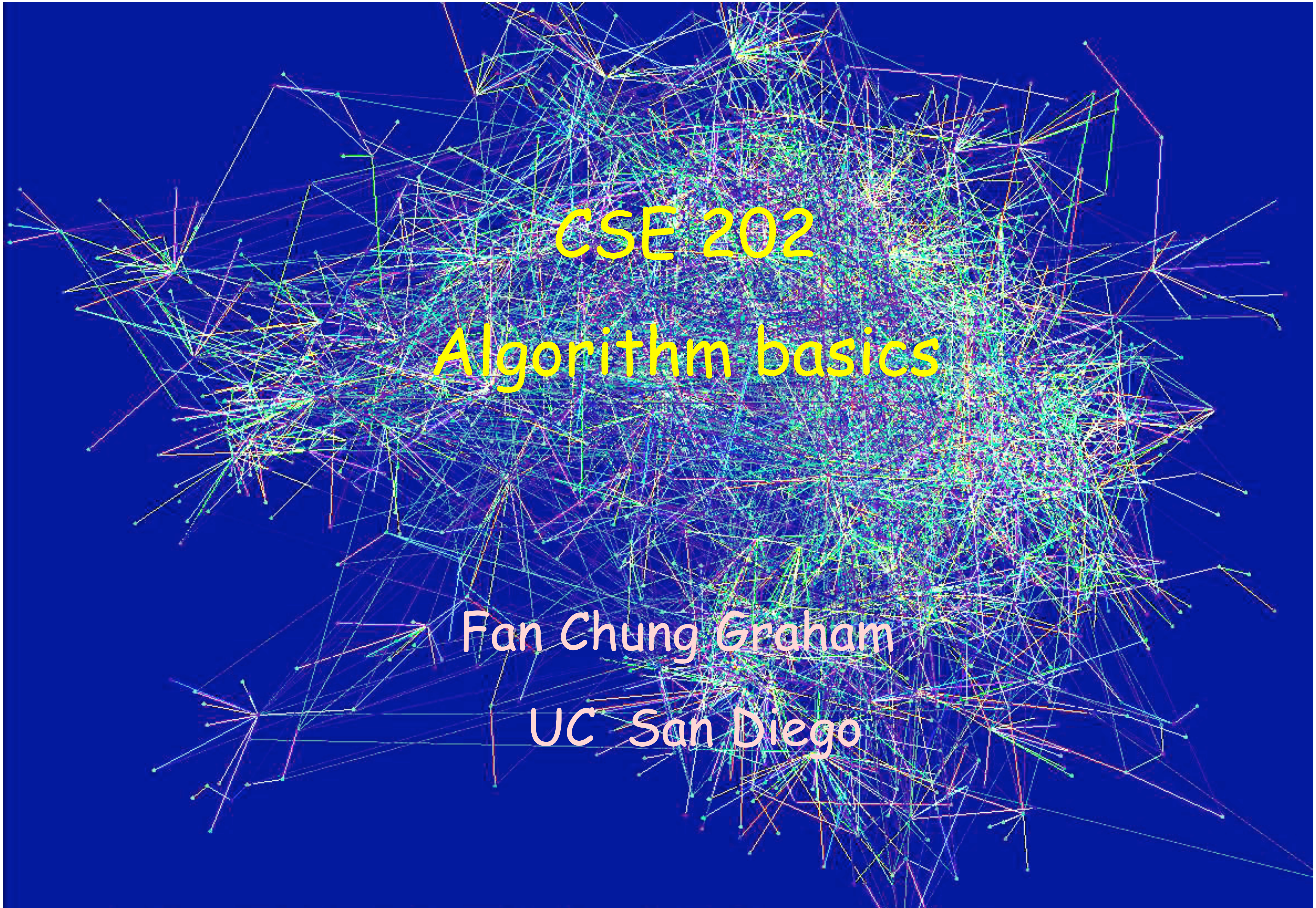
CSE 202

Algorithm basics

Fan Chung Graham

UC San Diego

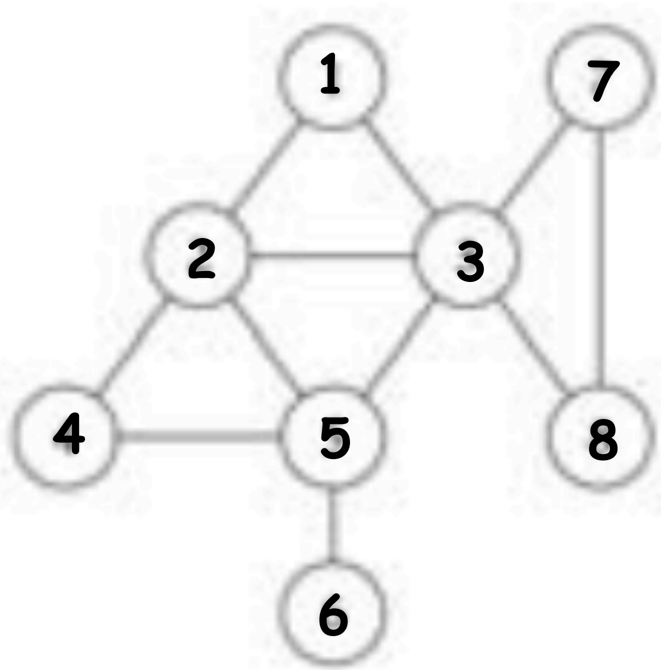


An induced subgraph of the collaboration graph (with Erdos number at most 2).
Made by Fan Chung Graham and Lincoln Lu in 2002.

Undirected Graphs

Undirected graph. $G = (V, E)$

- V = nodes.
- E = edges between pairs of nodes.
- Captures pairwise relationship between objects.
- Graph size parameters: $n = |V|$, $m = |E|$.



$V = \{ 1, 2, 3, 4, 5, 6, 7, 8 \}$

$E = \{ 1\text{-}2, 1\text{-}3, 2\text{-}3, 2\text{-}4, 2\text{-}5, 3\text{-}5, 3\text{-}7, 3\text{-}8, 4\text{-}5, 5\text{-}6 \}$

$n = 8$

$m = 11$

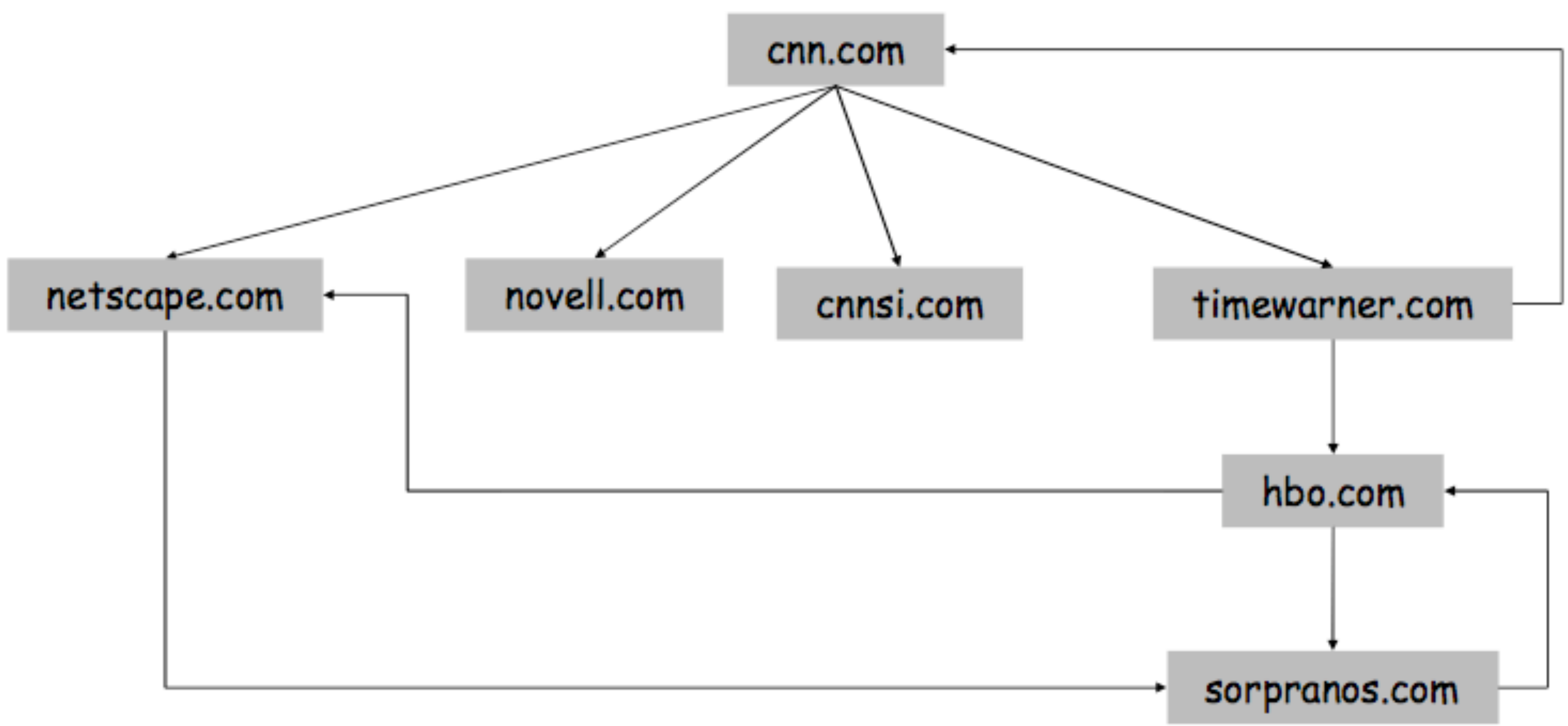
Some Graph Applications

Graph	*Nodes*	*Edges*
transportation	street intersections	highways
communication	computers	fiber optic cables
World Wide Web	web pages	hyperlinks
social	people	relationships
food web	species	predator-prey
software systems	functions	function calls
scheduling	tasks	precedence constraints
circuits	gates	wires

World Wide Web

Web graph.

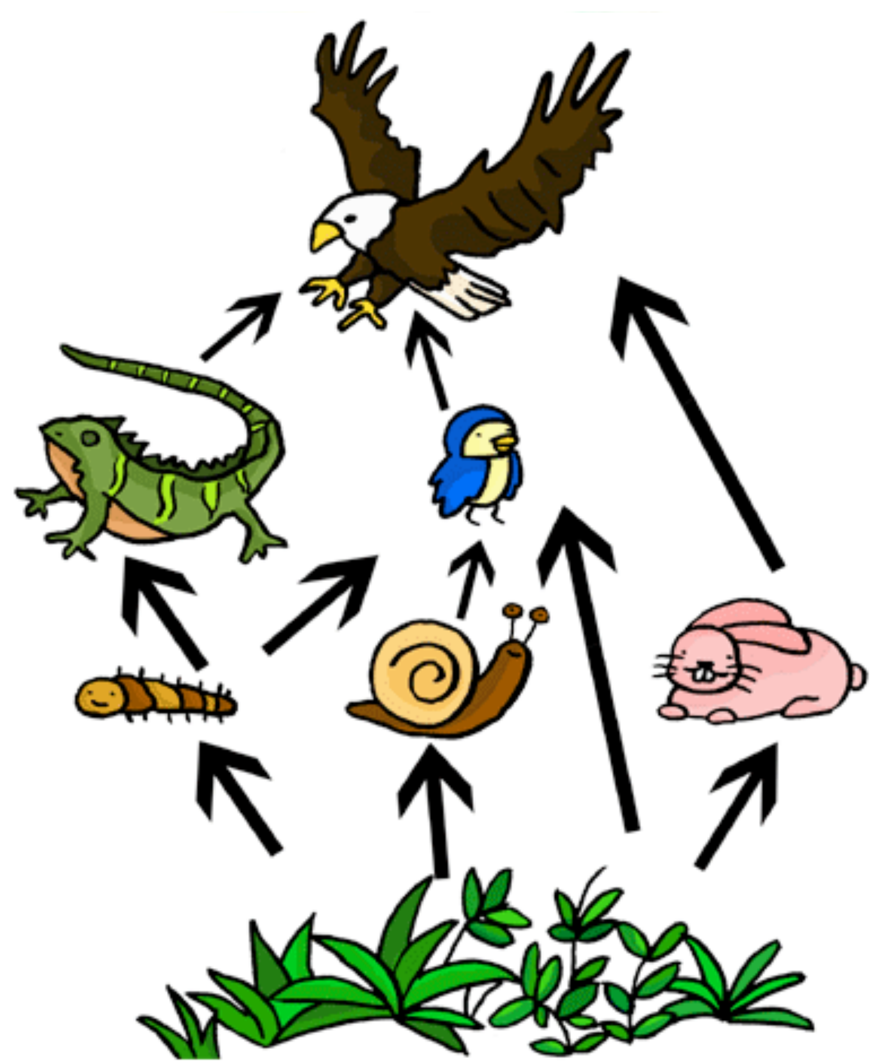
- Node: web page.
- Edge: hyperlink from one page to another.



Ecological Food Web

Food web graph.

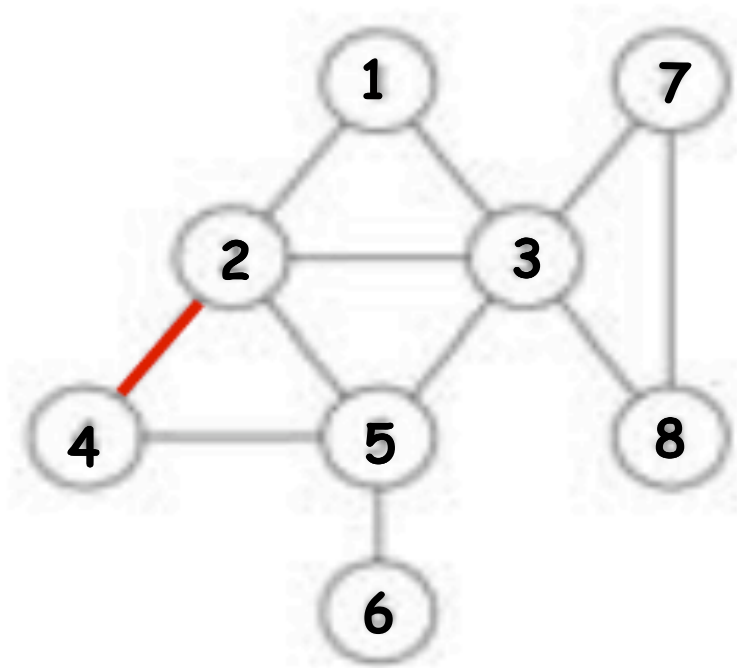
- Node = species.
- Edge = from prey to predator.



Graph Representation: Adjacency Matrix

Adjacency matrix. n-by-n matrix with $A_{uv} = 1$ if (u, v) is an edge.

- Two representations of each edge.
- Space proportional to n^2.
- Checking if (u, v) is an edge takes $\Theta(1)$ time.
- Identifying all edges takes $\Theta(n^2)$ time.



