Three popular science books, a dozen articles in *Science* and *Nature*, and 154 preprints at `xxx.arXiv.org/cond-mat` deal with *complex networks*, which in this context means the empirical and theoretical study of large graphs, focusing in particular on those possessing the following three qualitative properties, asserted to hold in many interesting real-world examples.

- the degree distribution has power-law tail
- local clustering of edges: graph is not locally tree-like
- small diameter – $O(\log(\text{number of vertices}))$.

The nature of that subject – typically not presented as rigorous mathematics – is most easily seen from the long survey papers

R. Albert and A.-L. Barabási, *Statistical mechanics of complex networks*, Rev. Mod. Phys. **74** (2002), 47–97.

S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of networks*, Adv. Phys. **51** (2002), 1079–1187.

M.E.J. Newman, *The structure and function of complex networks*, SIAM Review **45** (2003), 167–256.

and monograph

J.F.F. Mendes and S.N. Dorogovtsev, *Evolution of networks: From biological nets to the internet and WWW*, Oxford Univ. Press, 2003.

A shorter survey emphasizes rigorous mathematical results.

B. Bollobás and O. Riordan, *Mathematical results on scale-free random graphs*, Handbook of Graphs and Networks (S. Bornholdt and H.G. Schuster, eds.), Wiley, 2002.

For other mathematicians' views see notes on web by
Rick Durrett (Fall 2004)
David Aldous (Spring 2003).

Two basic models with numerous variants.

Lattice small worlds. Vertices of $\mathbb{Z}^d$ connected to $K \geq 2d$ near neighbors. Add random longer edges (v, w) with probability $a|v-w|^{-\beta}$.

Preferential attachment. Grow a directed graph. An arriving vertex receives K out-edges, from vertices v chosen randomly with probabilities proportional to $a + b \times$ in-degree(v).

What has been studied in such models?

- graph statistics: distribution of vertex degree; diameter; frequency of triangles and small subgraphs
- spread of epidemics etc
- decentralized routing schemes.

Illustrate using 4 recent papers.

Social network: vertex = person, edge = some specified relationship
- friend
- on board of same corporation
- sexual contact.

Collaboration network: edge is "co-author on paper". Lots of data available! Several attempts to fit probability models, but no convincing "good model".

Jon Kleinberg and D. Liben-Nowell: The Link Prediction Problem for Social Networks.

Idea: try to predict future collaborations.

Seek a predictor function $f_G(v, w)$ where v, w are vertices of G. Predict the next 1000 collaborations will be between the 1000 pairs (v, w) with maximum values of $f_G(v, w)$ in the current graph G.

Simplest such function is *graph distance*. Diagram compares this to 10 other predictors. Turned out best predictor was

$$f(v, w) = \sum_{z \in \mathcal{N}(v) \cap \mathcal{N}(w)} \frac{1}{\log |\mathcal{N}(z)|}$$

where $\mathcal{N}(v)$ = set of neighbors of v.

Suggests math problem: is there any probability model for which this is best predictor (not by fiat)?

P. Echenique et al: Dynamics of Jamming Transitions in Complex Networks.

• Packets generated at rate p, random source and destination.
• a vertex has $\{0, 1, 2, \ldots\}$ packets; can transmit first-in-queue packet to neighbor vertex in one time unit.
• Choose the neighbor i to minimize

$$hd(i, \text{destination}) + (1 - h)(\text{queue-length at } i)$$

for a parameter h we choose.

Paper takes a real-world (Internet Autonomous System) network and simulates this model. The diagram shows proportion of packets **not** delivered as function of attempted traffic p for different values of h.

Given a sparse graph, seek to identify **communities**, i.e. comparatively dense subgraphs. Many methods have been considered.

M. Latapy and P. Pons: Computing Communities in Large Networks using Random Walks

(a) Use standard *hierarchical clustering* scheme. Given a distance $D(A, B)$ between disjoint vertex subsets, start with the partition into singleton vertices. At each step merge some pair $\{A, B\}$ into $A \cup B$, where the pair is chosen to minimize

$$\text{ave}_{v \in A \cup B} D(\{v\}, A \cup B).$$

(b) Define $D(\cdot, \cdot)$ using L^2 distance for random walks. Let X_t^A be RW started uniformly on A;

$$D^2(A, B) = \sum_v (P(X_t^A = v) - P(X_t^B = v)^2 / \text{deg}(v)$$

for $t = \text{diam}(A) + \text{diam}(B)$, say.