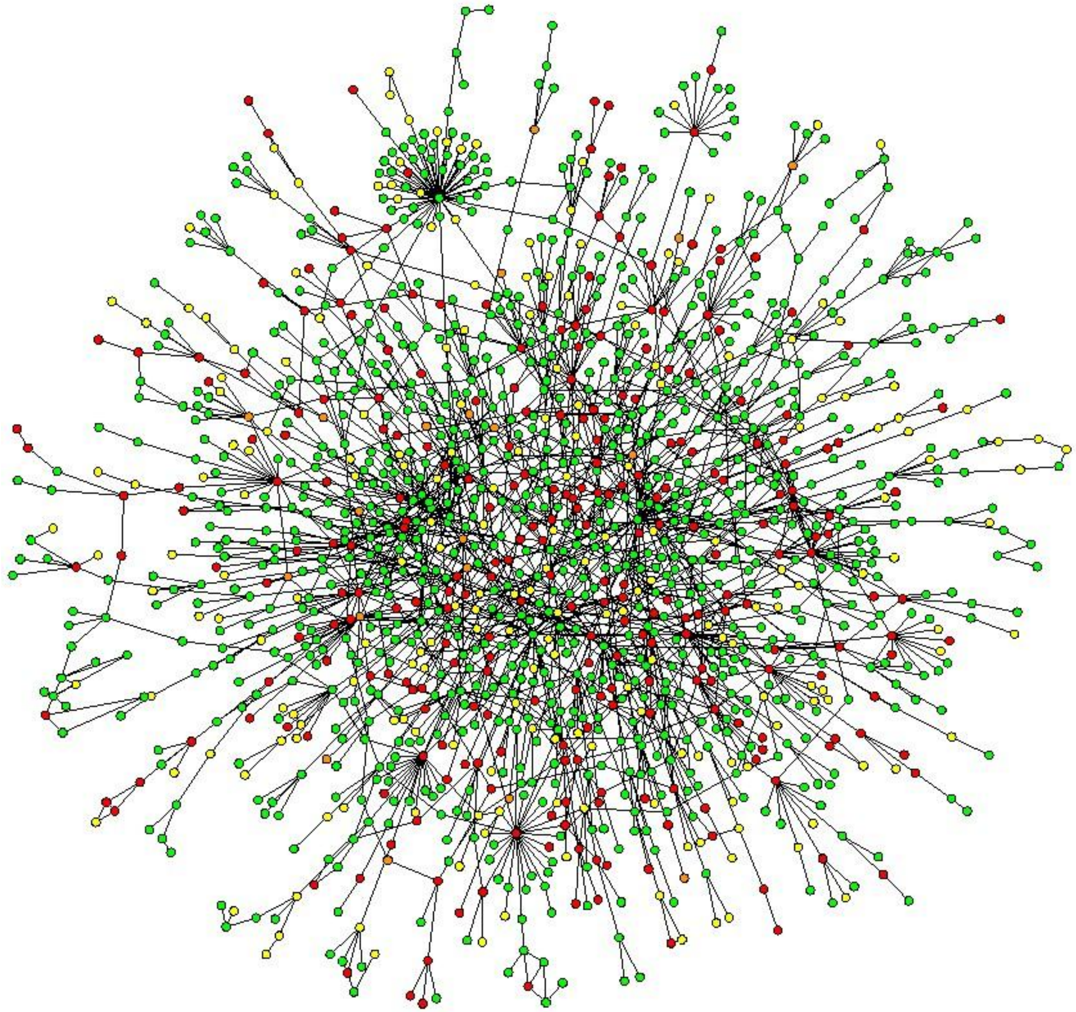