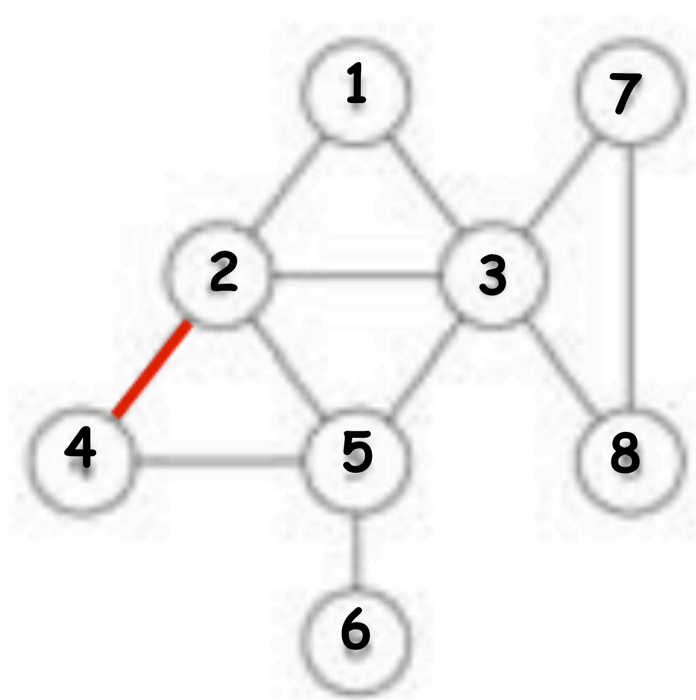
	1	2	3	4	5	6	7	8
1	0	1	1	0	0	0	0	0
2	1	0	1	1	1	0	0	0
3	1	1	0	0	1	0	1	1
4	0	1	0	0	1	0	0	0
5	0	1	1	1	0	1	0	0
6	0	0	0	0	1	0	0	0
7	0	0	1	0	0	0	0	1
8	0	0	1	0	0	0	1	0

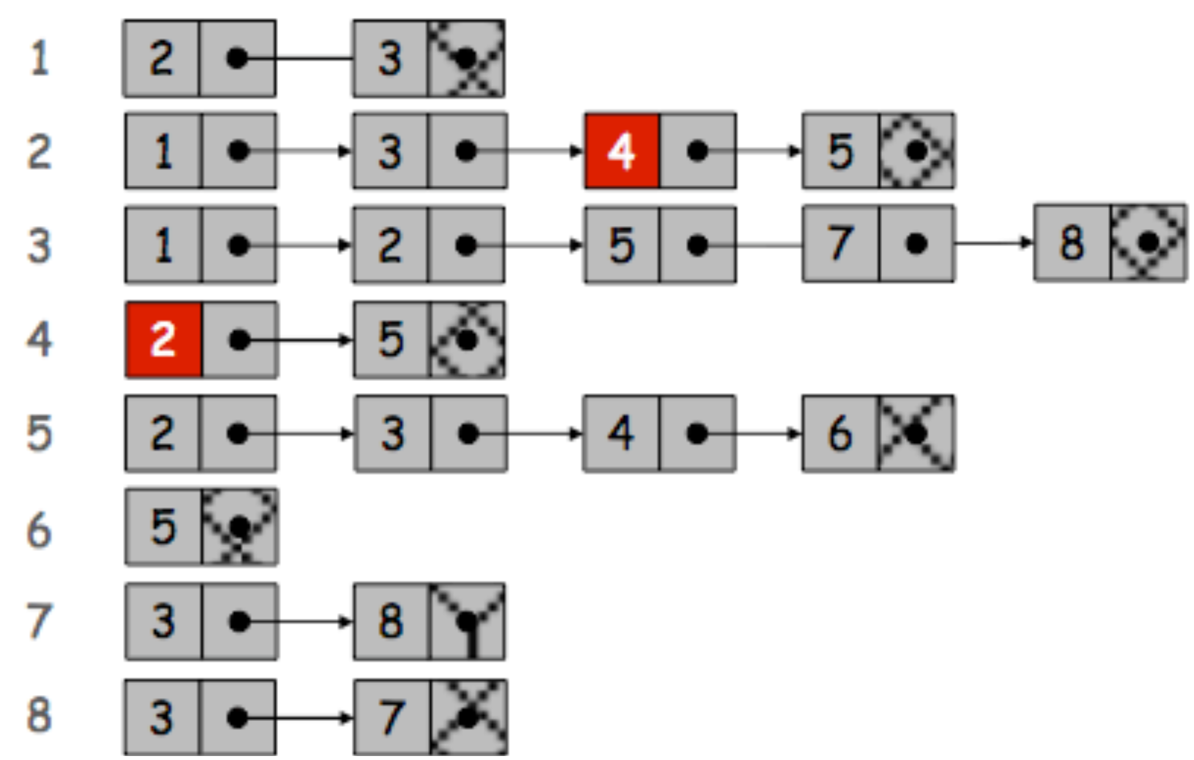
Graph Representation: Adjacency List

Adjacency list. Node indexed array of lists.

- Two representations of each edge.
- Space proportional to m + n.
- Checking if (u, v) is an edge takes O(deg(u)) time.
- Identifying all edges takes Θ(m + n) time.

degree = number of neighbors of u



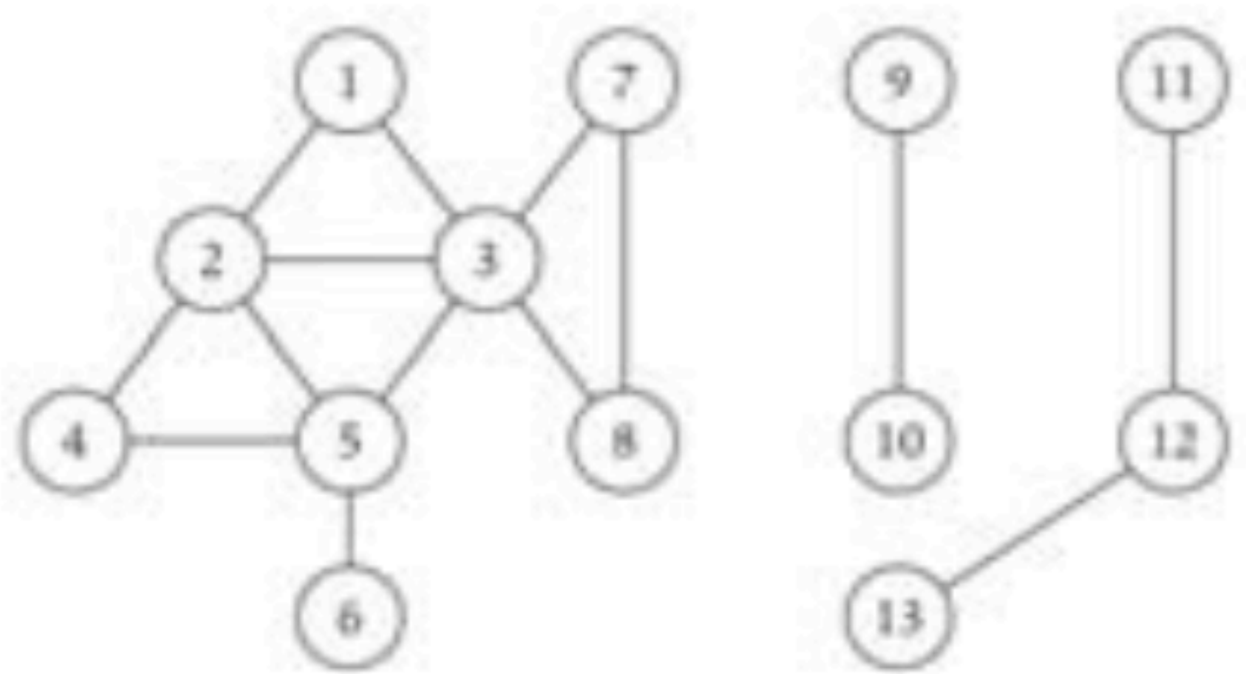


Paths and Connectivity

Def. A path in an undirected graph G = (V, E) is a sequence P of nodes $v_1, v_2, \ldots, v_{k-1}, v_k$ with the property that each consecutive pair v_i, v_{i+1} is joined by an edge in E.

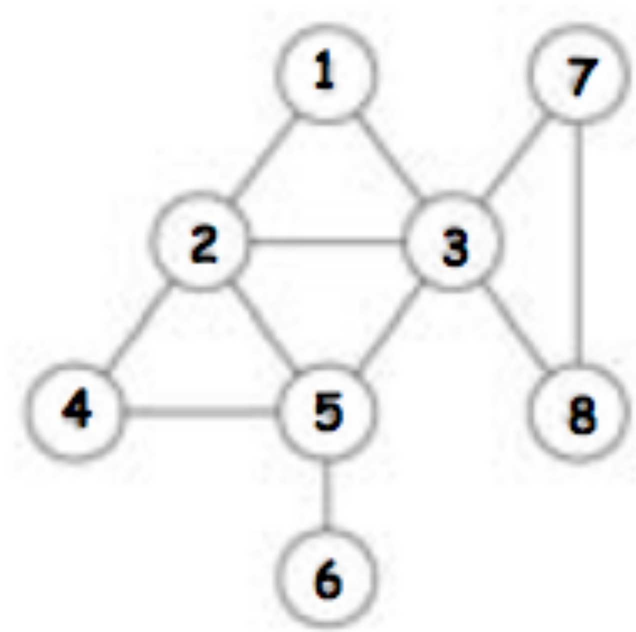
Def. A path is simple if all nodes are distinct.

Def. An undirected graph is connected if for every pair of nodes u and v, there is a path between u and v.



Cycles

Def. A cycle is a path $v_1, v_2, \ldots, v_{k-1}, v_k$ in which $v_1 = v_k$, $k > 2$, and th
first k-1 nodes are all distinct.



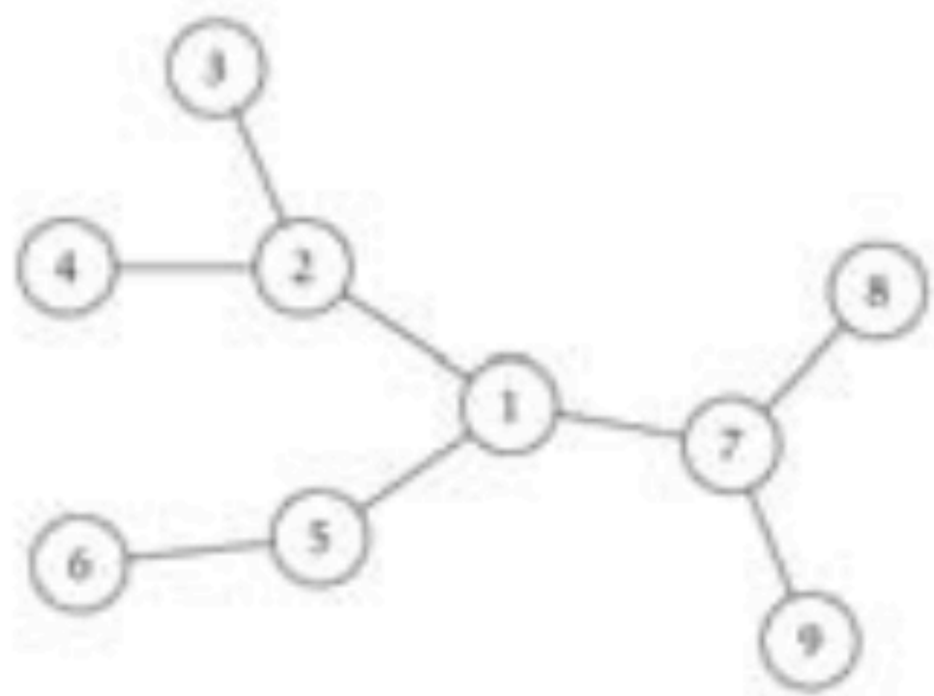
cycle C = 1-2-4-5-3-1

Trees

Def. An undirected graph is a tree if it is connected and does not contain a cycle.

Theorem. Let G be an undirected graph on n nodes. Any two of the following statements imply the third.

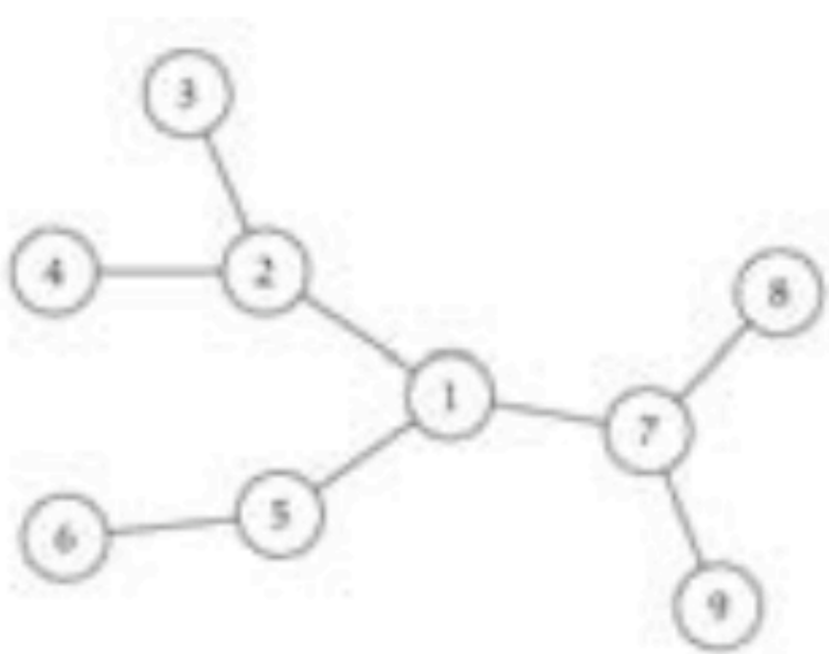
- G is connected.
- G does not contain a cycle.
- G has n-1 edges.



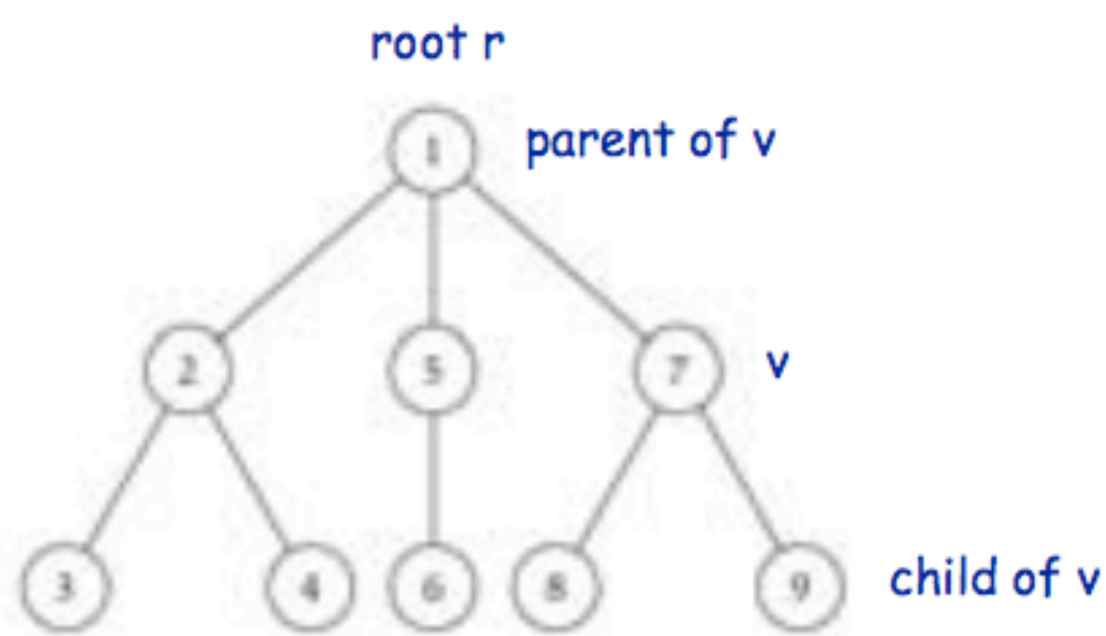
Rooted Trees

Rooted tree. Given a tree T, choose a root node r and orient each edge away from r.

