Networks in System Architecture

- Goals of this class:
- Background on graph theory and network representations - more to come in the following weeks
- Network applications to system representations and analyses
- Quantitative evaluations possible
- Examples

Why Study Networks?

- Networks capture relationships
- Networks have structure (possibly random)
- Various metrics exist that capture various aspects of this structure
- In some cases the structure or the metrics can be related to important properties of the system or its behavior
- A theory for engineering systems based on these methods is just getting started

Every (Network) Model Is a Choice of Level of Abstraction

- "High" abstraction
  - Summarize, generalize, compare
  - Don't need domain knowledge
- "Low" abstraction
  - Valid detail
  - Explainable differences
  - Need domain knowledge
- Models and analyses at many levels are needed

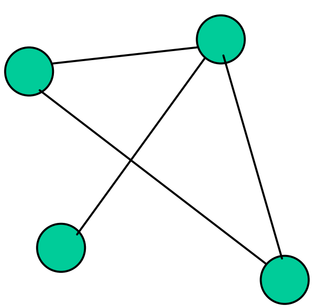
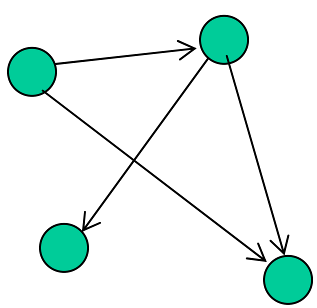
NRC Report on "Network Science"

Researchers across diverse domains share an implicit understanding that a network is more than topology alone. It also entails *connectivity, resource exchange*, and *locality* of action.

- *Connectivity*. A network has a well-defined connection topology in which each discrete entity (node) has a finite number of defined connections (links) to other nodes. In general, these links are dynamic.
- *Exchange*. The connection topology exists in order to exchange one or more classes of resource among nodes. Indeed, a link between two nodes exists if and only if resources of significance to the network domain can be directly exchanged between them.
- *Locality*. The exchanged resource is delivered, and its effects take place, only in local interactions (node to node, link to link). This locality of interaction entails autonomous agents acting on a locally available state.

Graphs and Networks

- A graph is a collection of nodes connected by arcs (directed, with arrows) or links (undirected, no arrows)
  - In graph theory, undirected links are called edges
- A network is a graph

Single Mode and Multimode

- Single mode network
  - Nodes are identical
- Multimode
  - Nodes can be in different classes
  - People
  - Events attended by those people
  - Equivalent to two or more single mode networks
- Common multimode is bipartite

Graph/Network "Rules"

- Links connect pairs of nodes
- Links can be directed or undirected
- Nodes can have any number of impinging links
- Dual graphs can be formed
  - Arcs become nodes
  - Nodes become arcs

Nodes Can Be

- Places
- Things
- Jobs, tasks, process steps
- Calculations or calculation steps

Arcs/links Can Be

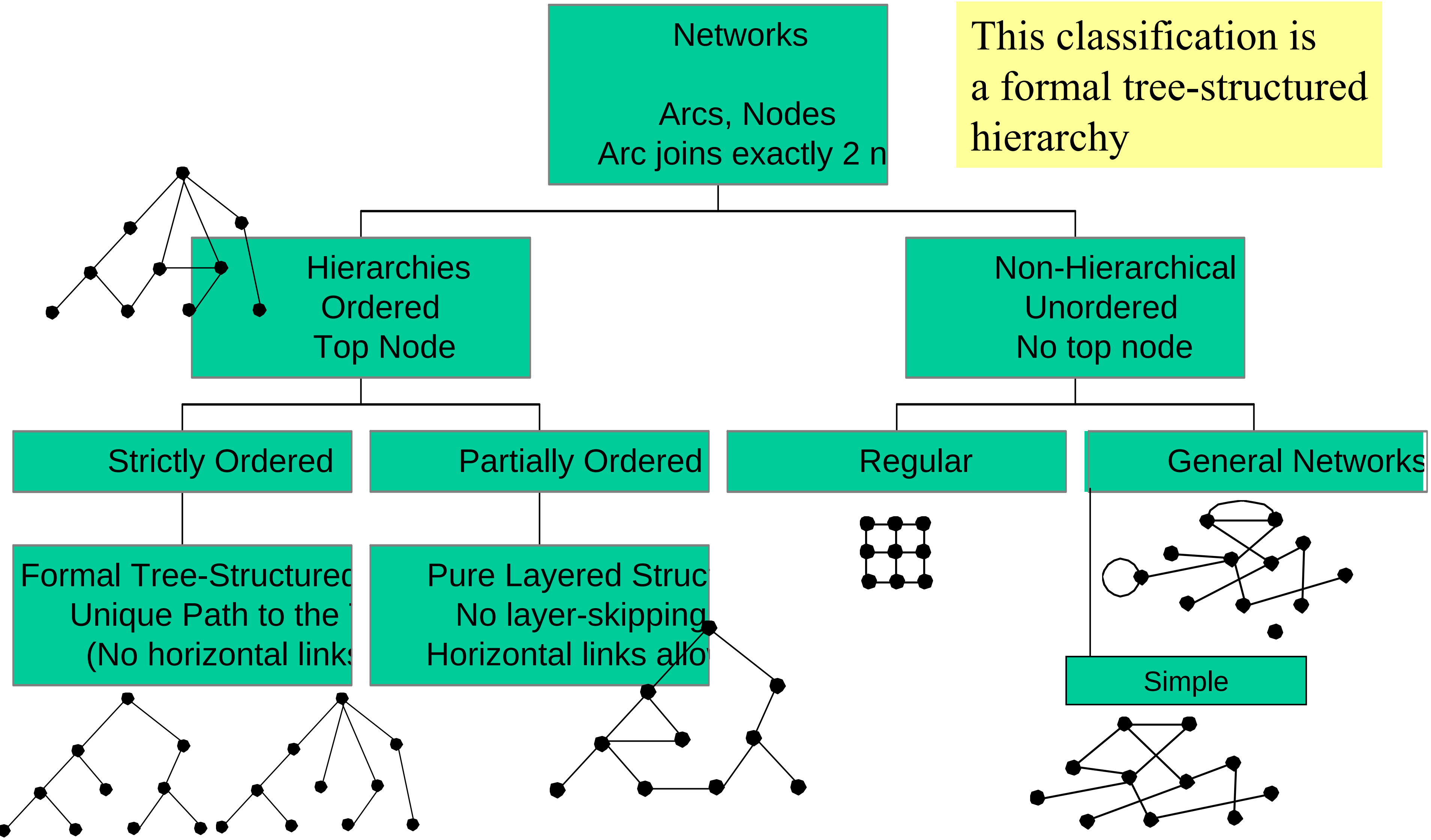
- Physical paths, mechanical joints
- Abstract or real relationships
  - Directed: A commands B, is the father of B, occurs before B…
  - Undirected: A lives near B, is on the same side as B…
- Indications of flow of material or information
- Annotated to represent capacity, direction, content
- Carriers of single or multiple geometric dimensions