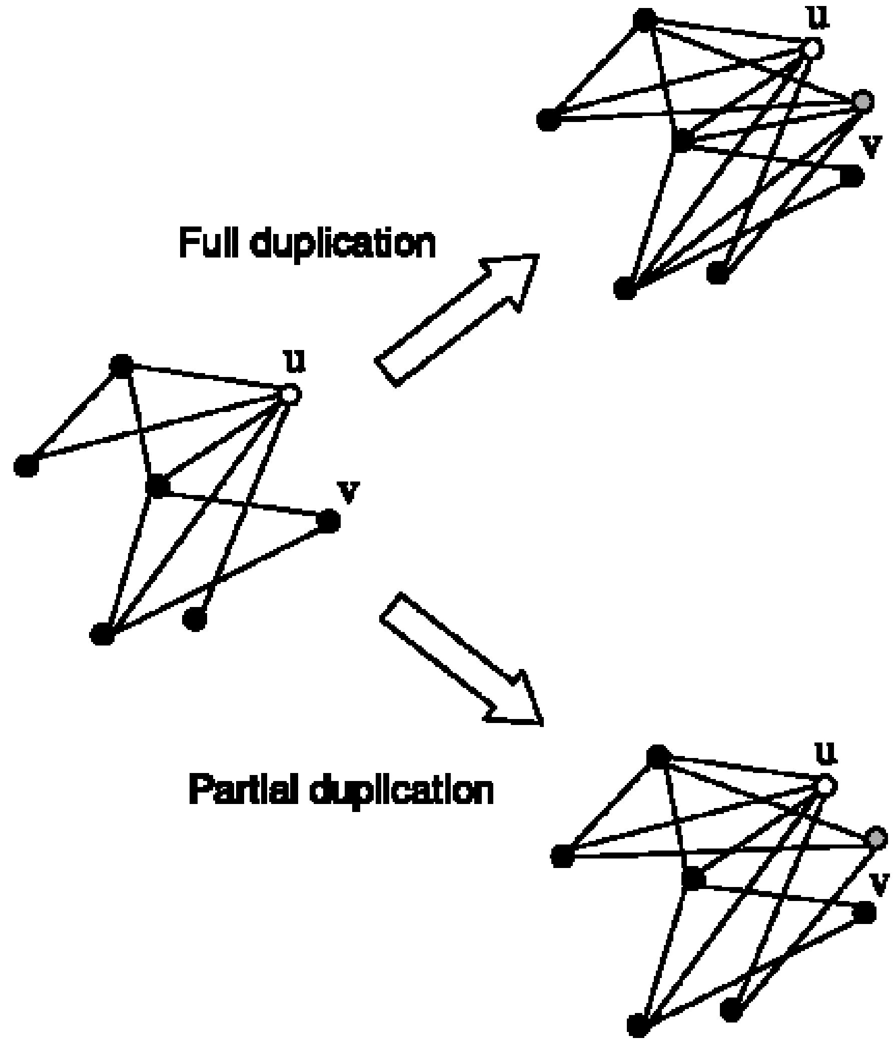