Importance. Models hierarchical structure.



a tree



the same tree, rooted at 1

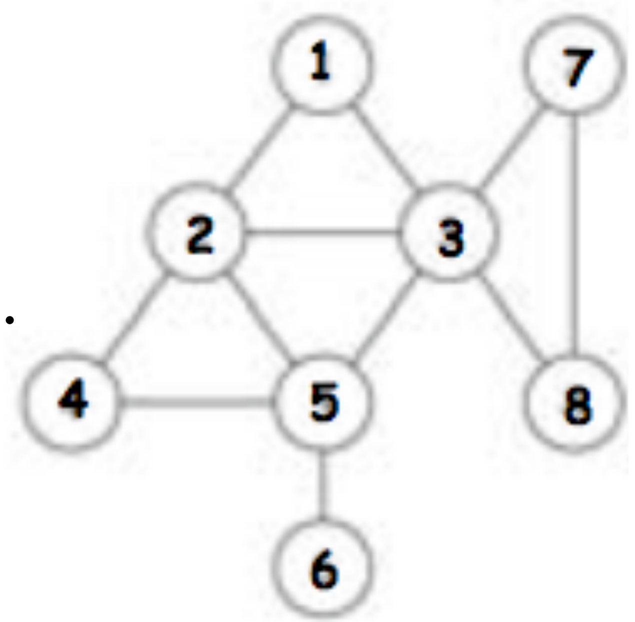
Connectivity

s-t connectivity problem. Given two node s and t, is there a path between s and t?

s-t shortest path problem. Given two node s and t, what is the length of the shortest path between s and t?

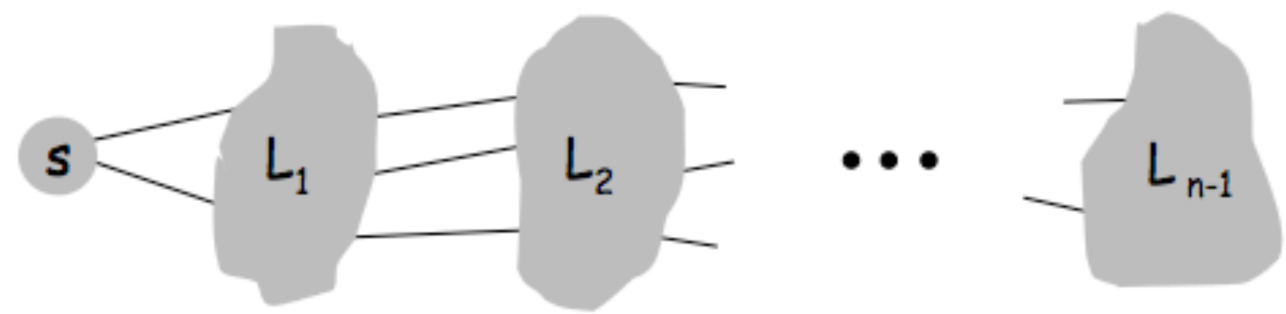
Applications.

- Friendster.
- Maze traversal.
- Kevin Bacon number.
- Fewest number of hops in a communication network.



Breadth First Search

BFS intuition. Explore outward from s in all possible directions, adding nodes one "layer" at a time.



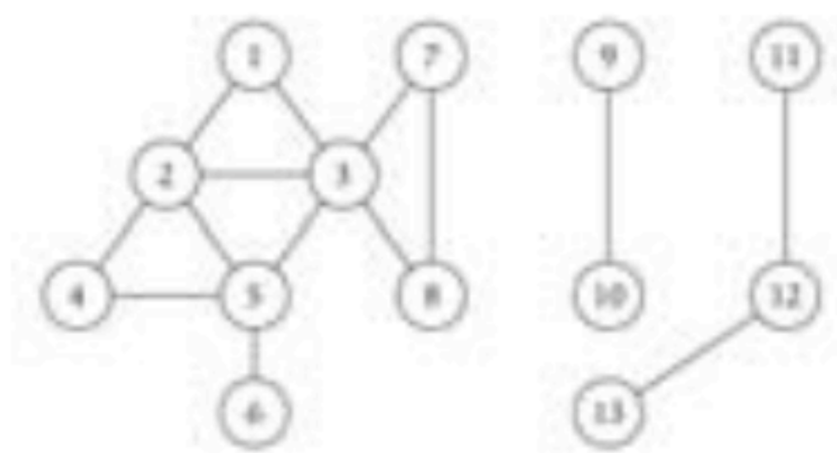
BFS algorithm.

- $L_0 = \{ s \}$.
- L_1 = all neighbors of L_0.
- L_2 = all nodes that do not belong to L_0 or L_1, and that have an edge to a node in L_1.
- L_{i+1} = all nodes that do not belong to an earlier layer, and that have an edge to a node in L_i.

Theorem. For each i, L_i consists of all nodes at distance exactly i from s. There is a path from s to t iff t appears in some layer.

Connected Component

Connected component. Find all nodes reachable from s.



Connected component containing node 1 = { 1, 2, 3, 4, 5, 6, 7, 8 }.

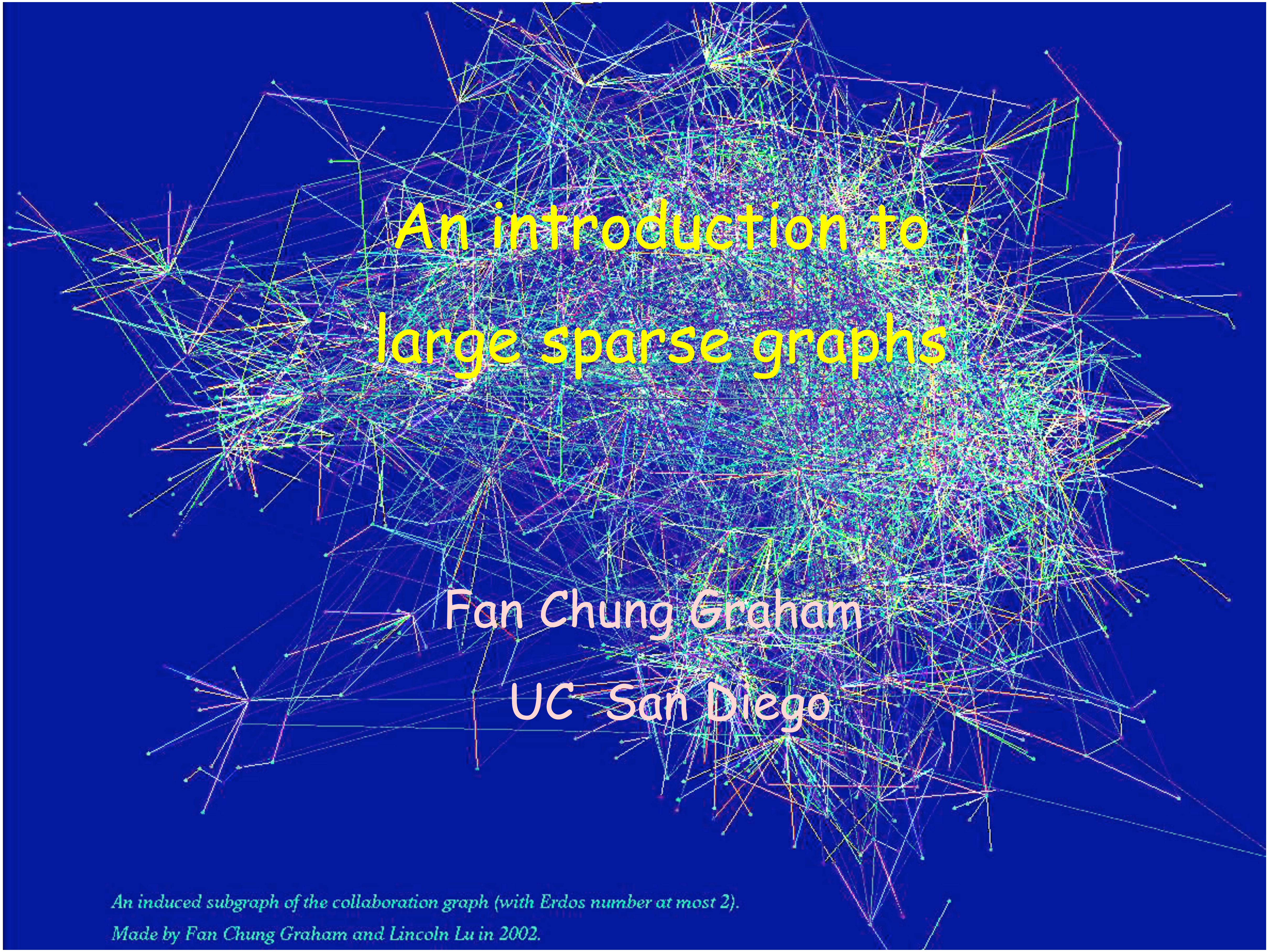
Algorithms: Breadth First Search BFS

Depth First Search DFS

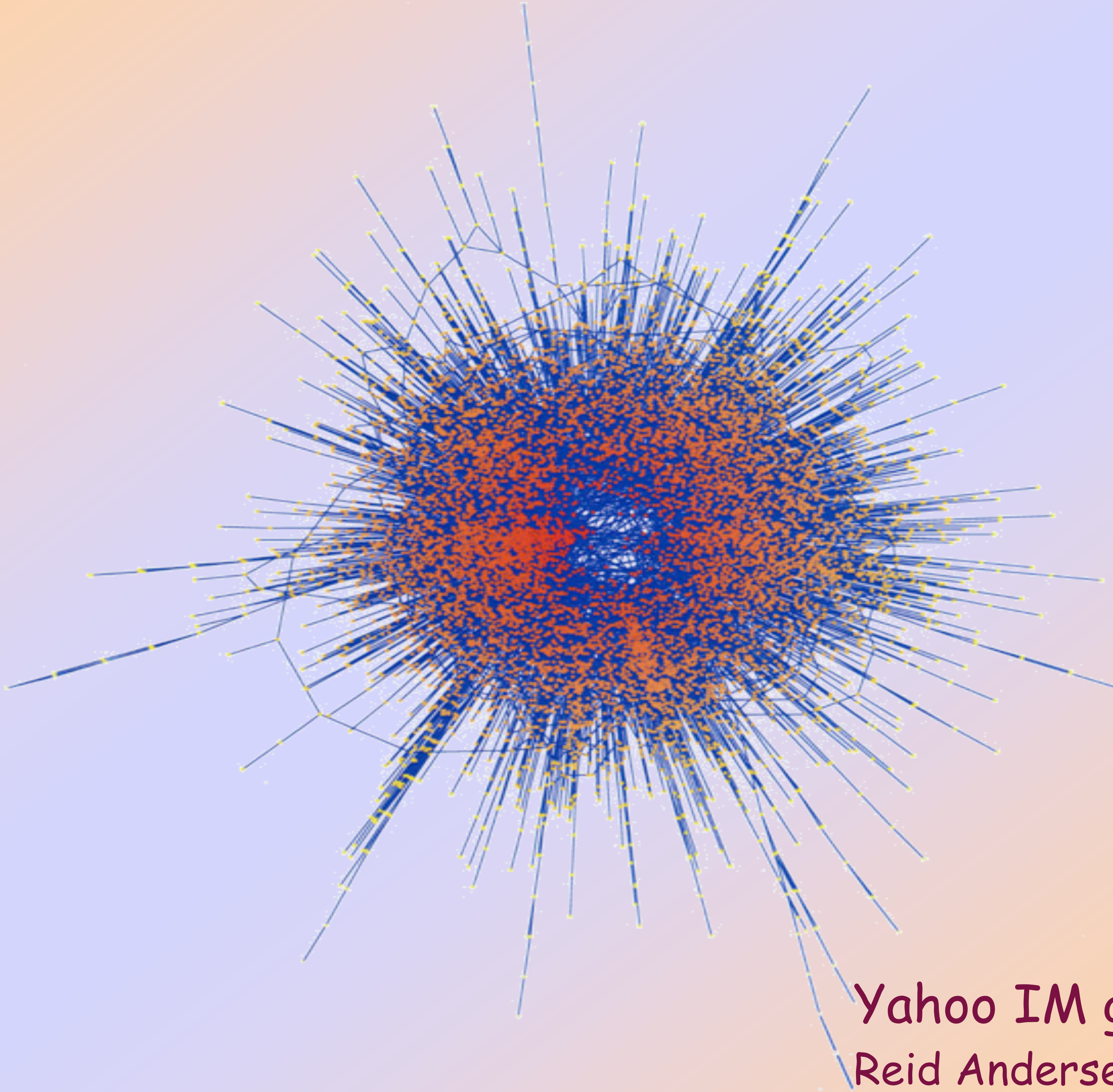
An introduction to large sparse graphs

Fan Chung Graham

UC San Diego

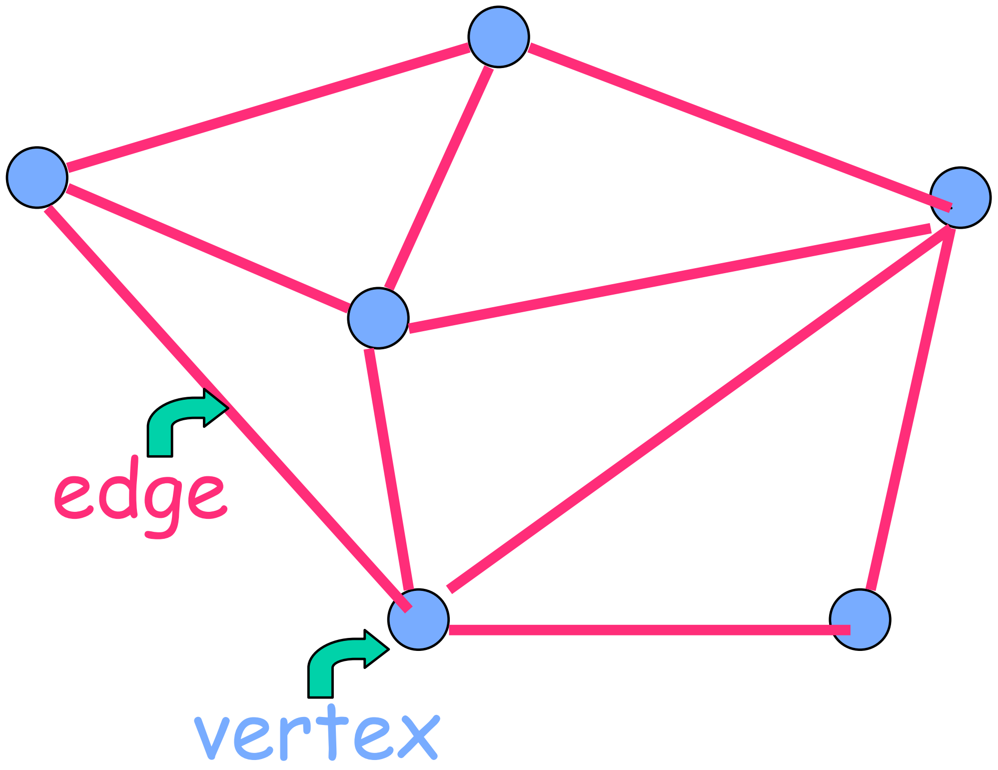


An induced subgraph of the collaboration graph (with Erdos number at most 2).
Made by Fan Chung Graham and Lincoln Lu in 2002.