Graphs Can Be Classified As…

- Metric (arcs have real lengths, and node locations obey the triangle inequality)
- Non-metric (the layout is purely logical)
- Planar (can be drawn so that no arcs cross)
- Connected or unconnected
  - Connected: a path exists between every pair of nodes
- Simple: no self-loops, ≤ 1 link between nodes
- Able/unable to support a looped path or a path that touches every node
- These are not mutually exclusive

A Classification of Networks

This classification is a formal tree-structured hierarchy



Advantages of Graph Representations

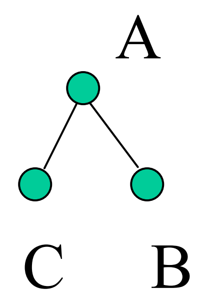
- Abstraction
- Sharp focus on relationships
- Ability to calculate many relevant properties of the modelled system, including many that accommodate huge graphs

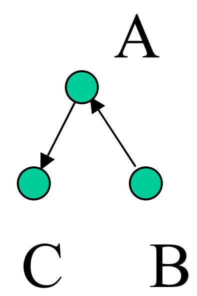
Disadvantages or Shortcomings of Typical Models

- Multiple properties often require separate graphs
- Nodes and arcs are usually treated as identical
  - Seeking abstraction
  - Not knowing enough about the system
  - Exceptions: arcs with costs, bounded flow or one-way flow, nodes classified as toll-takers, sources, sinks, and pass-throughs
- Theory does not deal with the graph as a whole the way, say, set theory does

Real and "Not Real" Networks

- Real: road network, mechanical assembly
- Not real(?): coauthor or movie actor network
- Real but documentable only in a statistical sense: transmission of flu or rumors
- Much depends on the word "also"
  - A writes a paper with B and also one with C
  - A reacts with B and also with C
  - A is bolted to B and also bolted to C
  - B sends signals to A which also sends signals to C
  - B eats A which also eats C





Graph Theory

- Euler is the pioneer
  – How many paths, routes, faces
  – Planar and not planar
- Shortest paths
- Likelihood of a giant cluster
- For random graphs, the authority is Erdös

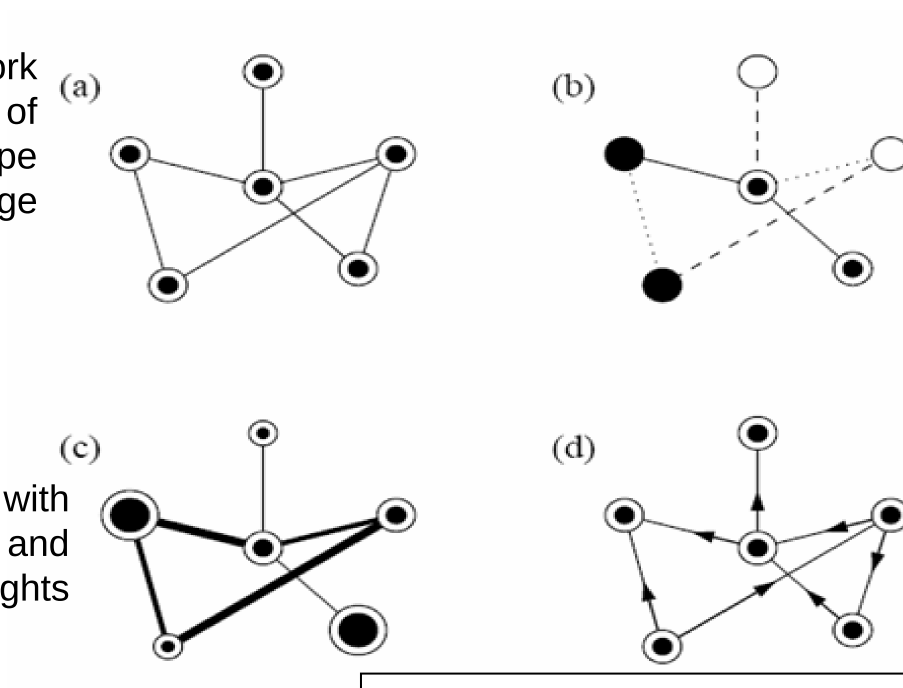
Some Theoretical or Canonical Graph Types

- Planar
- Random
- Grid structured
- Trees
- Hub and spokes
- Possessing Hamilton or Euler circuit
  – Hamilton: touching every node once
  – Euler: touching every arc once

Examples of various types of networks.

an undirected network with only a single type of vertex and a single type of edge

a network with a number of discrete vertex and edge types



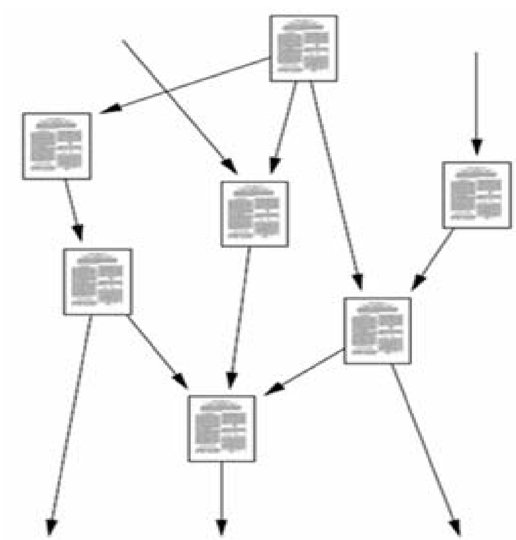
a network with varying vertex and edge weights

a directed network in which each edge has a direction

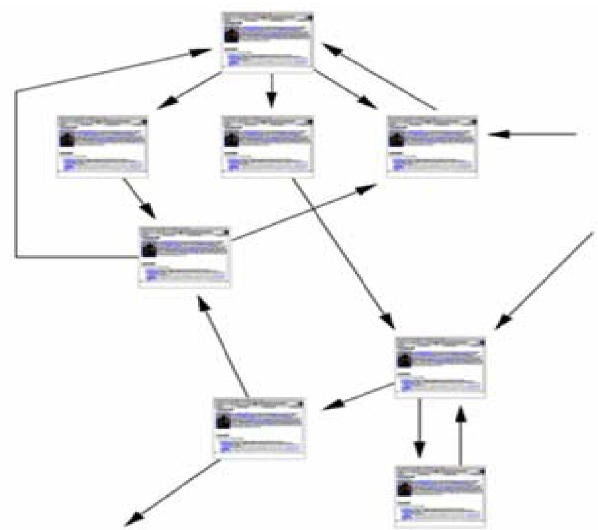
Acyclic and cylic difference following slide

Source: M. E. J. Newman, *The Structure and Function of Complex Networks*, SIAM Review, Vol. 45, No. 2, pp . 167–256, 2003 Society for Industrial and Applied Mathematics