Image from: http://www.bordalierinstitute.com/images/yeastProteinInteractionNetwork.jpg



Duplication Model

Image from: Chung, Lu, Dewey, and Galas. Duplication Models for Biological Networks. Journal of Computational Biology. 10 (5), 2003 pp. 677--687.

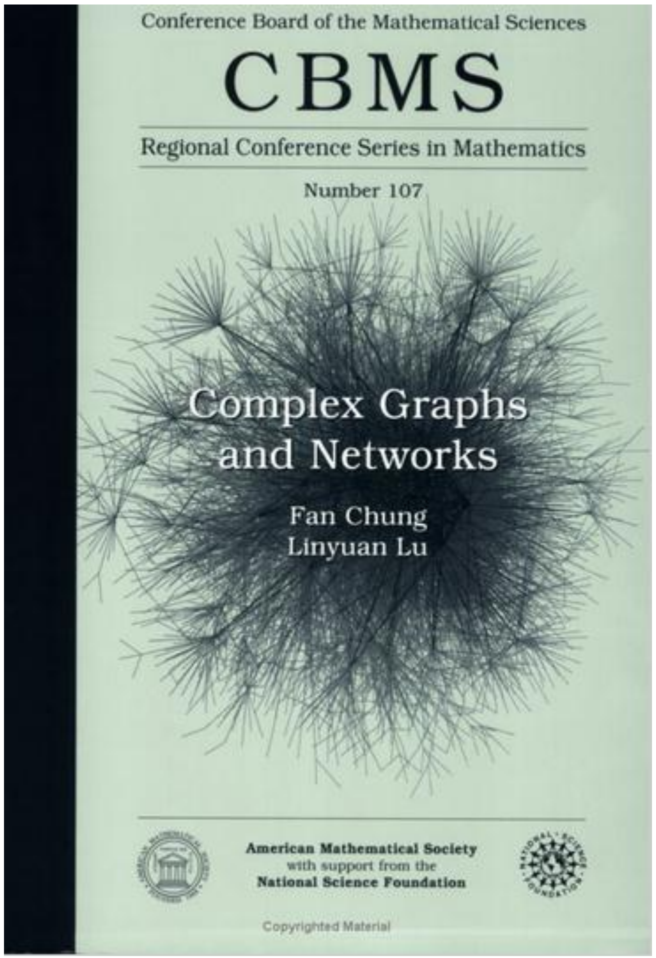
Partial Duplication Model

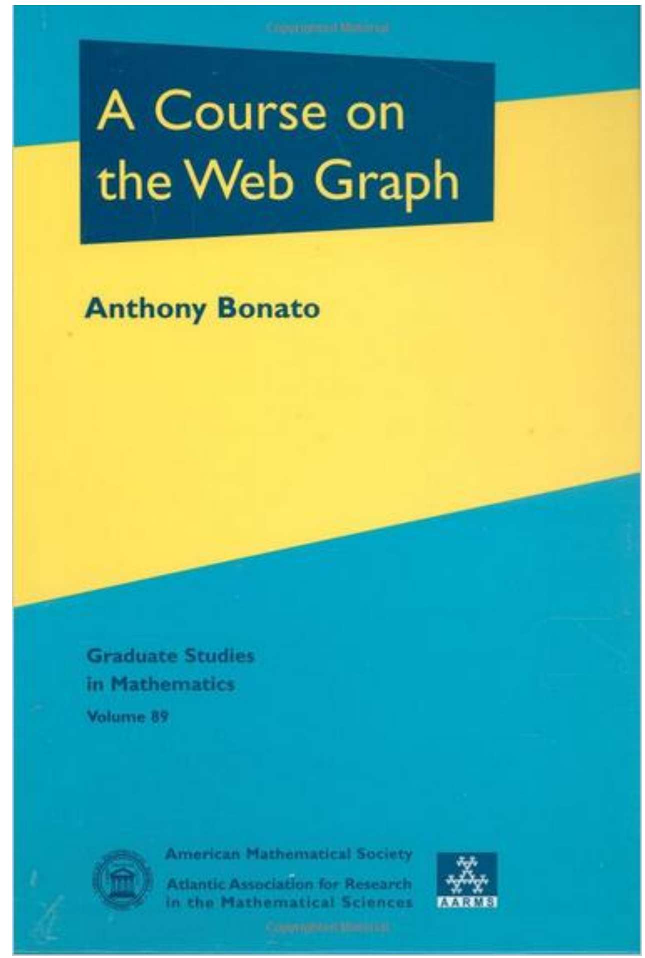
- Parameter: p
- Chung, Lu, Dewey, Galas (2003):
  Theorem. The partial duplication model generates power-law graphs with exponent satisfying

$$p(\beta - 1) = 1 - p^{\beta - 1}$$

So, if 0.5 < p < 1 then $\beta < 2$.