Yahoo IM graph
Reid Andersen 2005

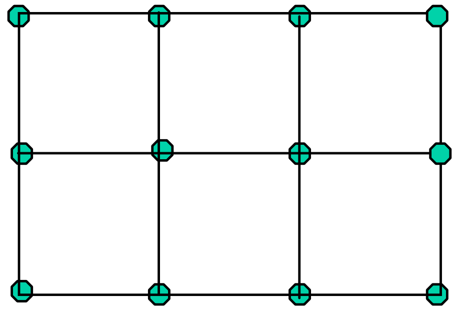
A graph G = (V,E)

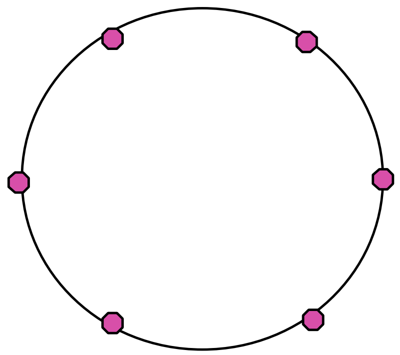


Graph Theory has 250 years of history.

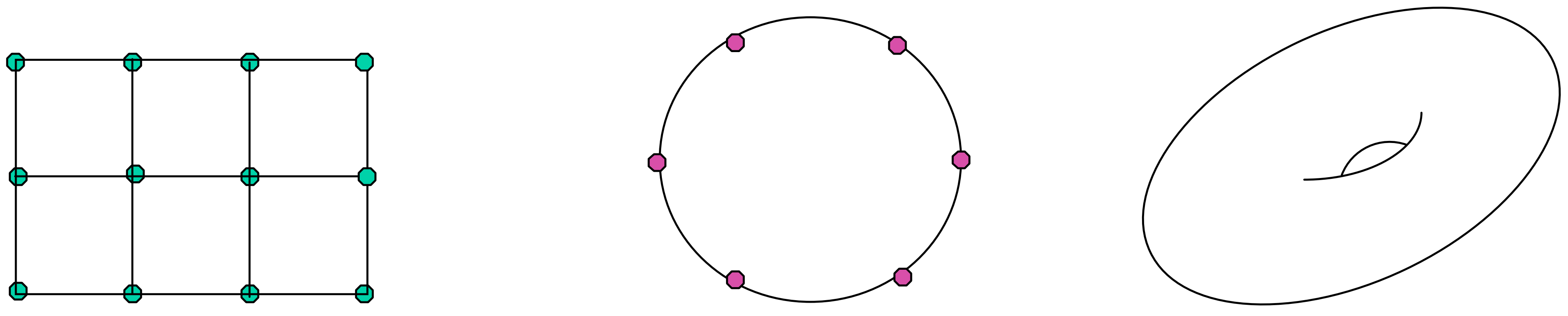
Leonhard Euler, 1707-1783

Geometric graphs

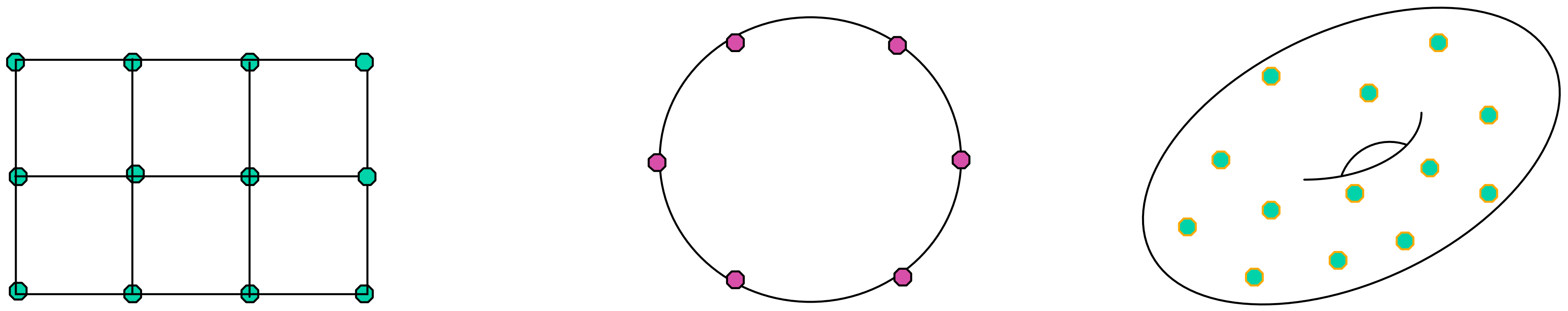




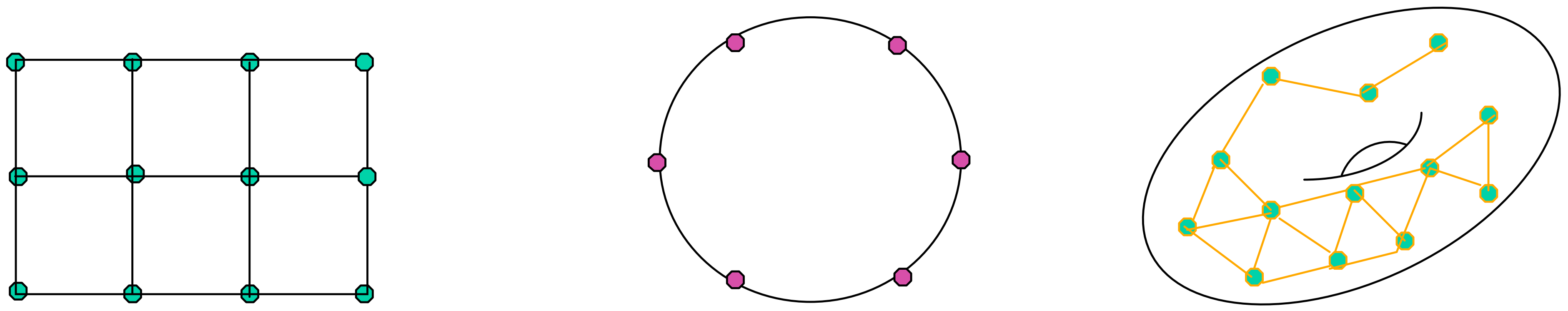
Geometric graphs



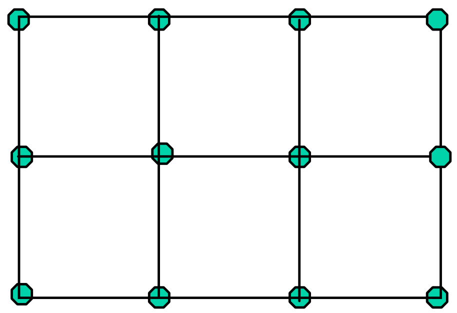
Geometric graphs

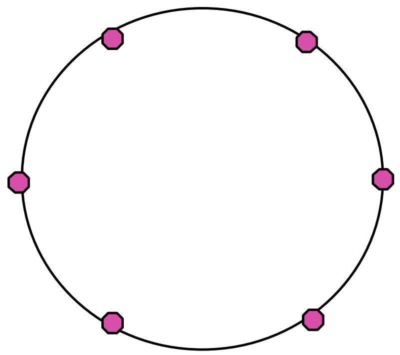


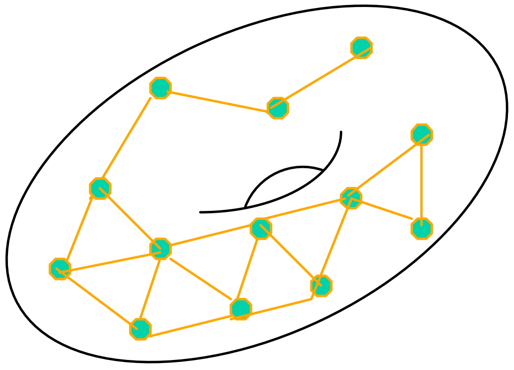
Geometric graphs

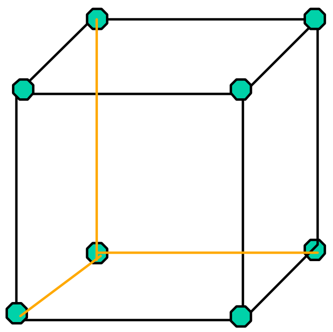


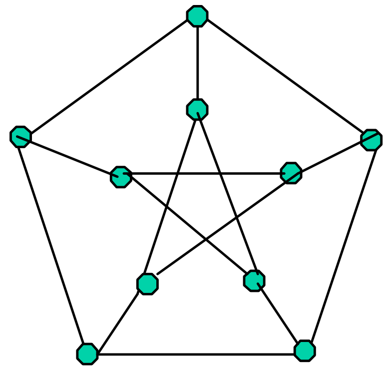
Geometric graphs





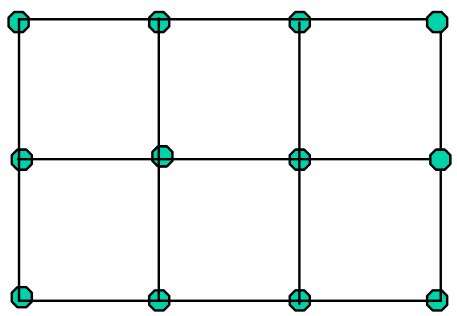


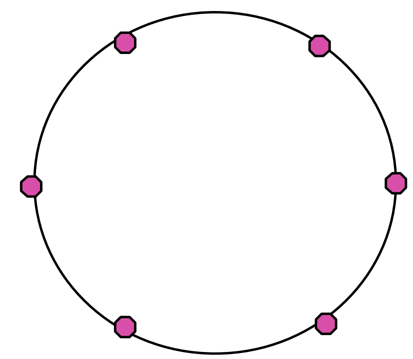


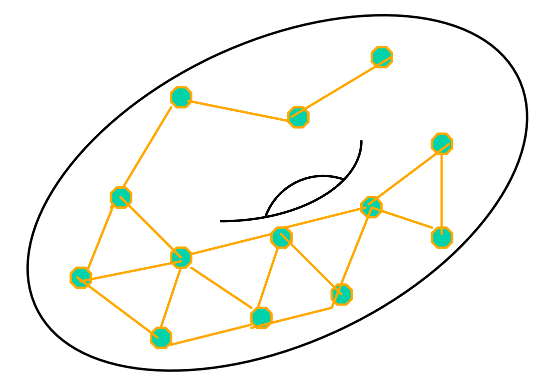


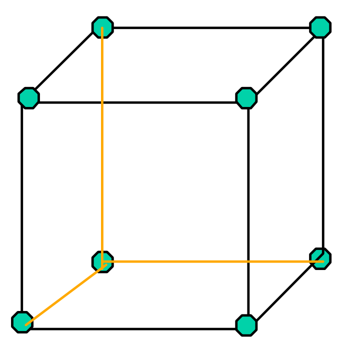
Algebraic graphs

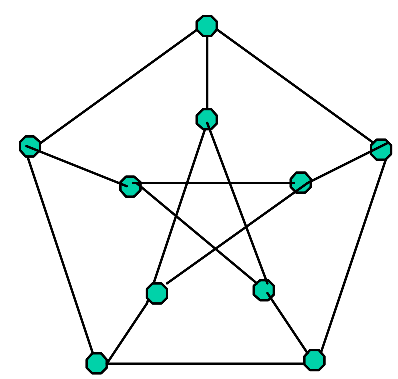
Geometric graphs

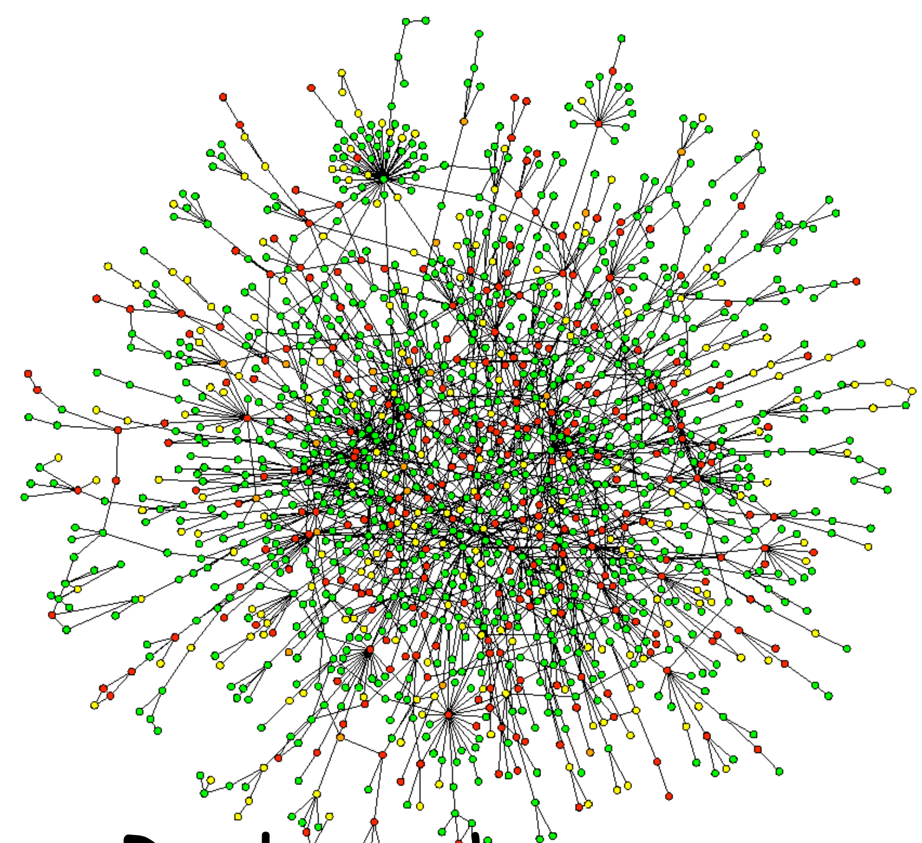






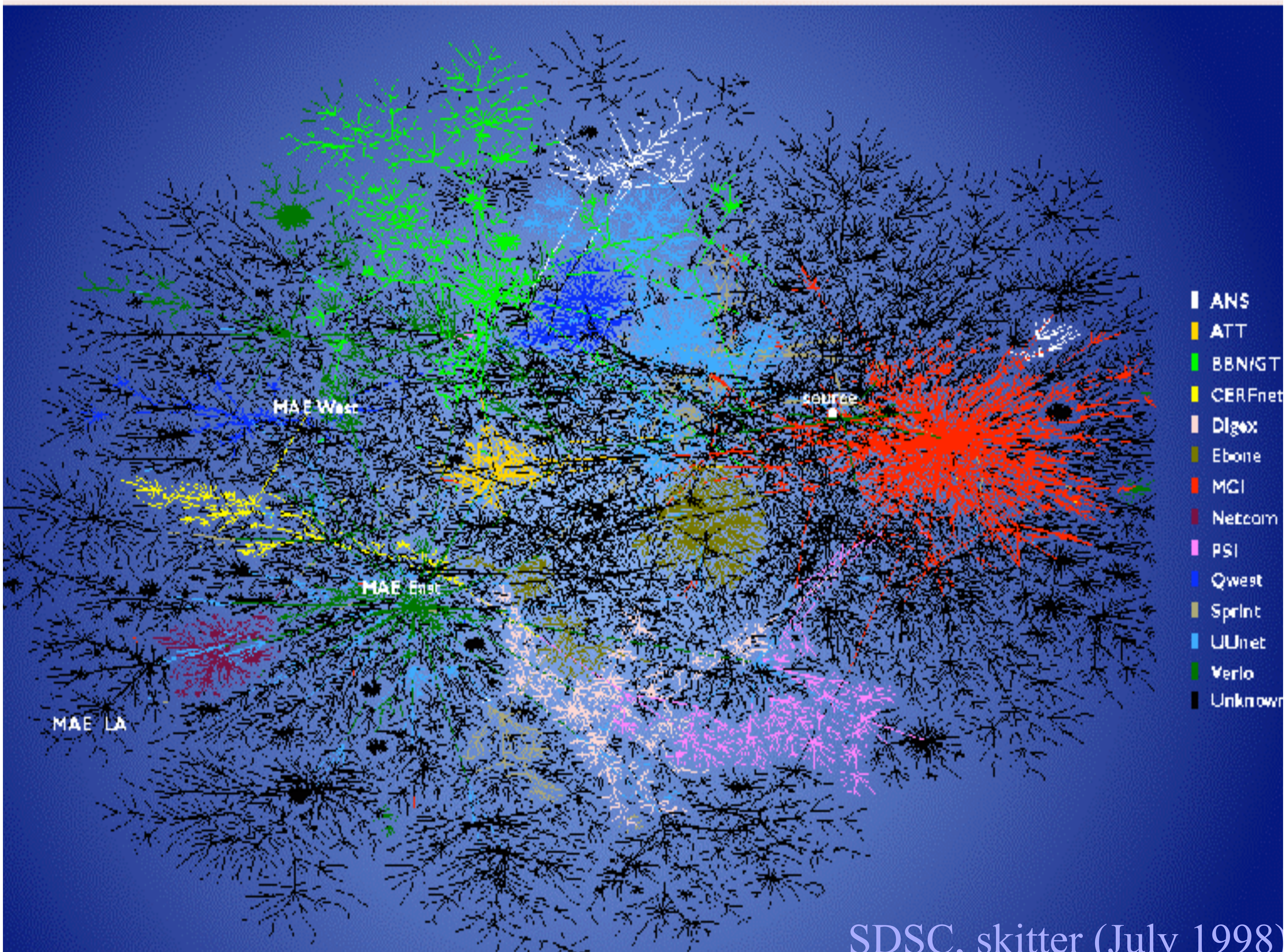






Algebraic graphs

Real graphs
(protein interactions
by Jawoong Jeong)



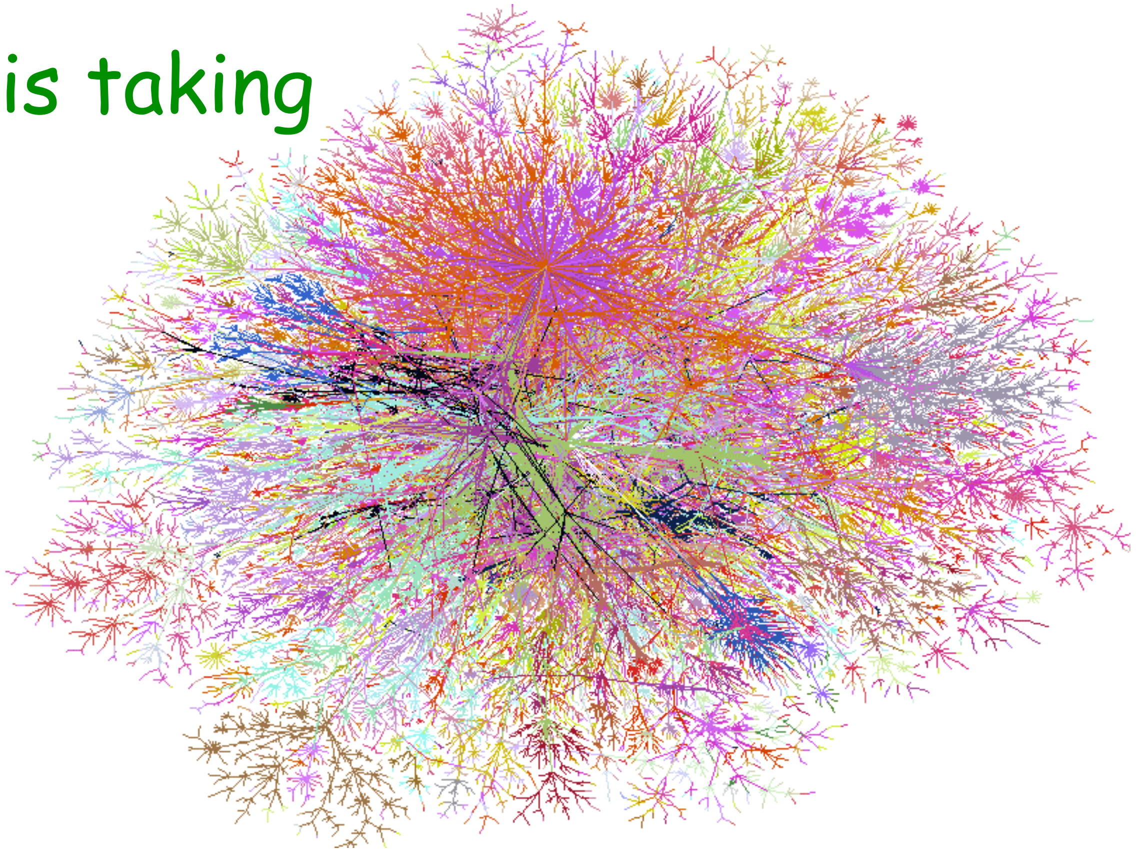
SDSC, skitter (July 1998)

Massive data

Massive graphs

Massive data → Massive graphs

The information we deal with is taking on a networked character.



What does a massive graph look like?

What does a massive graph look like?

- sparse
- clustered
- small diameter

What does a massive graph look like?

sparse

clustered

small diameter