The 2 best studied information networks; citation network and the World Wide Web



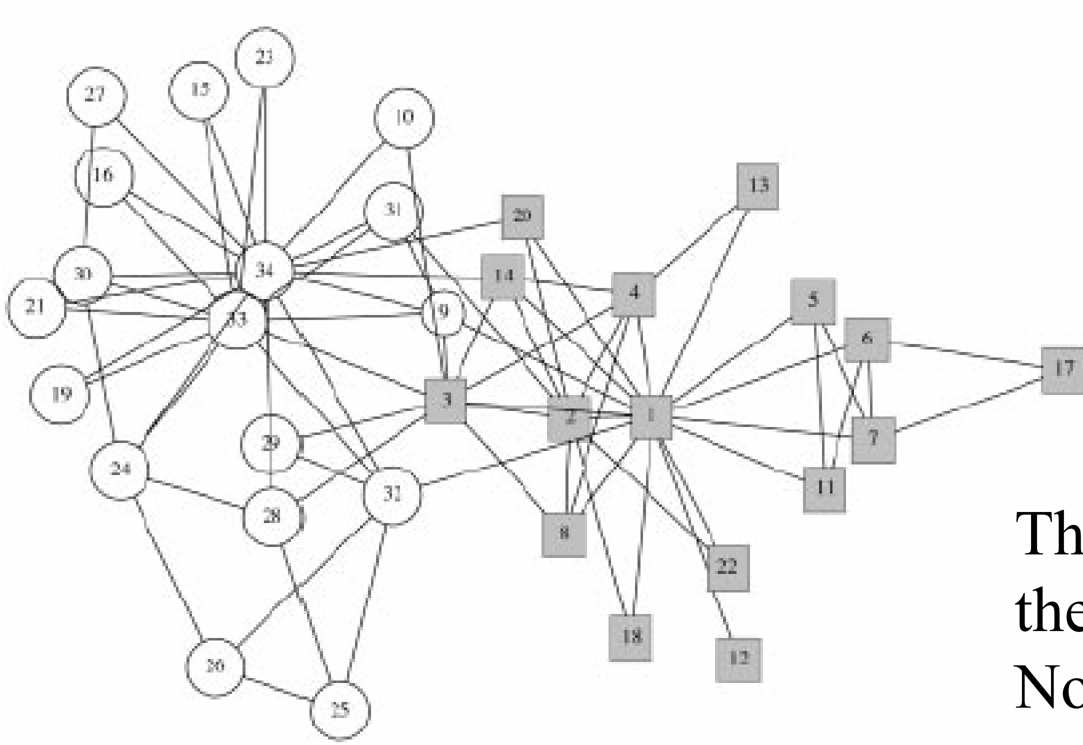
Citation network of academic papers in which the vertices are papers and the directed edges are citations of one paper by another. Since ***papers can only cite those that came before them the graph is acyclic -it has no closed loops***.



World Wide Web, a network of text pages accessible over the Internet, in which the vertices are pages and the directed edges are hyperlinks. There are ***no constraints on the Web that forbid cycles and hence it is in general cyclic***.

Source: M. E. J. Newman, *The Structure and Function of Complex Networks*, SIAM Review, Vol. 45, No. 2, pp . 167–256, 2003 Society for Industrial and Applied Mathematics

Another Well-Studied Network: The Karate Club



This not how the club split. Node 9 is

Community structure in social and biological networks

M. Girvan*†‡ and M. E. J. Newman*§

initial split of the network into two groups is in agreement with ~~the actual factions observed by Zachary~~, with the exception that node 3 is misclassified.

“Traditional” Network Theory

- Main paradigm is flows
- Shortest paths, max flow or lowest cost paths
- Assignment, enumeration
- Traveling salesman and other routing problems
- Many problems can be converted to graph notation and solved using network methods
- Basic text is *Network Flows* by Ahuja, Magnanti, and Orlin

Example Networks

- Road map
- Electric circuit or pipe system
- Structure of bridge or building, with load paths
- Organizational chart or social network
- Markov chain
- Control circuit feedback loop
- Phone system
- Chemical reaction
- Sequential event plan

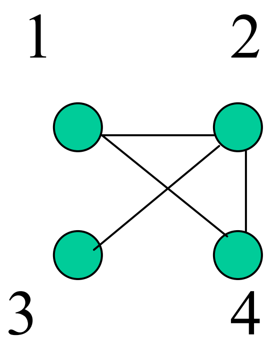
More Example Analyzable Networks

- Manufacturing process
- Assembly sequence
- Schedule
- Family tree
- Ecological food chain
- Taxonomy of living things, rocks, and other natural hierarchies
- Naval battle, military campaign

Possible Analyses

- Finding an ordering on the nodes
  - Schedule, seating arrangement, space allocation, assembly sequence
- Enumerating cut sets
- Analyzing electric circuits and other applications of linear algebra
- Calculating mechanism properties like mobility and constraint
- Calculating control system stability
- Estimating or calculating system complexity
- Paths: shortest, max capacity, least cost, critical, first passage time, etc. Second place, 3rd place, etc

Matrix Representation



»A=[0 1 0 1;1 0 1 1;0 1 0 0;1 1 0 0]

A =

0	1	0	1
1	0	1	1
0	1	0	0
1	1	0	0

»

A is called the adjacency matrix

$A(i,j)=1$ if there is an arc from i to j

$A(j,i)=1$ if there is also an arc from j to i

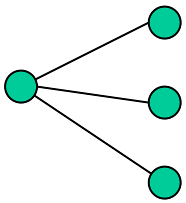
In this case A is symmetric and the graph is undirected

Basic Graph Definitions and Calculations

- Nodal degree, in-degree, out-degree
- Average degree
- Degree sequence
- Rewiring
- Degree correlation

Basic Facts About Undirected Graphs

- Let *n* be the number of nodes and *m* be the number of links
- Then *average nodal degree* is $<k>=2m/n$
- The *Degree sequence* is a list of the nodes and their respective degrees $D=[3\ \ 1\ \ 1\ \ 1]$
- The sum of these degrees is $\sum_{i=1}^{n} d_i = 2m$
- D=sum(A) in Matlab
- sum(sum(A)) = 2m



Rewiring

- A way to deliberately transform a graph
- Several ways this is done
  - Unhooking one end of a link and hooking it in somewhere else
  - Adding a new link
  - Pairwise rewiring that preserves the original degree sequence
    - This can disconnect the graph unless you take care to reject rewirings that do so

Rewiring - 2

Unhook-rehook links

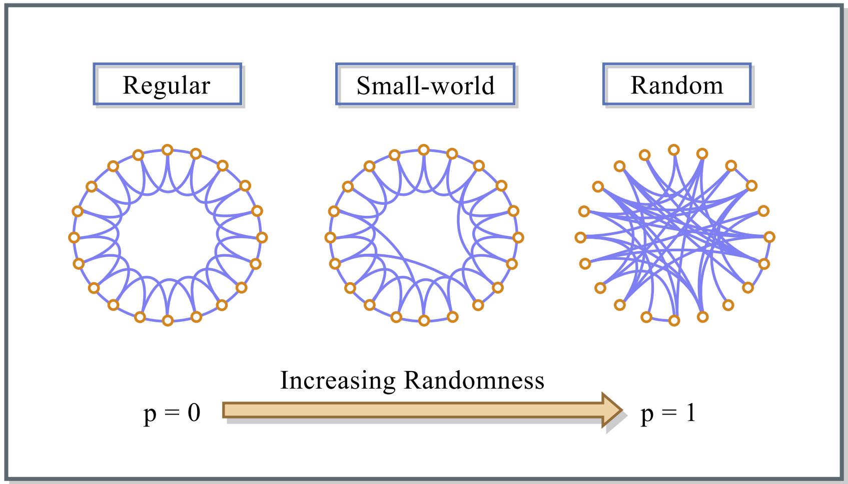


Figure by MIT OCW. After Watts & Strogatz, 1998.