Real-world networks: further reading





Real-world networks: further reading

- Chung and Lu. Complex Graphs and Networks. AMS. 2006;
  chapter1: http://www.ams.org/bookstore/pspdf/cbms-107-prev.pdf
- Bonato. A Survey of Models of the Web Graph. Proceedings of CAAN, 2004;
  http://www.math.ryerson.ca/~abonato/WEBSURV2.pdf

Graph Limits

- In 2003 people in Microsoft research started to define notions of “convergence” for sequences of graphs with increasing size…

- A book on this topic has been published in 2012

Copyrighted Material

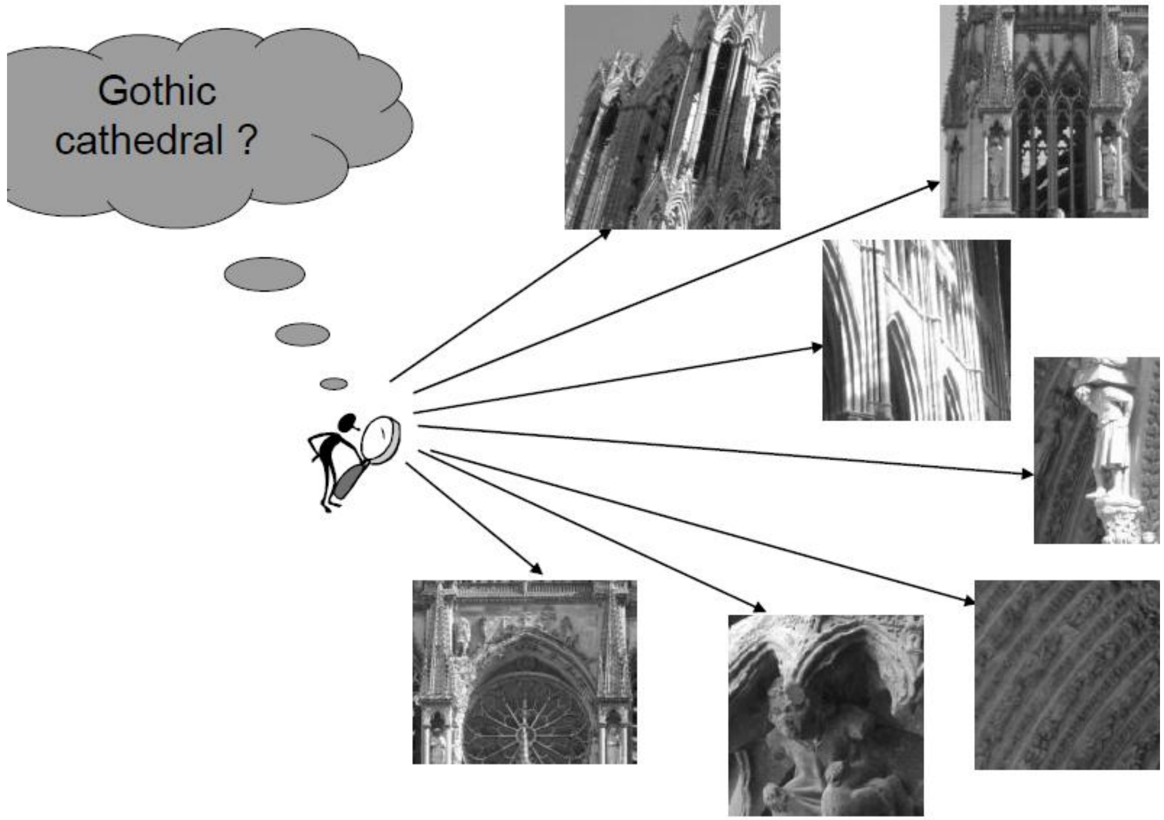
American Mathematical Society

Colloquium Publications
Volume 60

UM PUBLICATIONS

Large Networks and Graph Limits

László Lovász



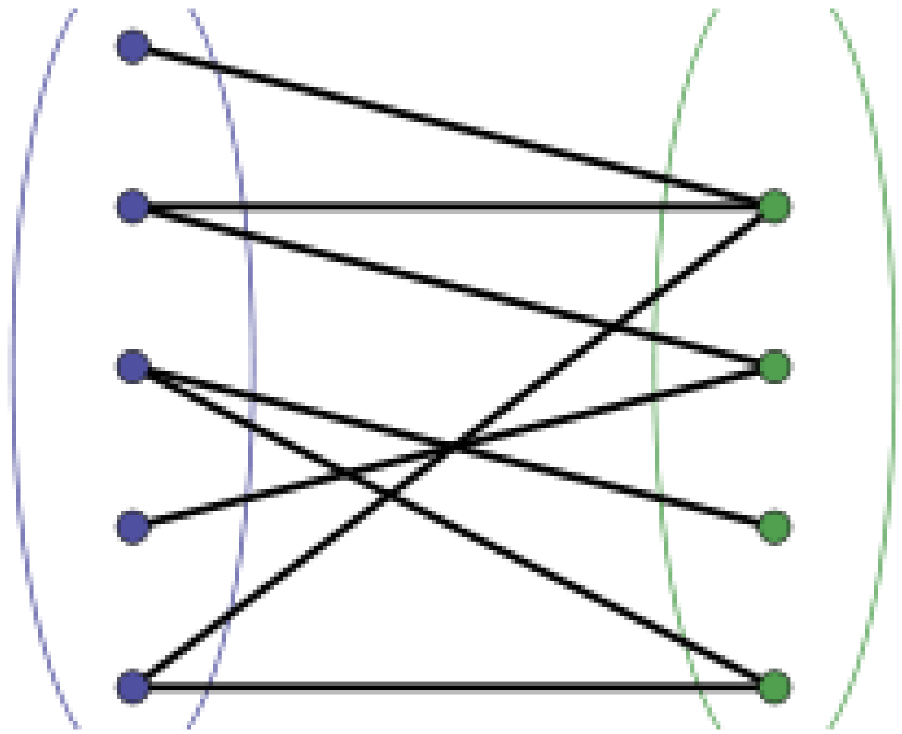
GRAPH PROPERTY TESTING

Image from: Oded Goldreich. A Brief Introduction to Property Testing. Manuscript, 2010.

Graph Property Testing

- For graphs G and H on the same vertex set,
$$d(G,H) = \frac{|E(G)\Delta E(H)|}{|V(G)|^2}$$

- For a property P (i.e. a class of graphs),
$$d(G,P) = \min\{d(G,H) : H \in P\}$$



Bipartite graph

Graph Property Testing

- A property P (e.g., being bipartite)
- The algorithm (called "tester") is allowed to ask queries of the following type:
  - Are vertices u and v adjacent?
- The algorithm should distinguish the cases:
  - G is in P
  - G has distance > ε from all graphs in P

Graph Property Testing

- A property tester for property P:
  1. A randomized decision algorithm
  2. Is given n and $\varepsilon > 0$
  3. Asks q queries
  4. If G has P accepts with probability > 2/3
  5. If $d(G,P) > \varepsilon$ rejects with probability > 2/3

Important parameter: q = query complexity

Graph Property Testing

- Theorem (Alon-Krivelevich 2002). Query complexity of testing bipartiteness is O($1/\varepsilon^2$)
- Algorithm works by "sampling"
- The analysis uses the probabilistic method