prohibitively large

dynamically changing

incomplete information

What does a massive graph look like?

sparse

clustered

small diameter

prohibitively large

dynamically changing

incomplete information

Hard to describe !

Harder to analyze !!

Some prevailing characteristic of large realistic networks

- Small world phenomenon

    Small diameter/average distance

    Clustering

- Power law degree distribution

A crucial observation

Massive graphs satisfy the power law.

Discovered by several groups independently.

- Barabási, Albert and Jeung, 1999.
- Broder, Kleinberg, Kumar, Raghavan, Rajagopalan and Tomkins, 1999.
- M Faloutsos, P. Faloutsos and C. Faloutsos, 1999.
- Abello, Buchsbaum, Reeds and Westbrook, 1999.
- Aiello, Chung and Lu, 1999.

The history of power law

- Zipf's law, 1949. (The n^{th} most frequent word occurs at rate 1/n)
- Yule's law, 1942. (City population follows a power law.)
- Lotka's law, 1926. (distribution of authors in chemical abstracts) 1907-1916
- Pareto, 1897 (Wealth distribution follows a power law.)

Natural language

Bibliometrics

Social sciences

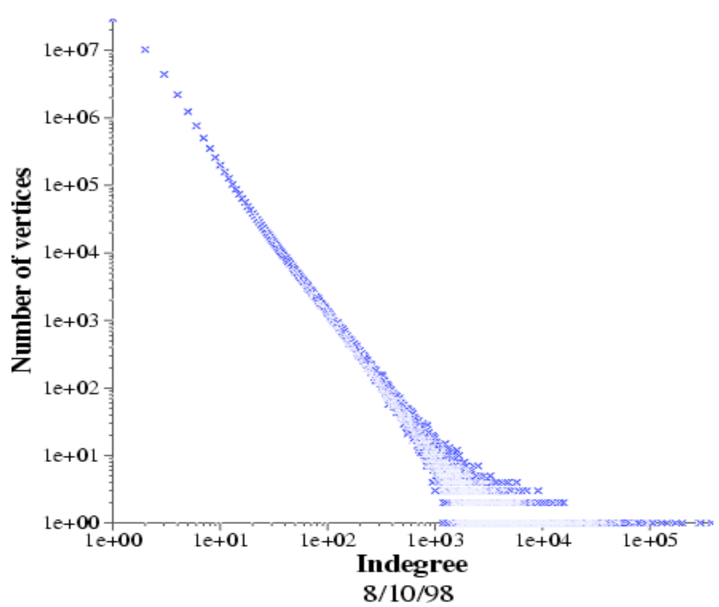
Nature

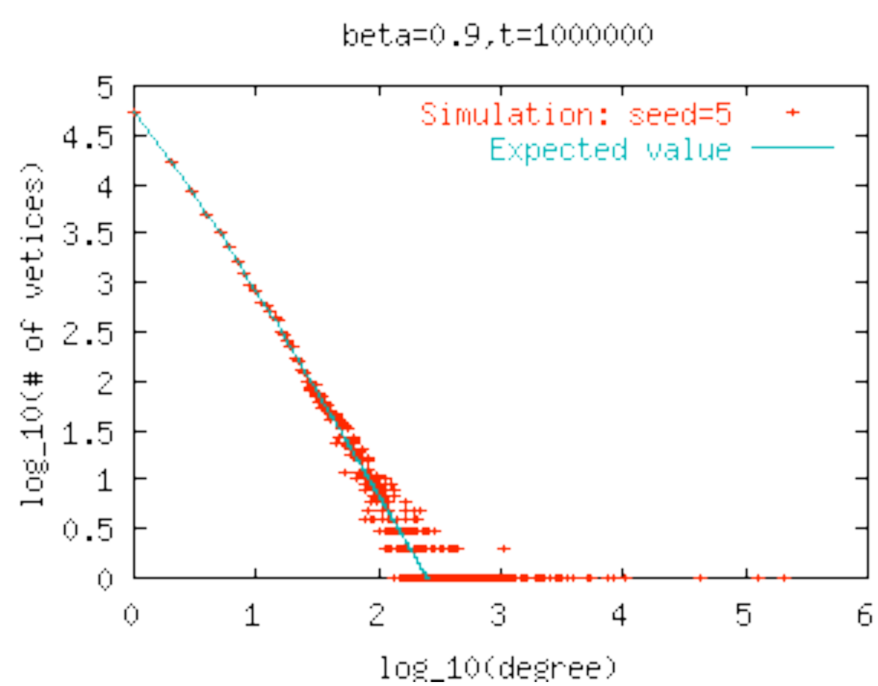
Massive graphs satisfy the power law.

Power decay degree distribution.

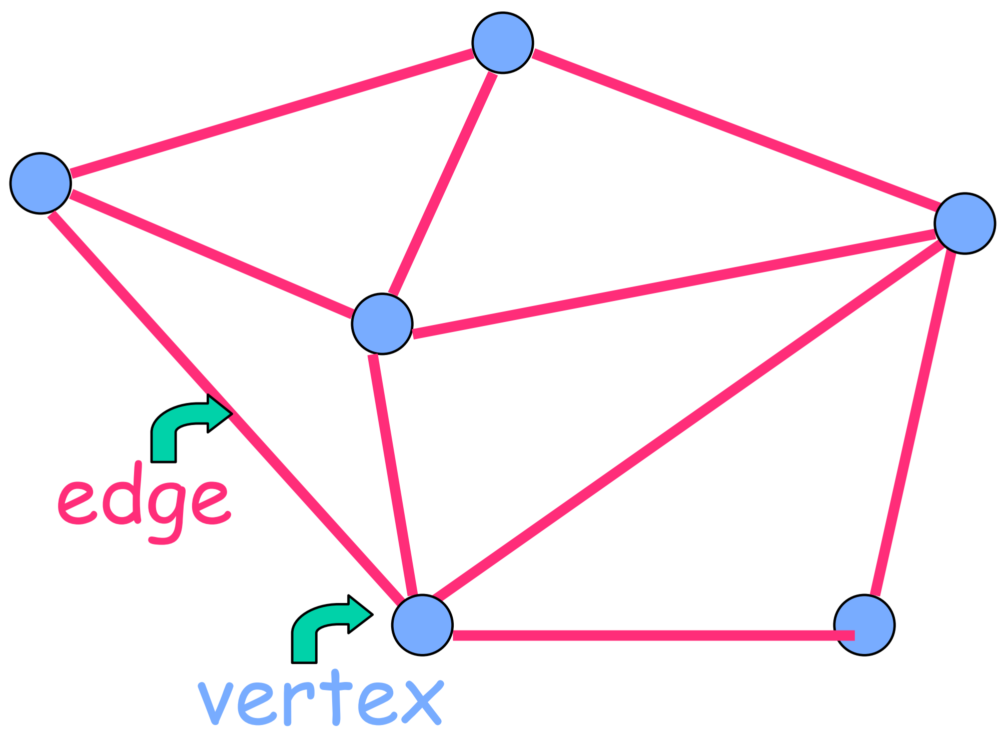
The degree sequences satisfy the power law:

The number of vertices of degree j is proportional to $j^{-\beta}$ where β is some constant ≥ 2.