Preserving degree

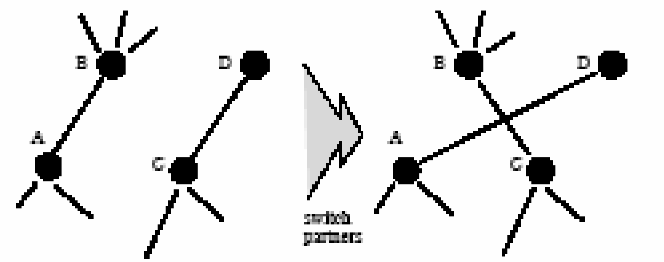
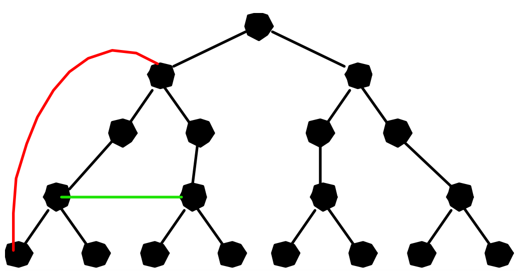


FIG. 1. One elementary step of the local rewiring algorithm. A pair of edges A—B and C—D is randomly selected. They are then rewired in such a way that A becomes connected to D, and C - to B, provided that none of these edges already exist in the network, in which case the rewiring step is aborted, and a new pair of edges is selected. The last restriction prevents the appearance of multiple edges connecting the same pair of nodes.

Detection of Topological Patterns in Complex Networks: Correlation Profile of the Internet

Sergei Maslov[1], Kim Sneppen[2,3], Alexei Zaliznyak[1]

Add links



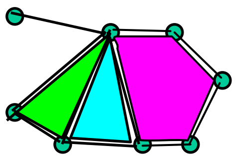
Information exchange and the robustness of organizational networks

Peter Sheridan Dodds*†, Duncan J. Watts*‡§, and Charles F. Sabel¶

More Definitions and Calculations

- Geodesic - shortest path between two nodes
- Average path length = avg of all geodesics
- Graph diameter = longest geodesic
- Clustering coefficient = fraction of all possible triangles actually present
- More on these in later classes

Planar Graph Example



$n =$ number of nodes

$m =$ number of links

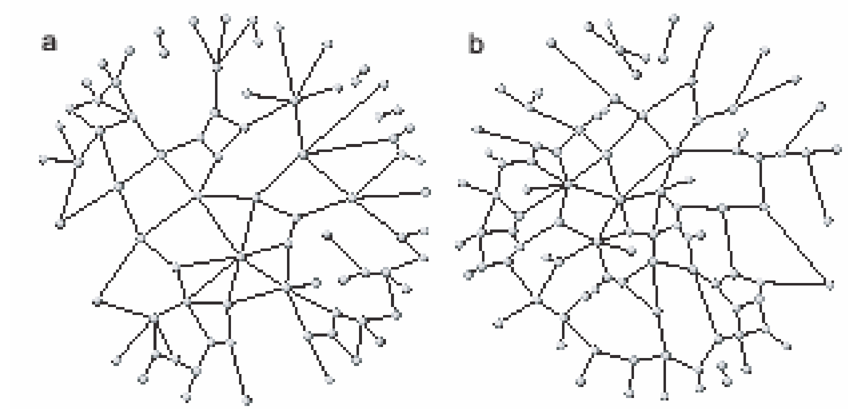
$f =$ number of facets

$f = m - n + 2$

$E \le 3n - 6$

$\therefore\ f_{\max} = 2n - 4$

Meshness Ratio $M = f / f_{\max}$



Trails made by ants in planar sand piles have average nodal index <k> = 2.2 and M ~ 0.1

Note: for connected planar graphs:

$$0 \le M \le 1$$

$$< k >_{\max} = \frac{6n-12}{n} \xrightarrow[n \to \infty]{} 6$$

$$< k >_{\min} = \frac{2n-2}{n} \xrightarrow[n \to \infty]{} 2$$

$$< k > \xrightarrow[n \to \infty]{} 4M + 2$$

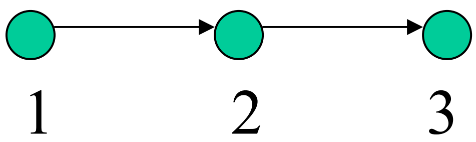
“Efficiency and Robustness in Ant Networks of Galleries,”
J. Buhl, J. Gautrais1, R.V. Sol´e, P. Kuntz, S. Valverde2, J.L. Deneubourg, and G. Theraulaz, Eur. Phys. J. B **42**, 123–129 (2004)

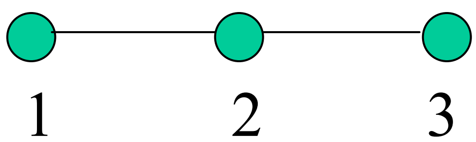
Matlab Routines

- On Stellar there are several Matlab routines that calculate many of these simple statistics
- Of course you can write your own.
- I have downloaded many Matlab graph theory routines but many seem to have bugs or do not work the way I expect.
- Always test any routine (your own or someone else’s) on some simple graphs first

Simple Test

```
g =

    0    1    0
    0    0    1
    0    0    0

»isconnected(g)
ans =

    0 % i.e., false

»g=g+g' %' = transpose
g =

    0    1    0
    1    0    1
    0    1    0

»isconnected(g)
ans =
   1 % i.e., true
```



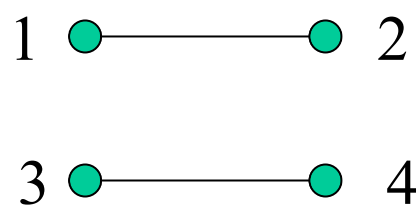


"Yeah. Sorry. I only built it for undirected graphs."

Another

Funny I didn't catch that before.
Line 49 of dfs.m should have a return statement.