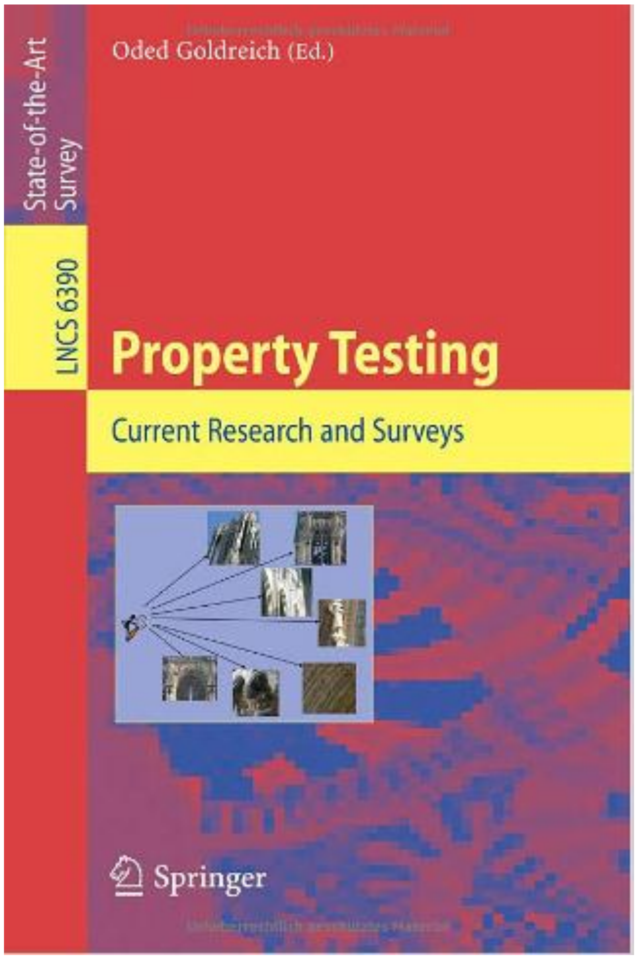
Graph Property Testing

- A graph property is hereditary if it is closed under removal of vertices
  - k-colourable graphs
  - Planar graphs
  - Chordal graphs
- Theorem (Alon, Shapira 2005). Every hereditary graph property is testable with query complexity independent of n

Graph Property Testing

- Which graph properties have query complexity polynomial in 1/ ε ?
  - being k-colourable, for fixed k
  - Being “induced P_3”-free
- Which don’t?
  - Being triangle-free

Graph Property Testing: further reading



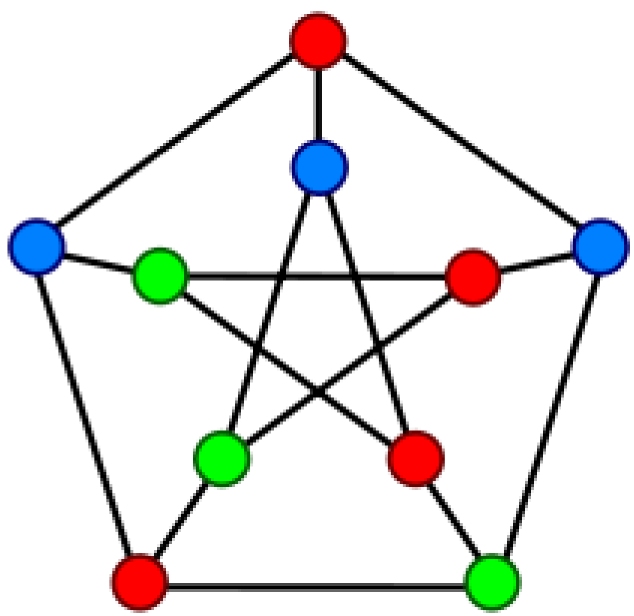
- Alon and Shapira, Homomorphisms in Graph Property Testing, 2006. http://people.math.gatech.edu/~asafico/nesetril.pdf