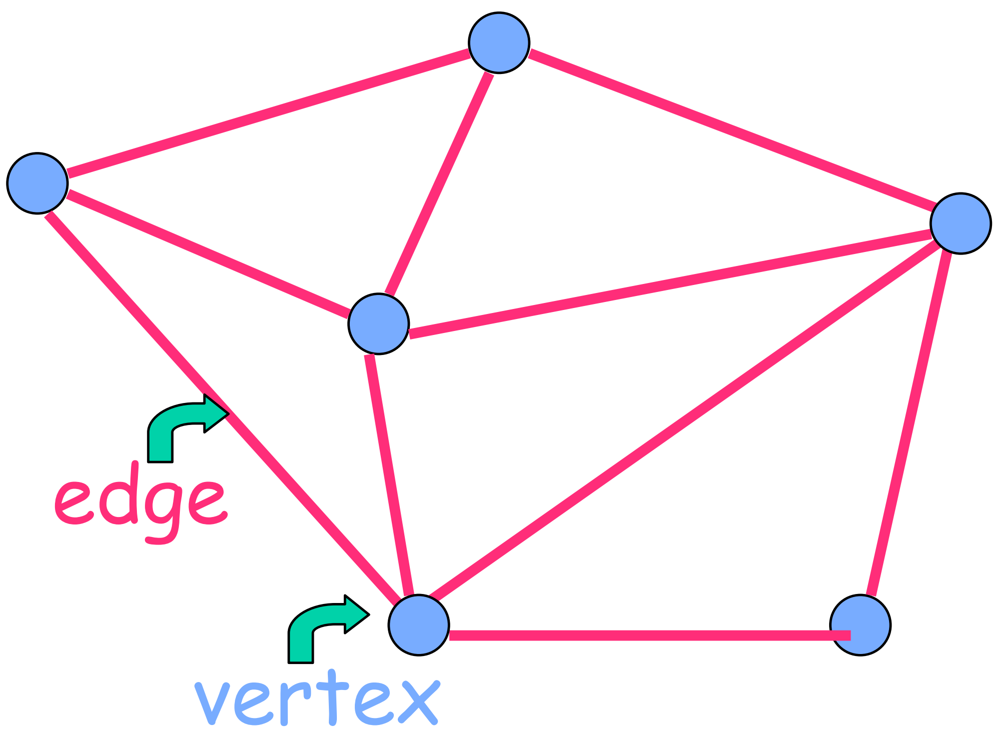


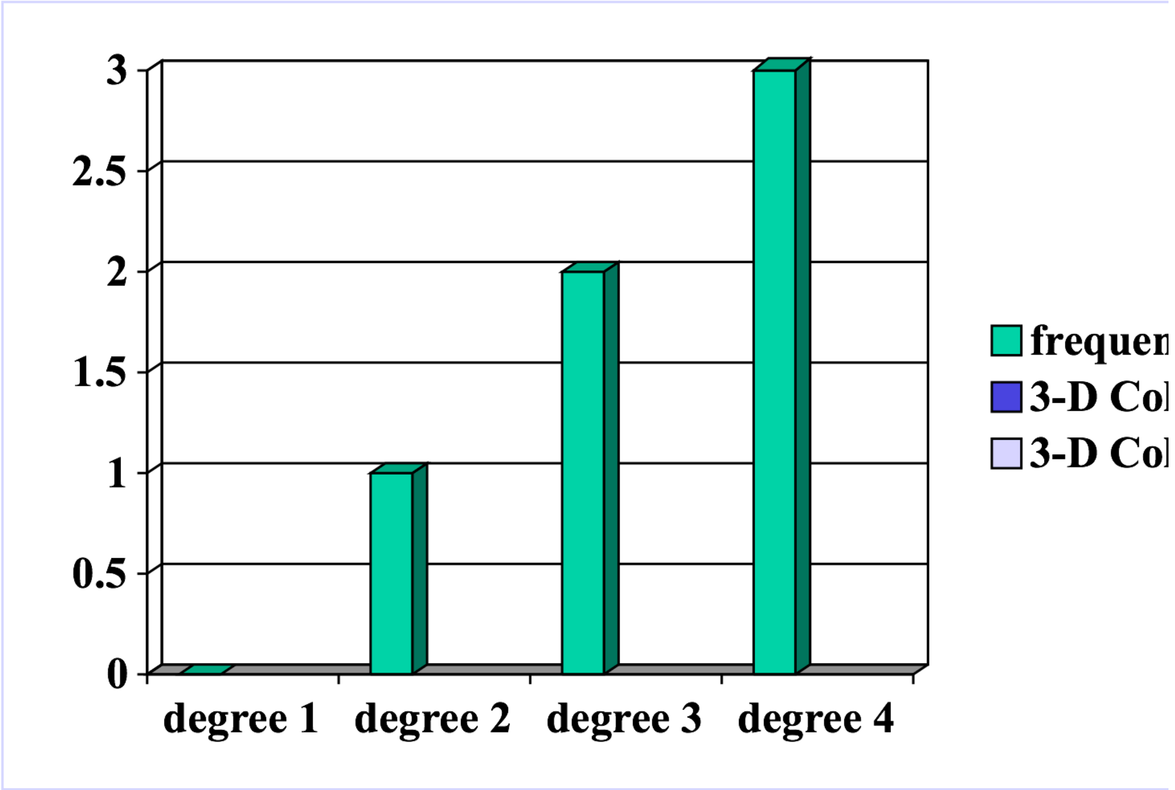
A graph G = (V,E)



Degree sequence (4,4,4,3,3,2)=(d_i), d_i: degree of v_i

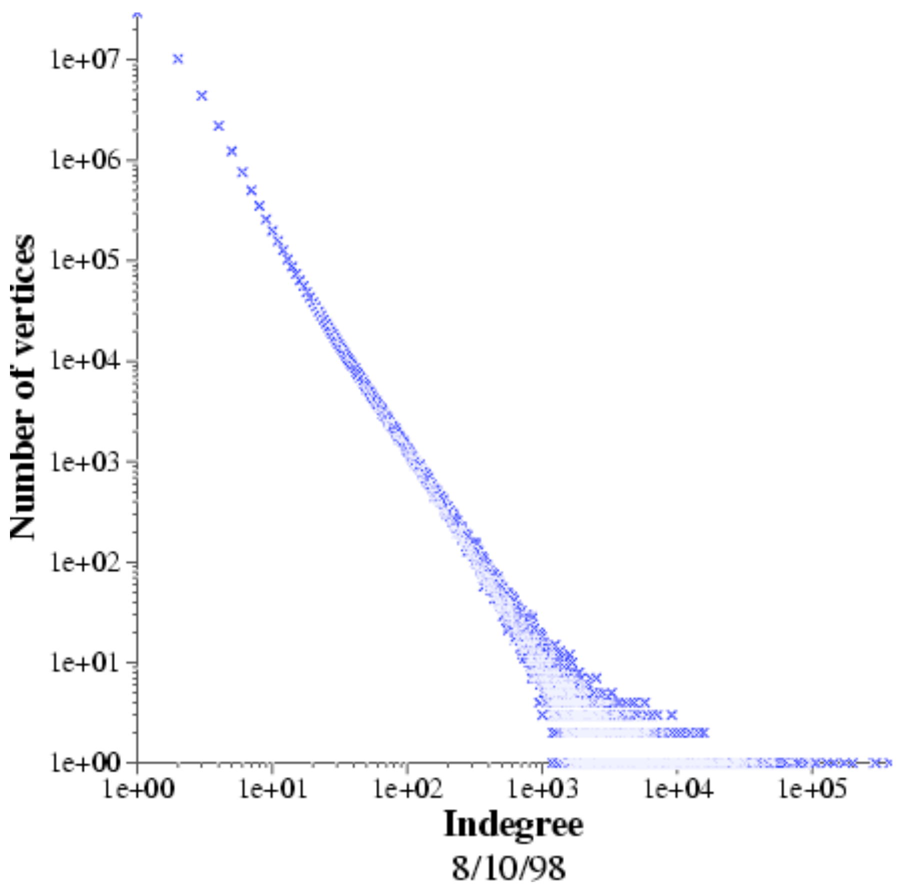
Degree distribution (0,0,1,2,3)=(f_i),
f_i: no. of vertices with degree i.

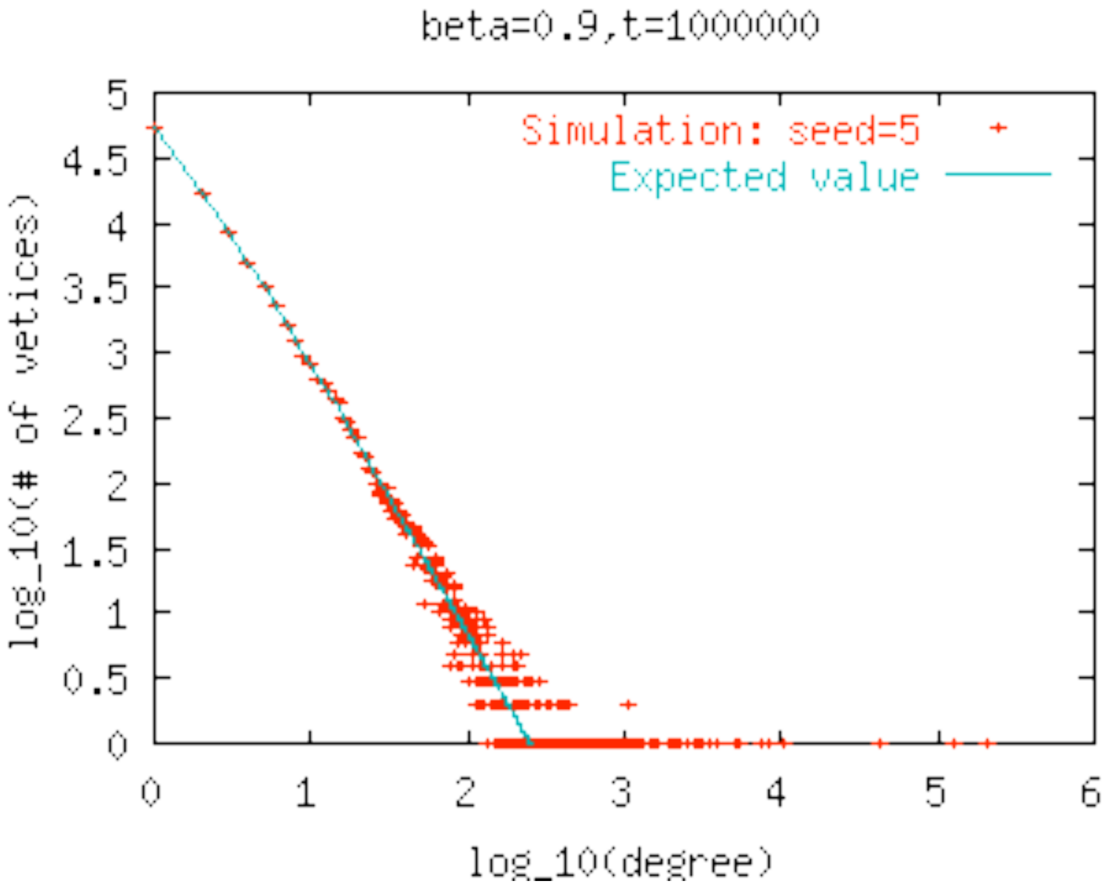




Degree distribution (0,0,1,2,3)=(f_i),
f_i: no. of vertices with degree i.