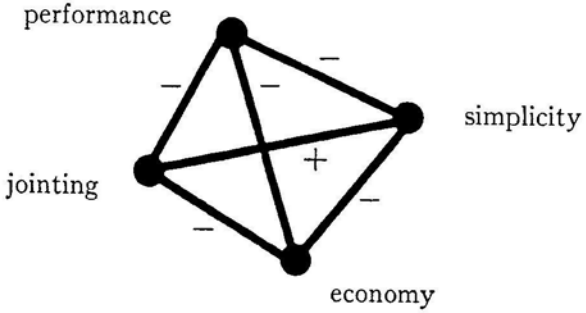
```
On 1/15/06, Daniel Whitney  wrote:
> Kevin
>
> Sorry to bother you with what I suppose is a simple question:
>
> Can you tell me why connected_graph returns true for the adj_mat
>
> »adj_mat=[0 1 0 0;1 0 0 0;0 0 0 1;0 0 1 0]
>
> adj_mat =
>
>      0     1     0     0
>      1     0     0     0
>      0     0     0     1
>      0     0     1     0
>
> »connected_graph(adj_mat,0)
>
> ans =
>
>      1
>
> »
>
> This graph is obviously disconnected.  Your
> routine draw_graph confirms this - see attached
> jpg.
```



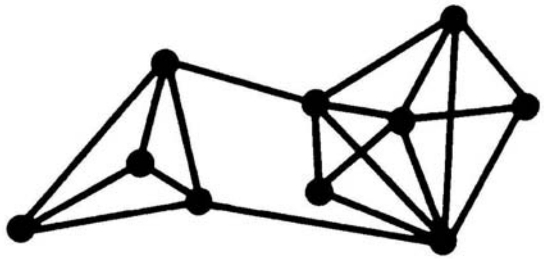
Link Between Networks and Systems

- Hierarchical descriptions
- General network descriptions
- Depiction of the decomposition process as a tree
- Depiction of the synthesis process as clustering
- Early thinkers: Simon, Alexander

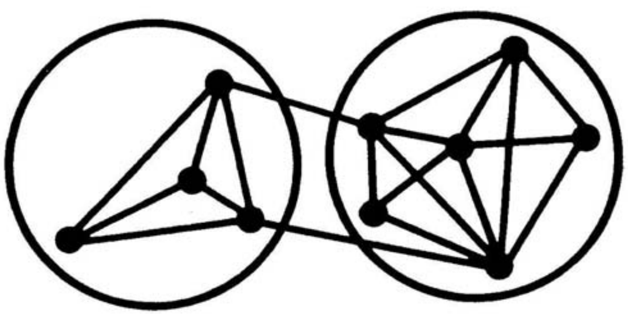
Alexander’s Depiction of System Design



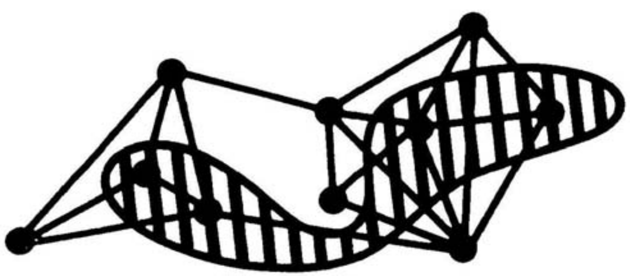
Requirements conflict



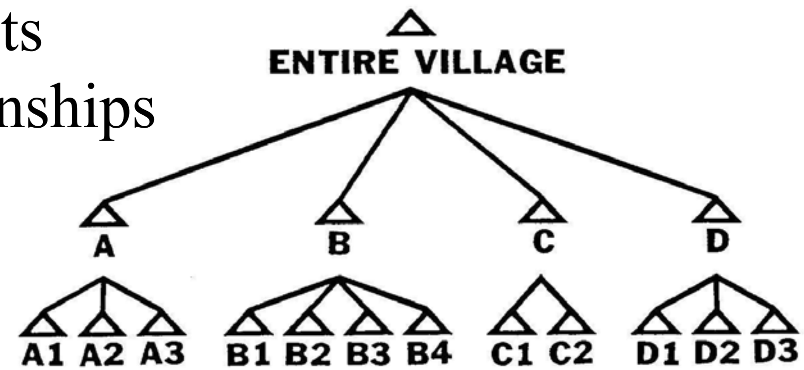
A system and its internal relationships



Good partition



Bad partition



Result of clustering many conflicting requirements

C. Alexander
Notes on the Synthesis of Form,
Cambridge: Harvard U Press, 1964

Simon and Alexander

- Simon: modularity related to evolution, survival, and complexity (1962)
  - Can’t survive (due to propagating failures) unless there is some independence between modules
  - The modules form into (nested/closed) hierarchies
  - Evolution can proceed in separate activities
  - Also: decomposable systems are less complex
- Alexander: modularity related to efficient design procedures (1964)
  - Can’t make design decisions (due to interactions) unless they are clustered meaningfully and dealt with in bulk
  - Design can proceed as independent steps
  - The right clusters may not be the obvious visible ones
- Both: perfect clean decomposition is impossible

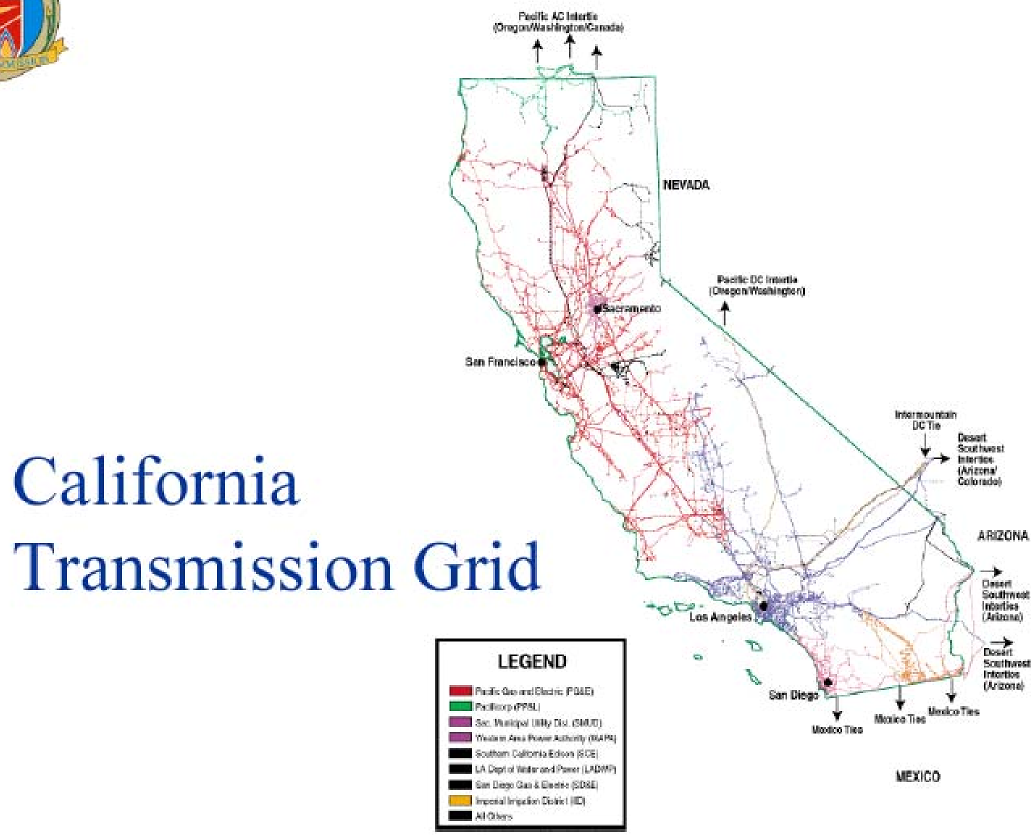
“New” Network Theory

- Not new in social science
- Main paradigm is connections and their implications: connectedness of the graph, statistics of nodal index k = links/node
- Identification of common statistical properties in a wide variety of graphs of real systems, both natural and man-made
- Clusters, cliques
- Scale free and power law properties
- Analyses of percolation, spread of diseases and rumors, vulnerability to attack