STRONG PERFECT GRAPH THEOREM

Perfect graphs

- For any graph G

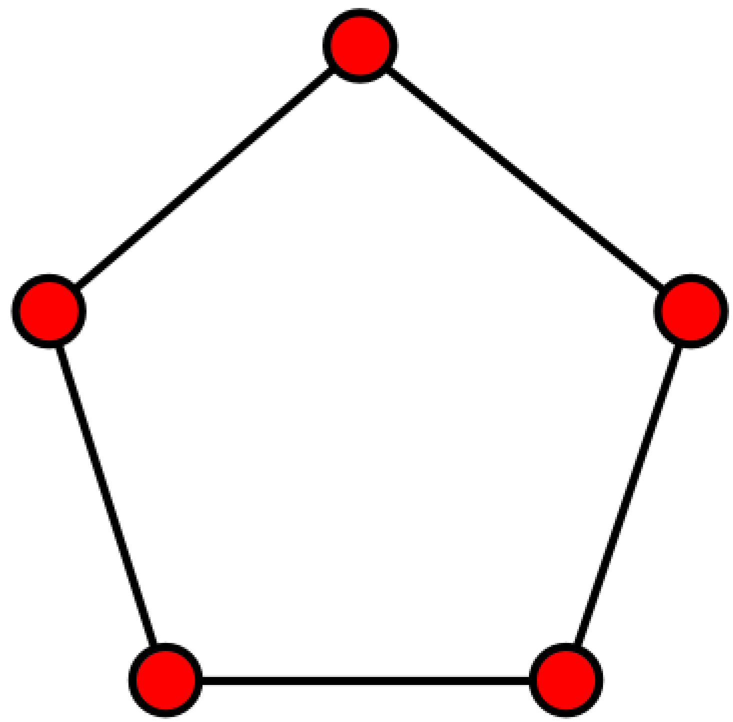
  Clique size of G $\leq$ colouring number of G



Perfect Graphs

- Graph G is perfect if for any induced subgraph H,

  Clique size of H = colouring number of H



Berge's conjectures

- Berge 1961:
- Weak perfect graph conjecture: A graph is perfect if its complement is perfect.
  - Proved by Lovász 1972
- Strong perfect graph conjecture: A graph is perfect if it does not contain an odd$^{>3}$ cycle or its complement as an induced subgraph

Annals of Mathematics, **164** (2006), 51–229

The strong perfect graph theorem

By Maria Chudnovsky, Neil Robertson,* Paul Seymour,**
and Robin Thomas***

Strong Perfect Graph Theorem: further reading

- Seymour, How the proof of the strong perfect graph conjecture was found, 2003 (informal report)
  http://users.encs.concordia.ca/~chvatal/perfect/pds.pdf

WEAK 3-FLOW CONJECTURE

Z_3-flows

- A Z_3-flow for an undirected graph G is an orientation of edges so that for each vertex, number of incoming edges minus number of outgoing edges is divisible by 3.

- Tutte showed: Graph G has a Z_3-flow if and only if it has a nowhere-zero 3-flow

Tutte's 3-flow conjecture

- Tutte (1950's) conjectured: Every 4-edge-connected graph has a Z_3-flow.
- Jaeger (1988) conjectured: there exists a k such that every k-edge-connected graph has a Z_3-flow.
- Thomassen (2012) proved for $k = 8$.
- Recently improved to $k = 6$.

Tutte's 3-flow conjecture: further reading

- Laszlo Miklos Lovasz, Tutte's flow conjectures, 2012
  http://tlovering.files.wordpress.com/2012/06/laszloessay.pdf

Sources for Pictures

- Fan Chung’s homepage: http://math.ucsd.edu/~fan/
- Chung and Lu. Complex Graphs and Networks. AMS. 2006
- Wikipedia
- Bonato. A Course on the Web Graph. AMS. 2008
- Lovász. Large Networks and Graph Limits. AMS. 2012
- Chung, Lu, Dewey, and Galas. Duplication Models for Biological Networks. Journal of Computational Biology. 10 (5), 2003, 677--687
- Goldreich. A Brief Introduction to Property Testing. Manuscript, 2010
- http://networkx.github.com/
- http://www.bordalierinstitute.com/

Thank you !