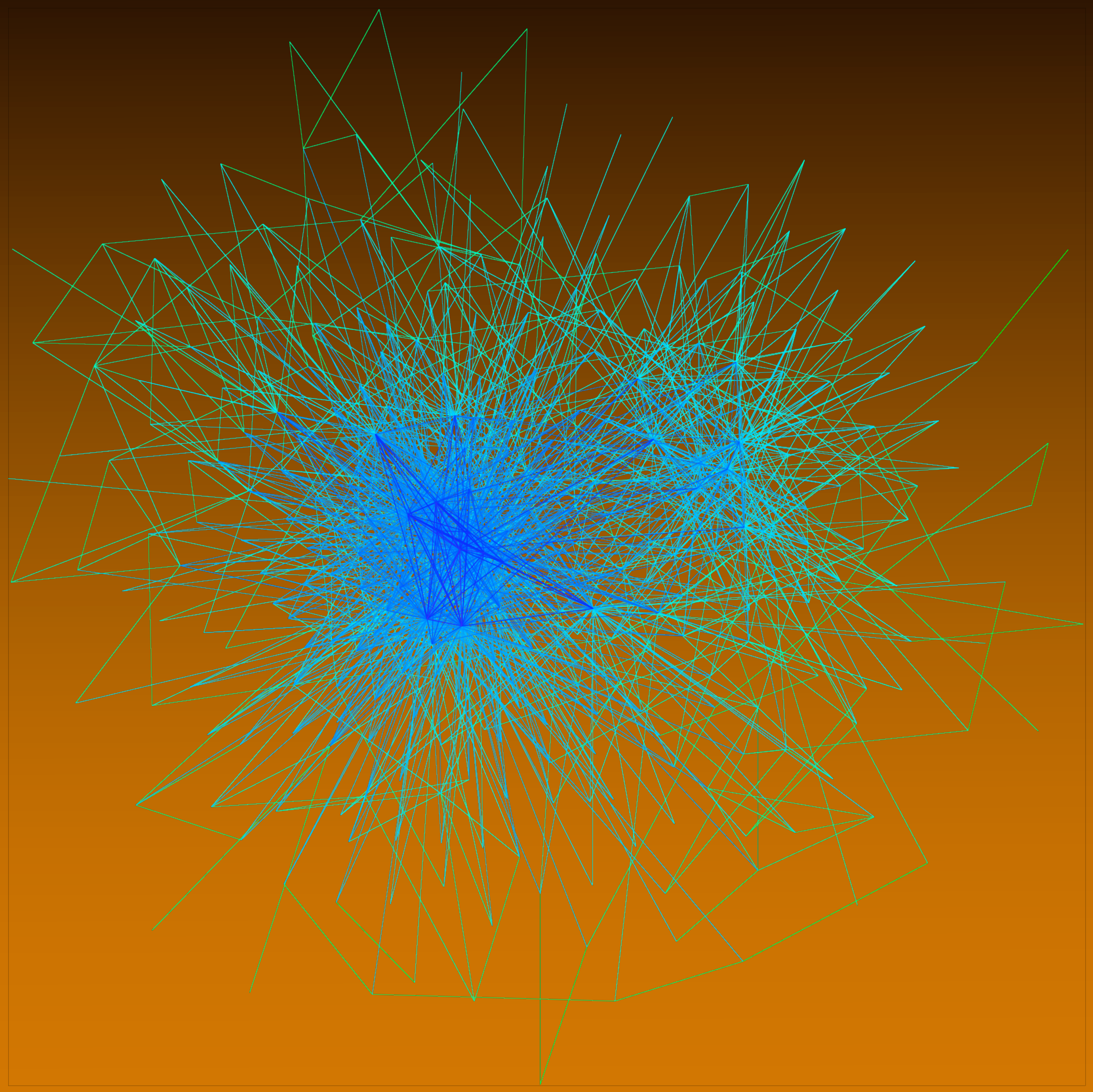
Comparisons



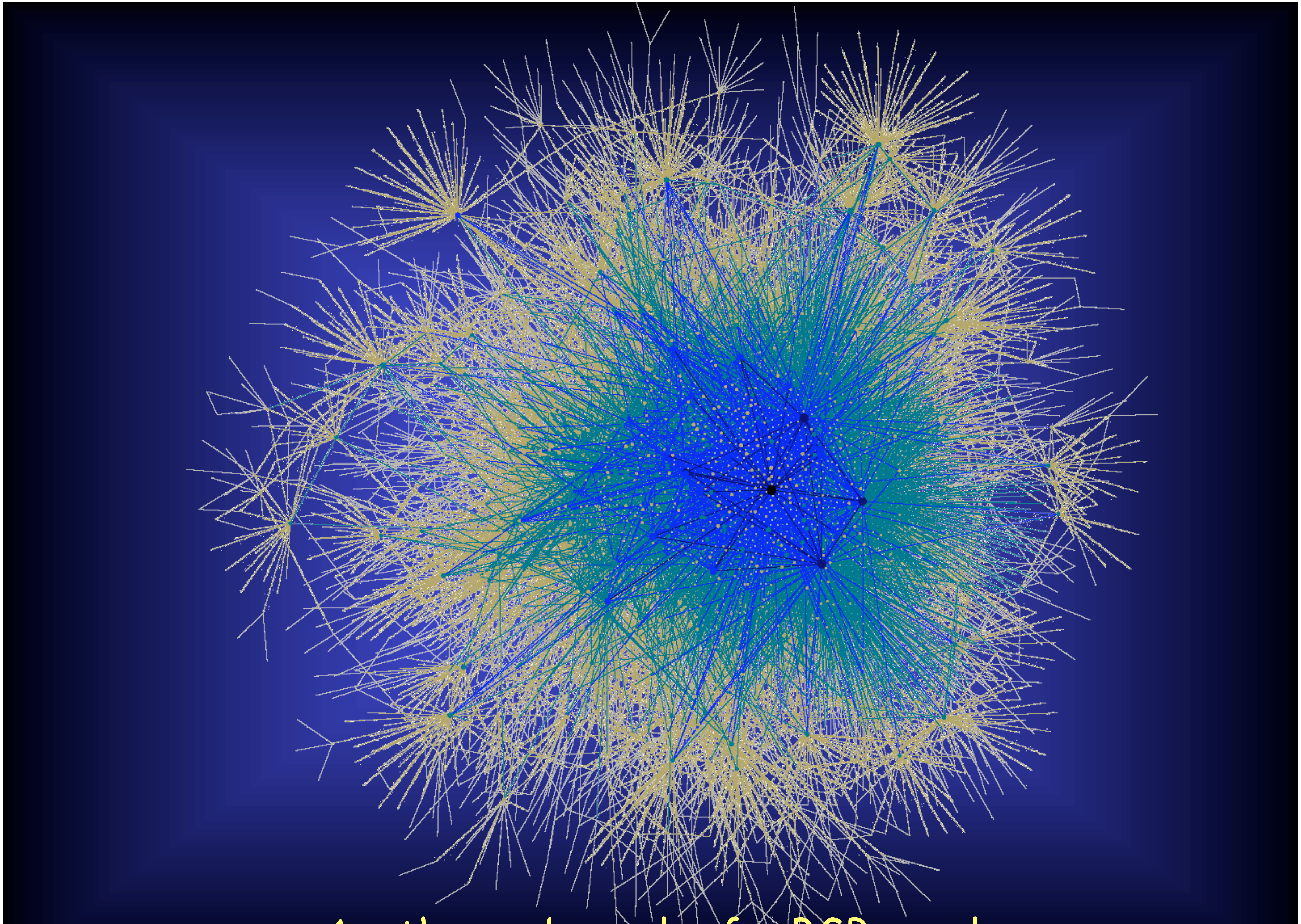


From real data

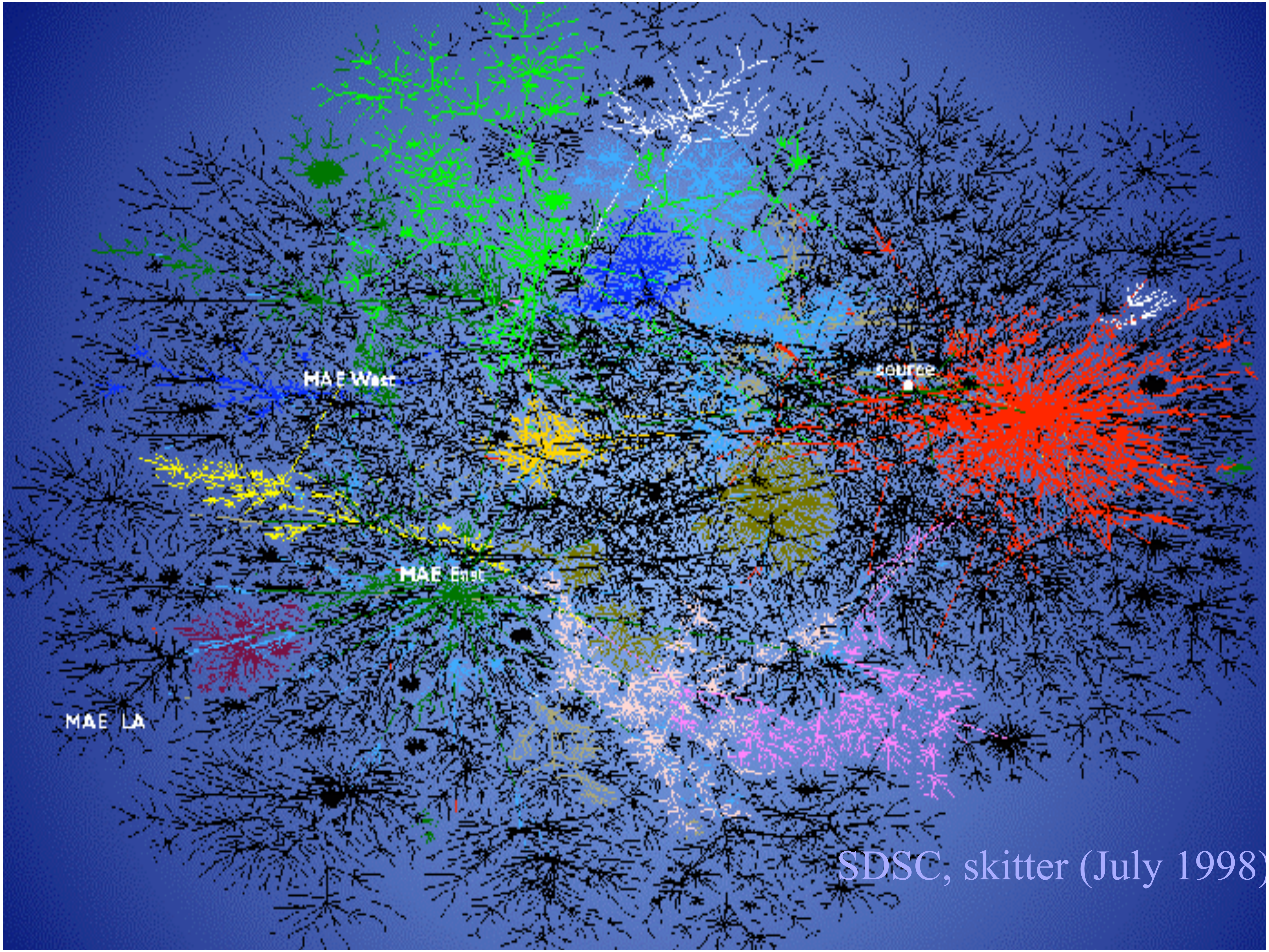
From simulation

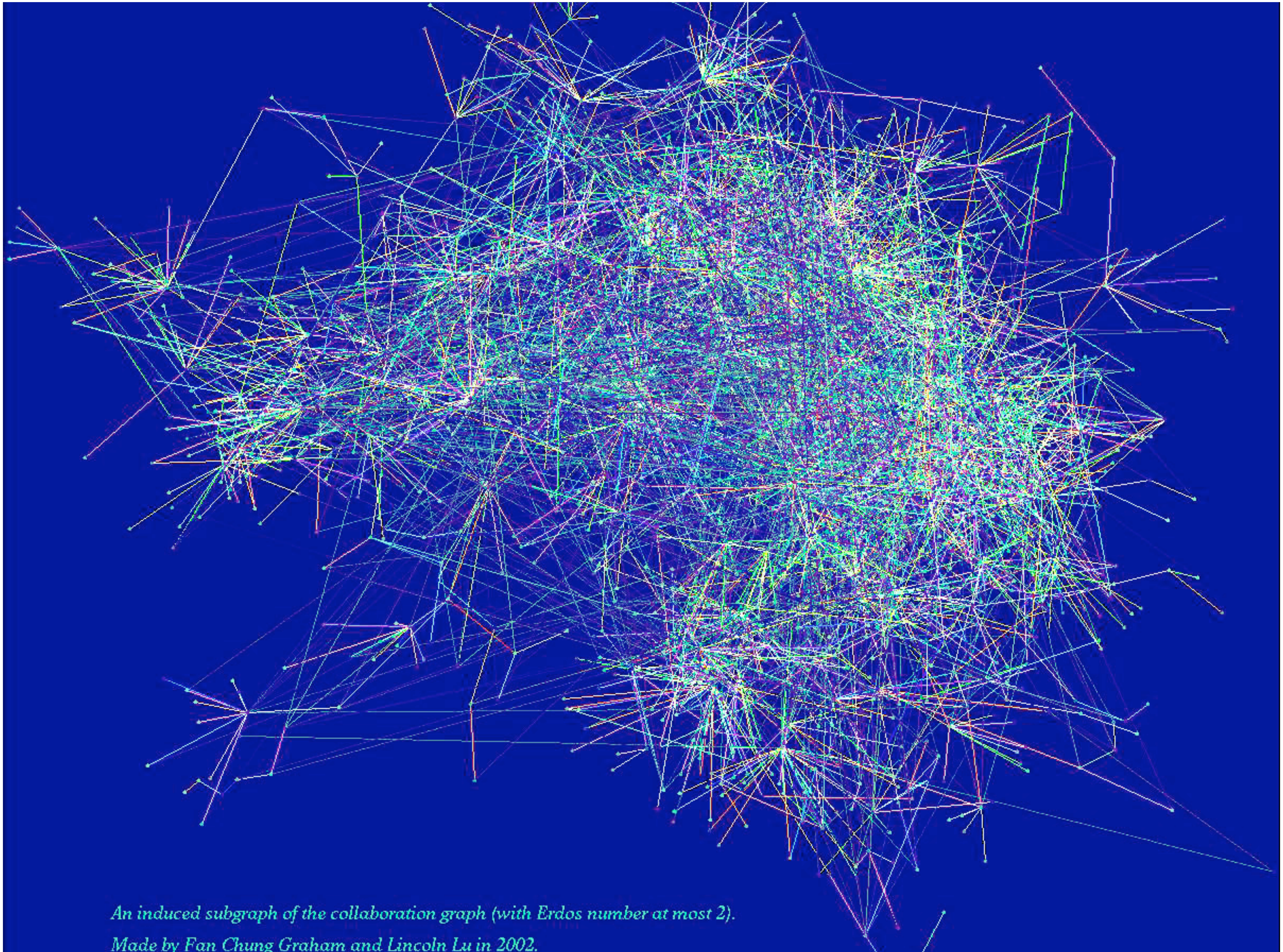


A subgraph of a BGP graph

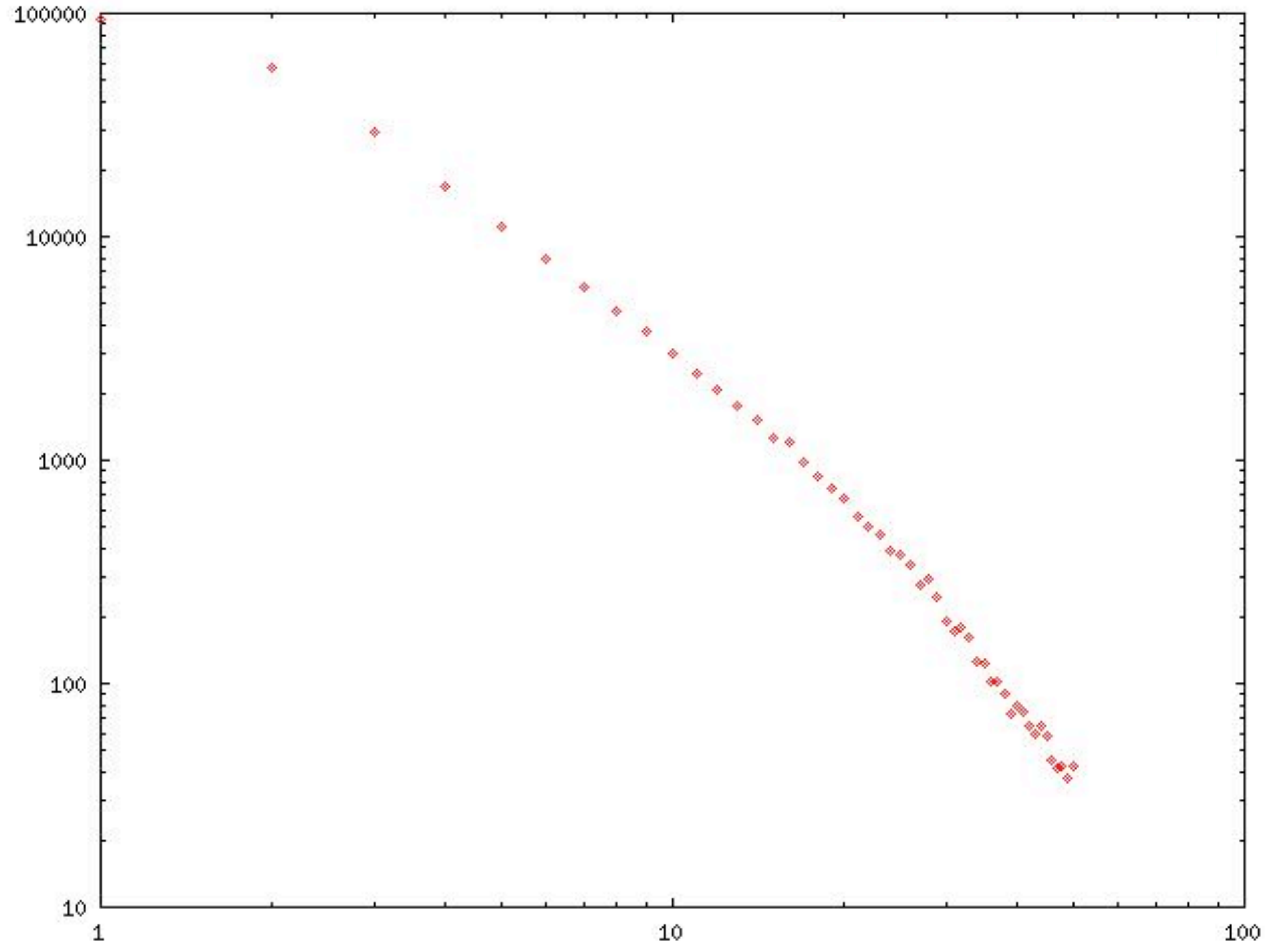


Another subgraph of a BGP graph


MAE West
source
MAE East
MAE LA
SDSC, skitter (July 1998)



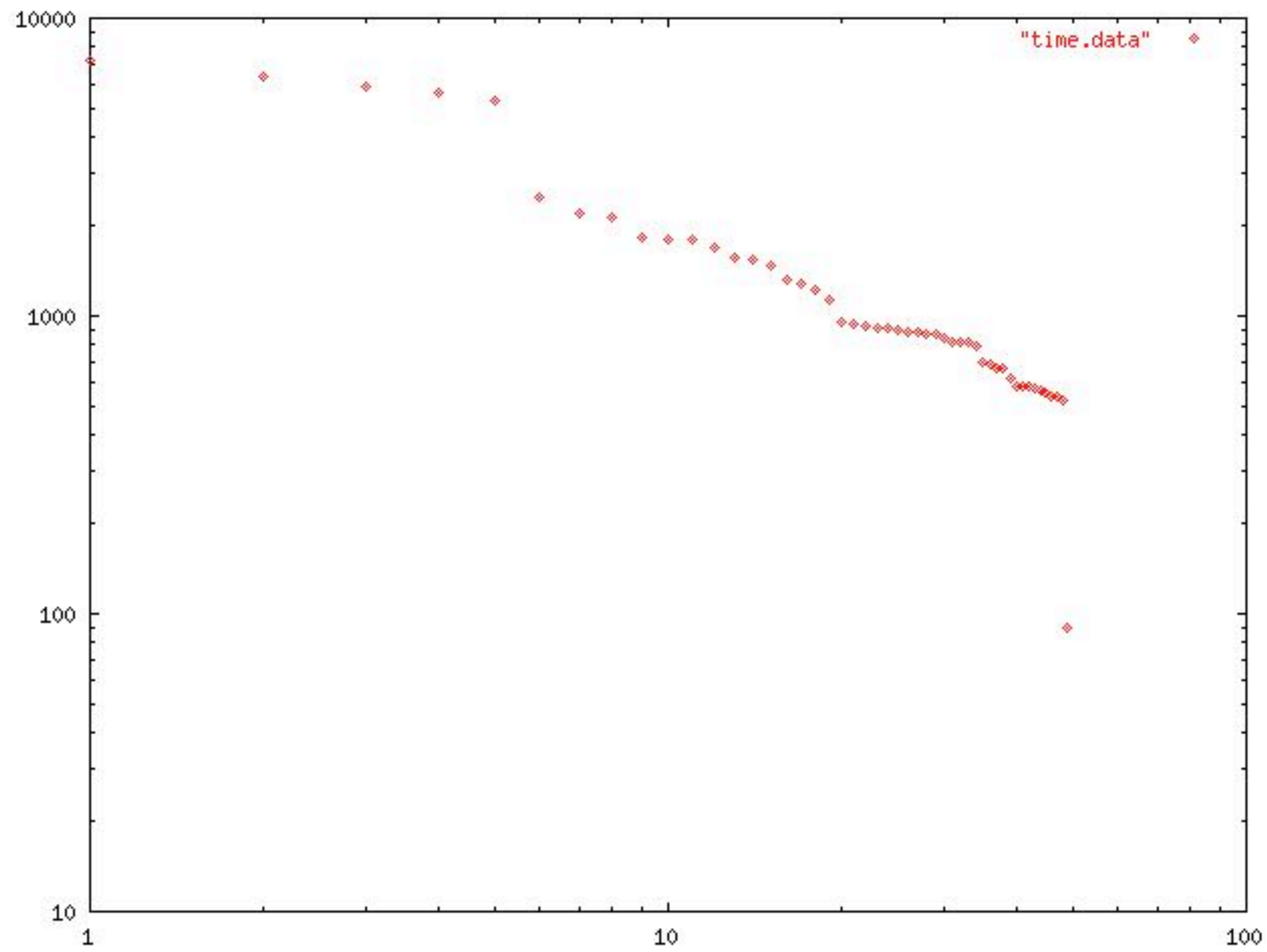
An induced subgraph of the collaboration graph (with Erdos number at most 2).
Made by Fan Chung Graham and Lincoln Lu in 2002.