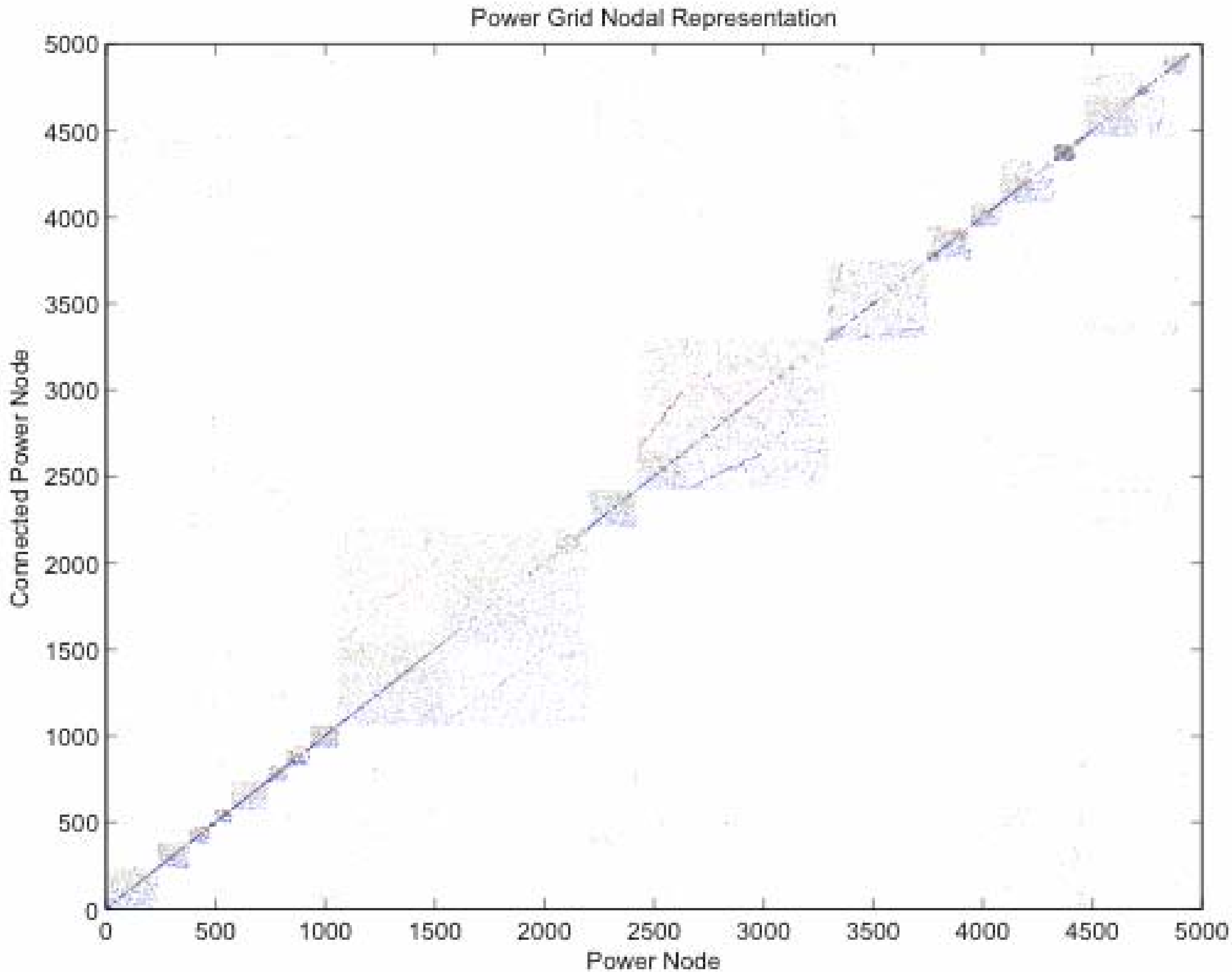
Diameter and Structure

- How many plane changes are needed to fly from city A to city B and how does this change as the network grows?
- In a point to point network, the number might grow linearly with the number of cities in the network, unless there is a link between most pairs (max n(n-1)/2 links needed)
- In a hub-spokes network, the number may hardly grow at all even if almost no cities are directly linked
- Of course, the distance flown and time spent flying and waiting are longer for hub-spokes

Part of Western Power Grid



Drawn by Spencer Lewis for last year's class

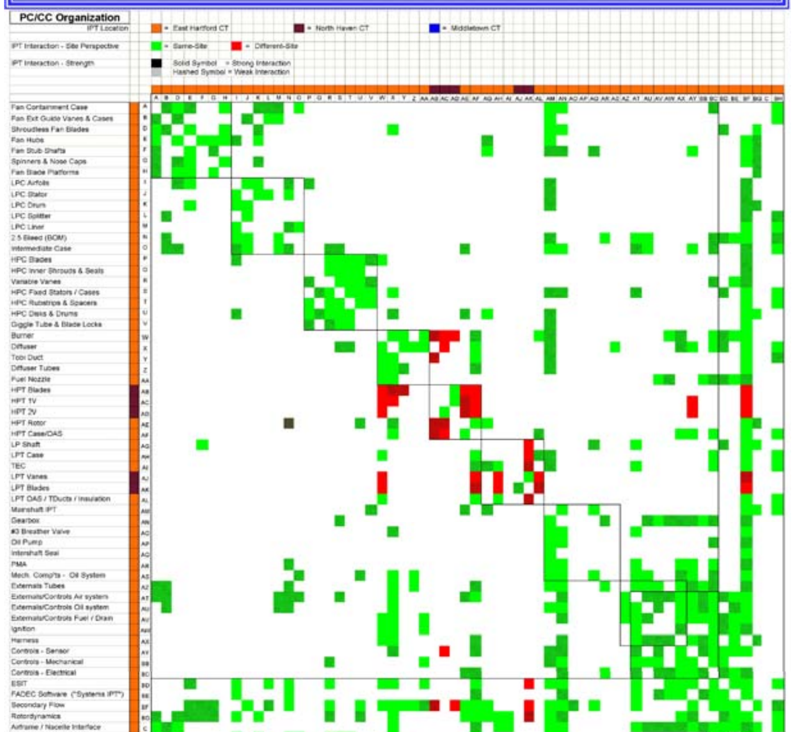


Glynn & Pelland’s Maps of Decisions and Locations

PW4098 effort involving 59 IPTs with a total of 605 IPT interactions (369 strong)

Under Product Center / Component Center Structure
559 of 605 (92%) interactions occur co-located on same site
46 of 605 (8%) interactions have to span sites

Vertical (and horizontal) strips of color indicate the IPT site
The matrix coloring indicates if the interacting IPTs are co-located by site
Indicates IPTs are NOT co-located on same site
Indicates IPT’s ARE co-located on same site