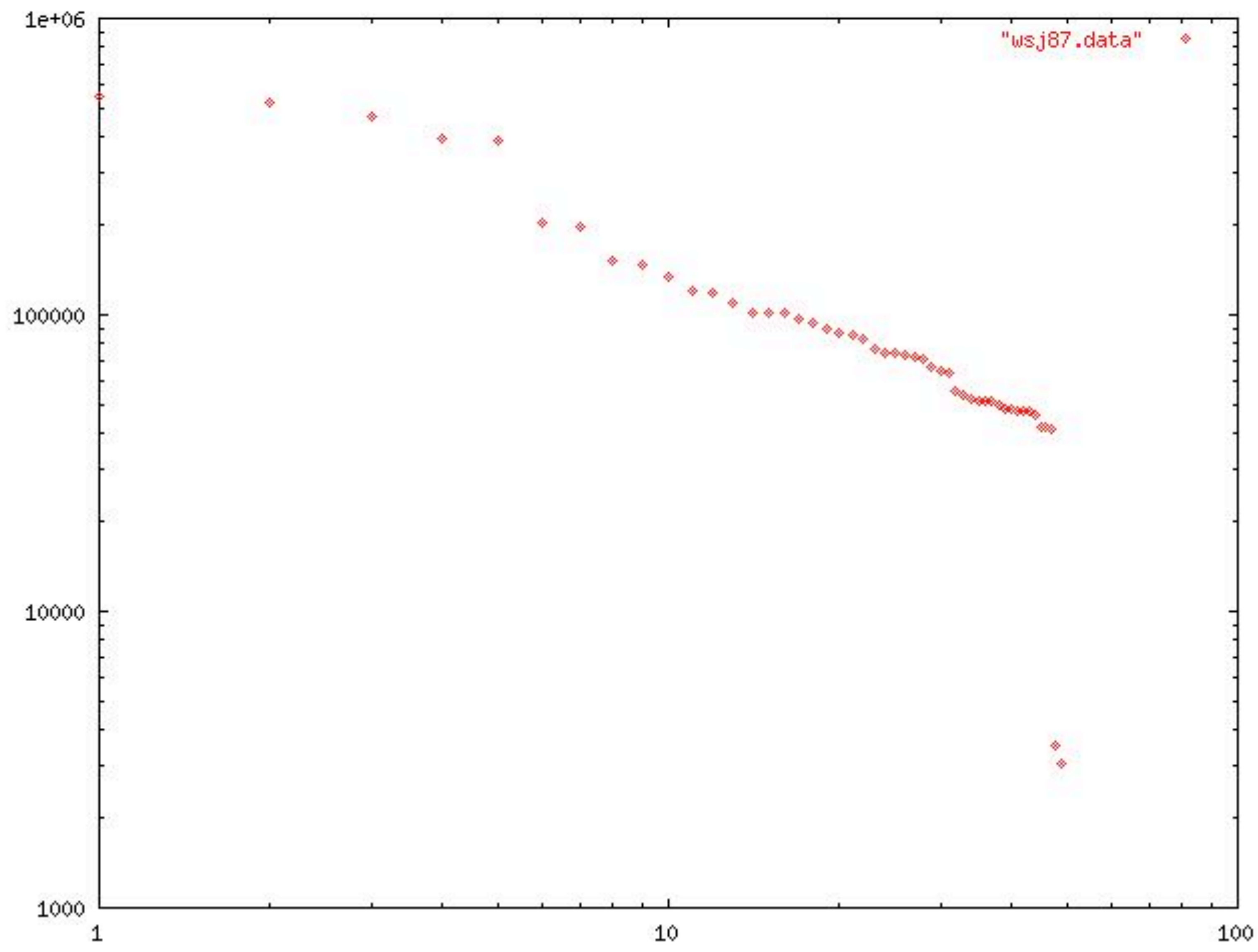
The collaboration graph is a power law graph, based on data from Math Review with 337451 authors with power 2.55

Collaboration graph (Math Review)

- 337,000 authors
- 496,000 edges
- Average 5.65 collaborations per person
- Average 2.94 collaborators per person
- Maximum degree 1401. Guess who?
- A giant component of size 208,000
- 84,000 isolated vertices

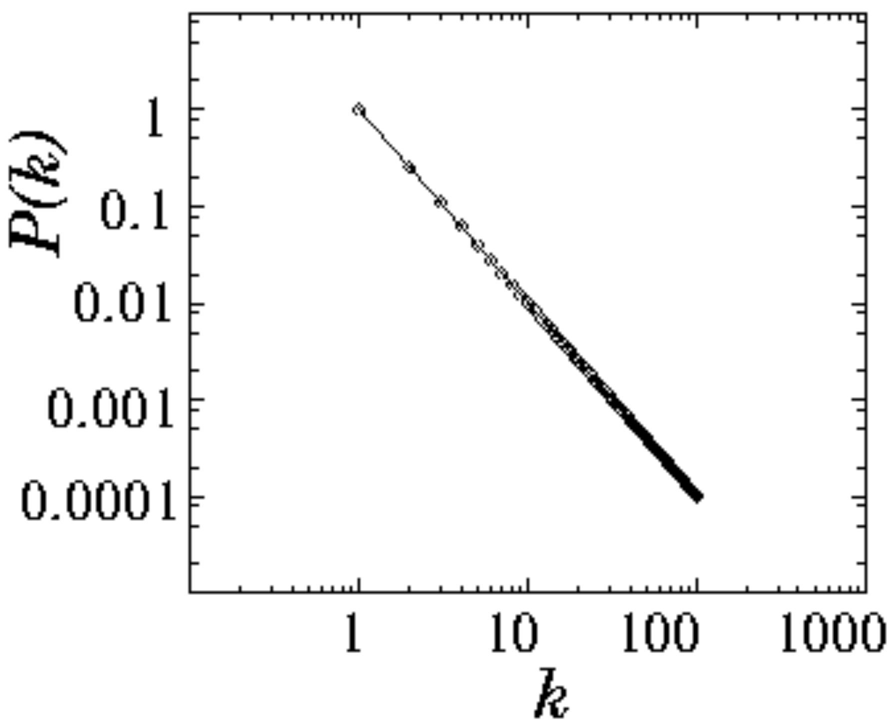


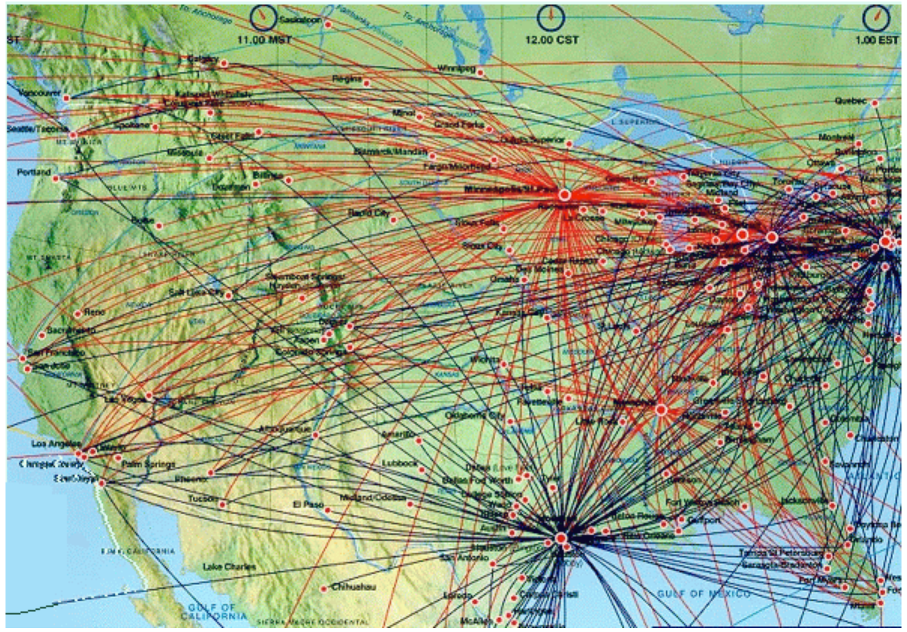
Ocurrences of words in TIME magazine articles
245412 terms.



Occurrences of words in WSJ Collection, a 131.6 MB collection of 46449 newspaper articles (19 million terms). Top 50 terms are included here.

Airline transportation networks are power graphs





Exponents for large power law networks

$$P(k) \sim k^{-\beta}$$

Networks	WWW	Actors	Citation Index	Power Grid	Phone calls
β	~2.1 (in) ~2.5 (out)	~2.3	~3	~4	~2.1

Numerous qustions

- What is a random graph? Which random graphs can best model real networks?
- Local growth rules versus global behavior?
- Communities and clustering
- network games, dynamics ……
- Applications----- routing protocals
  biological networks
  network performance
  …

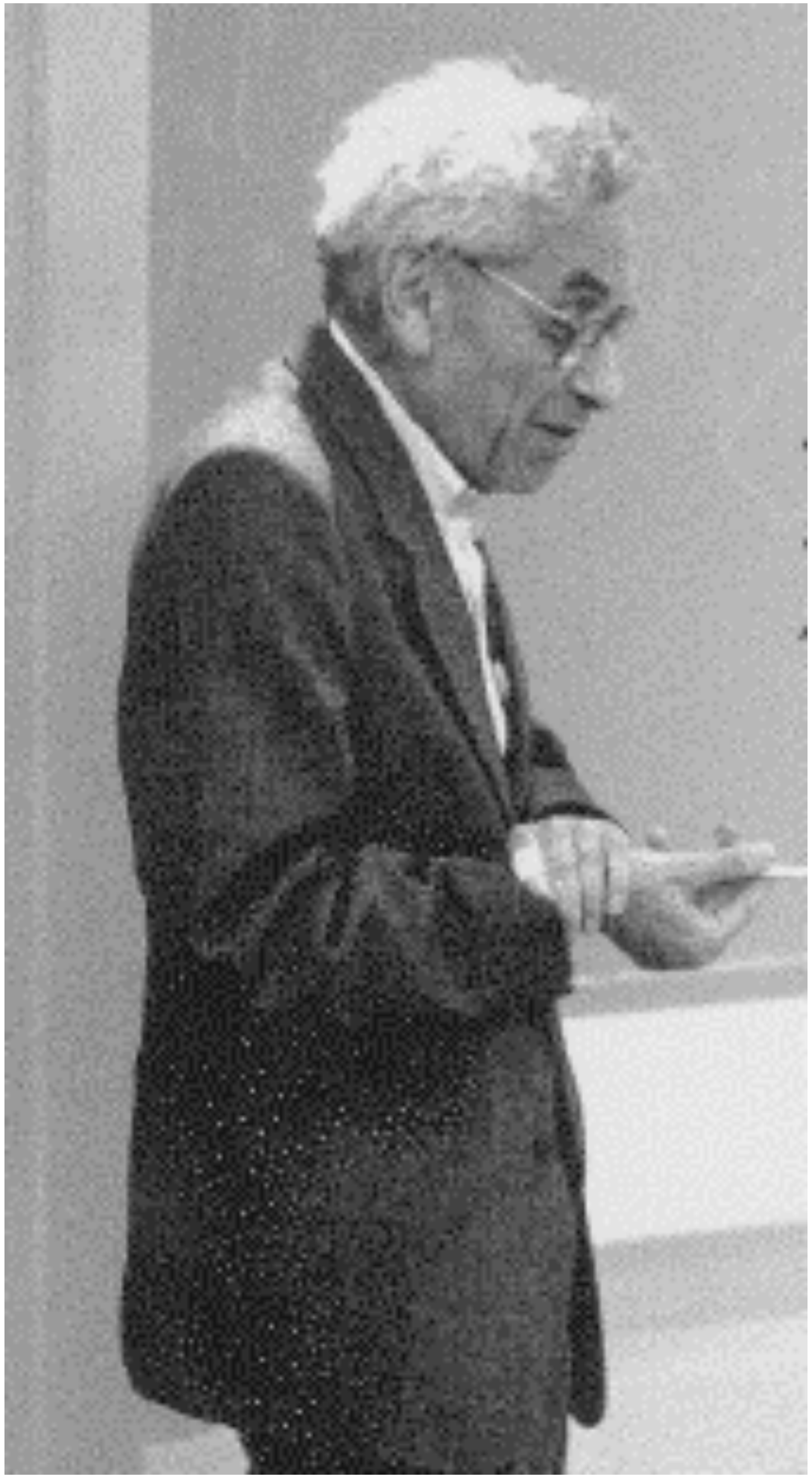
Questions:

- For a given sequence of integers, does it represent the degree sequence of some graph?

  Known. An old theorem of Erdos+Gallai 1960.

- For a given degree sequence of a subgraph, what is the mostly likely degree distribution of the host graph?

Hope I know! Depends on your random graph model!!



CSE 202

Matching algorithms

Fan Chung Graham

UC San Diego

An induced subgraph of the collaboration graph (with Erdos number at most 2).
Made by Fan Chung Graham and Lincoln Lu in 2002.

Gale-Sharpley Algorithm:

function stableMatching {

Initialize all *m* $\in$ M and *w* $\in$ W to *free*

while $\exists$ *free* man *m* who still has a woman w to propose to

{ w = m's highest ranked such woman

if w is *free,* (m, w) become *engaged*

else some pair (m', w) already exists

if w prefers m to m', (m, w) become *engaged* and

m' becomes *free*

else (m', w) remain *engaged*

}

}