PW4098 effort involving 59 IPTs with a total of 605 IPT interactions (369 strong)

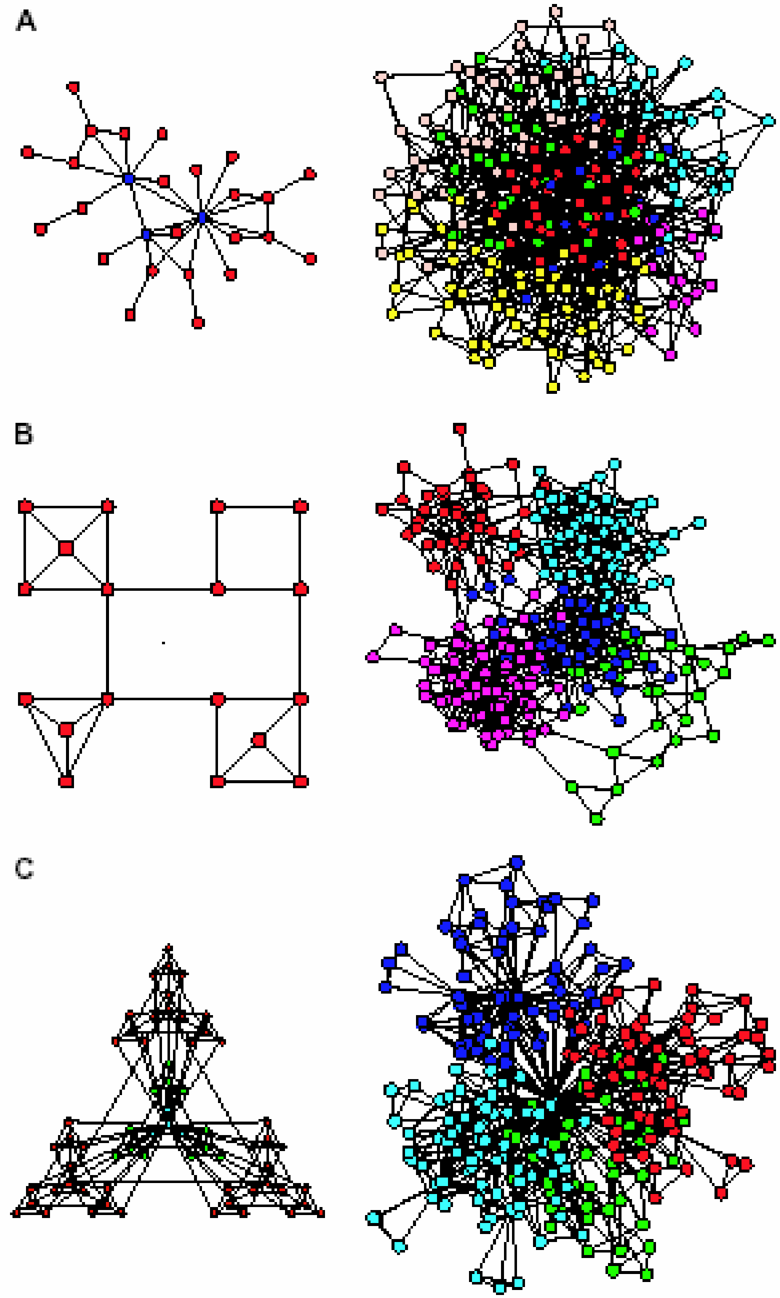
Under Current Module Center Structure
366 of 605 (60%) interactions occur co-located on same site
239 of 605 (40%) interactions have to span sites

Vertical (and horizontal) strips of color indicate the IPT site
The matrix coloring indicates if the interacting IPTs are co-located by site
Indicates IPTs are NOT co-located on same site
Indicates IPT’s ARE co-located on same site

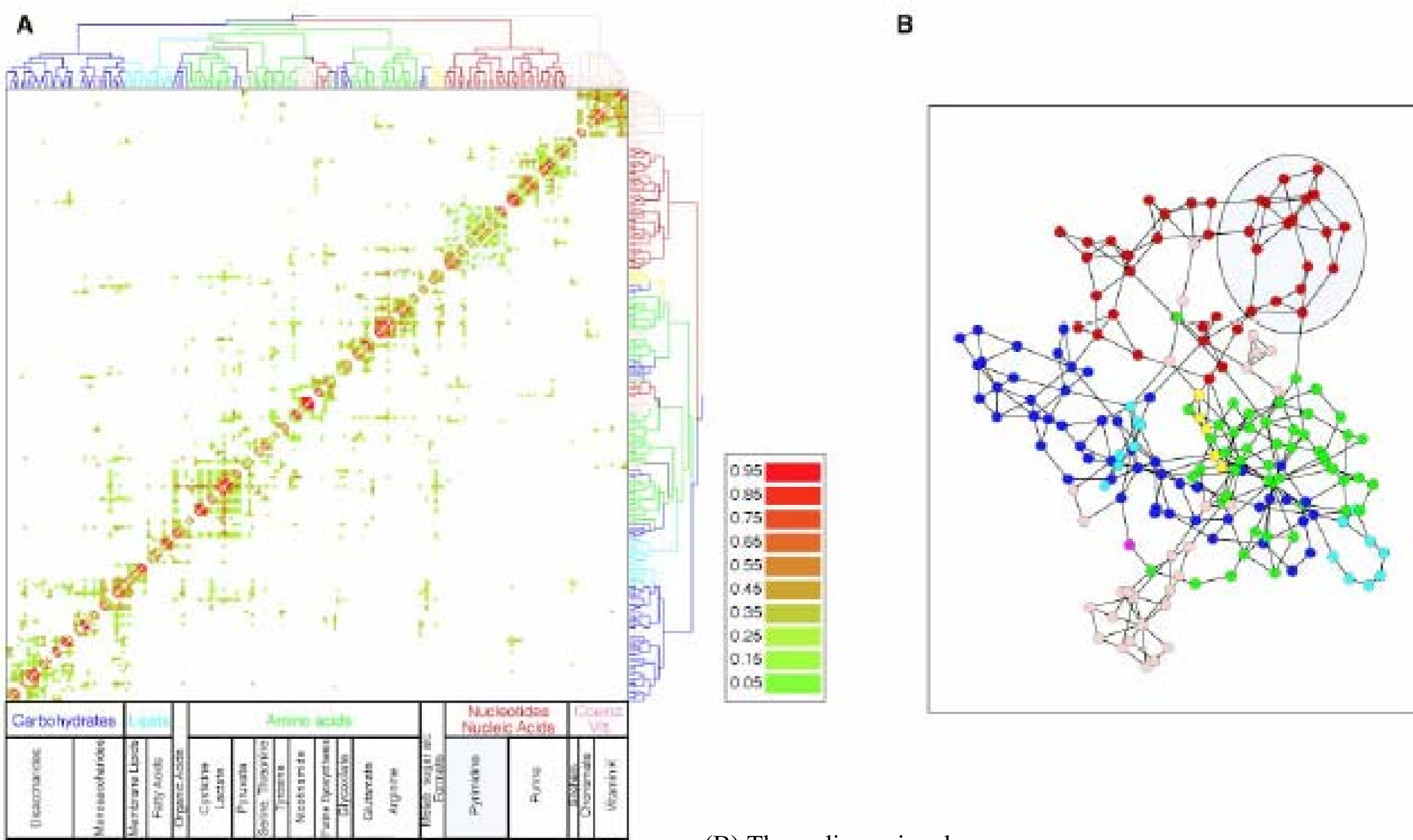
Hierarchical Organization of Modularity in Metabolic Networks

E. Ravasz, A. L. Somera,
D. A. Mongru, Z. N. Oltvai,
A.-L. Baraba´si
SCIENCE VOL 297
30 AUGUST 2002 p 1551

Spatially or chemically isolated functional modules composed of several cellular components and carrying discrete functions are considered fundamental building blocks of cellular organization, but their presence in highly integrated biochemical networks lacks quantitative support. Here, we show that the metabolic networks of 43 distinct organisms are organized into many small, highly connected topologic modules that combine in a hierarchical manner into larger, less cohesive units, with their number and degree of clustering following a power law. Within *Escherichia coli*, the uncovered hierarchical modularity closely overlaps with known metabolic functions. The identified network architecture may be generic to system-level cellular organization.



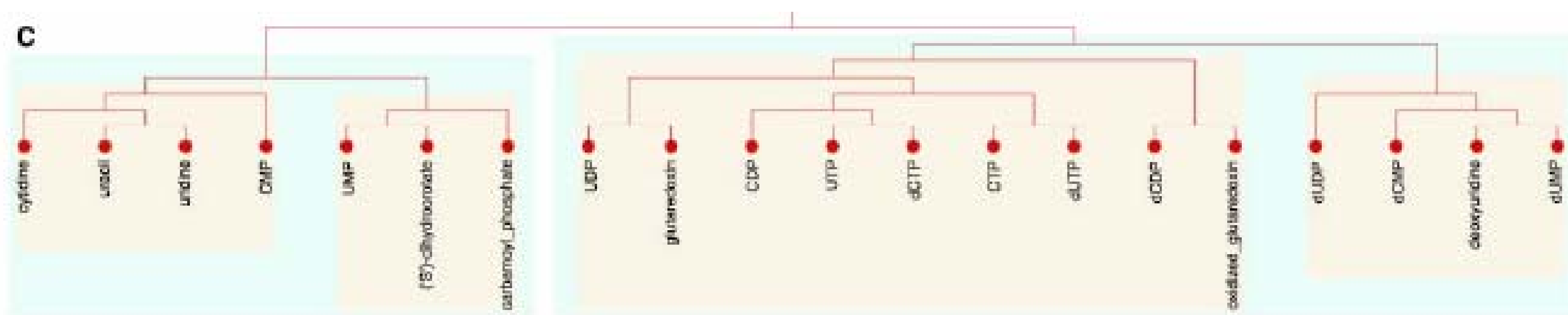
A. Scale free
B. Modular, not scale free
C. Nested modular, scale free



(A) The topologic overlap matrix corresponding to E. coli metabolism, together with the corresponding hierarchical tree (top) that quantifies the relation between the different modules.

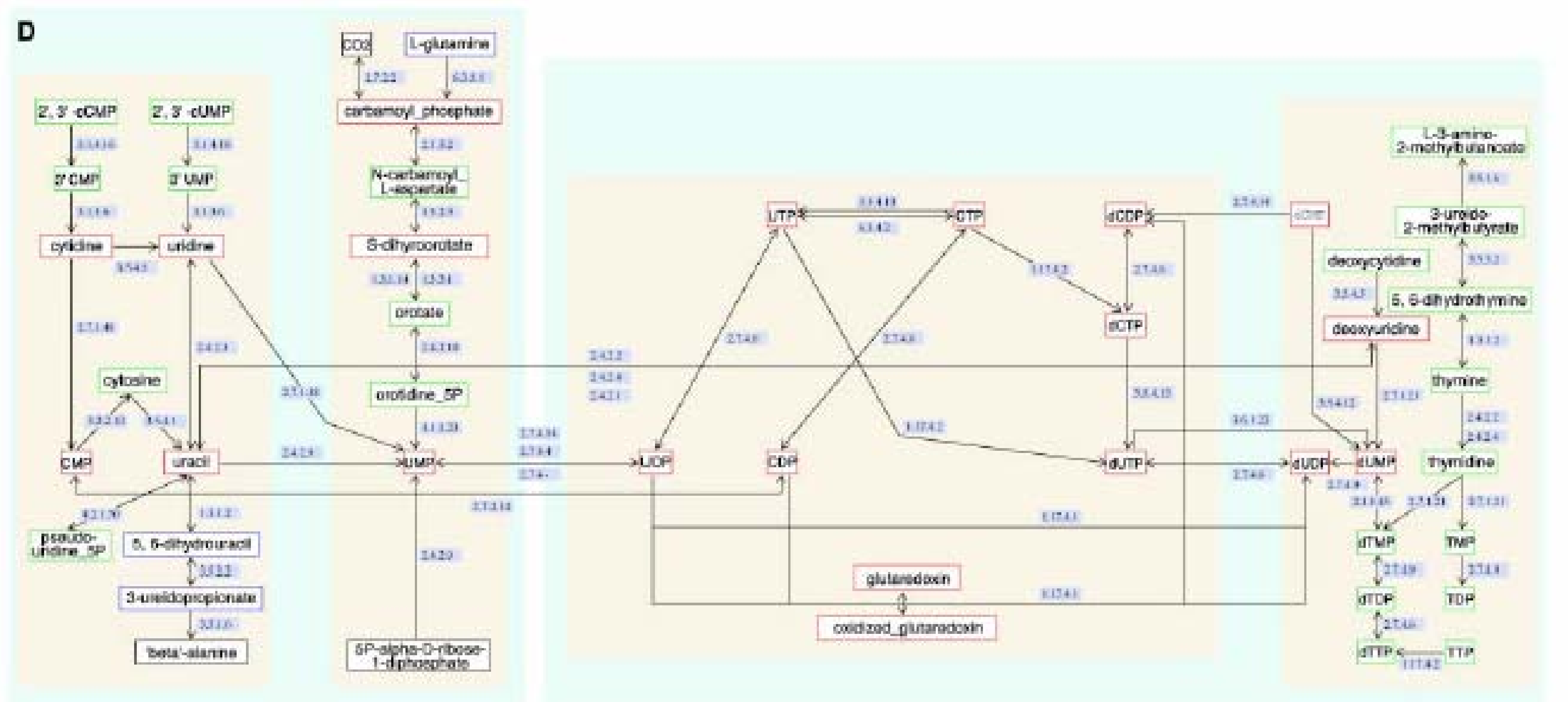
(B) Three-dimensional representation of the reduced E. coli metabolic network. Each node is color coded by the predominant biochemical class to which it belongs and is identical to the color code applied to the branches of the tree shown in (A).

Hierarchy of Metabolism



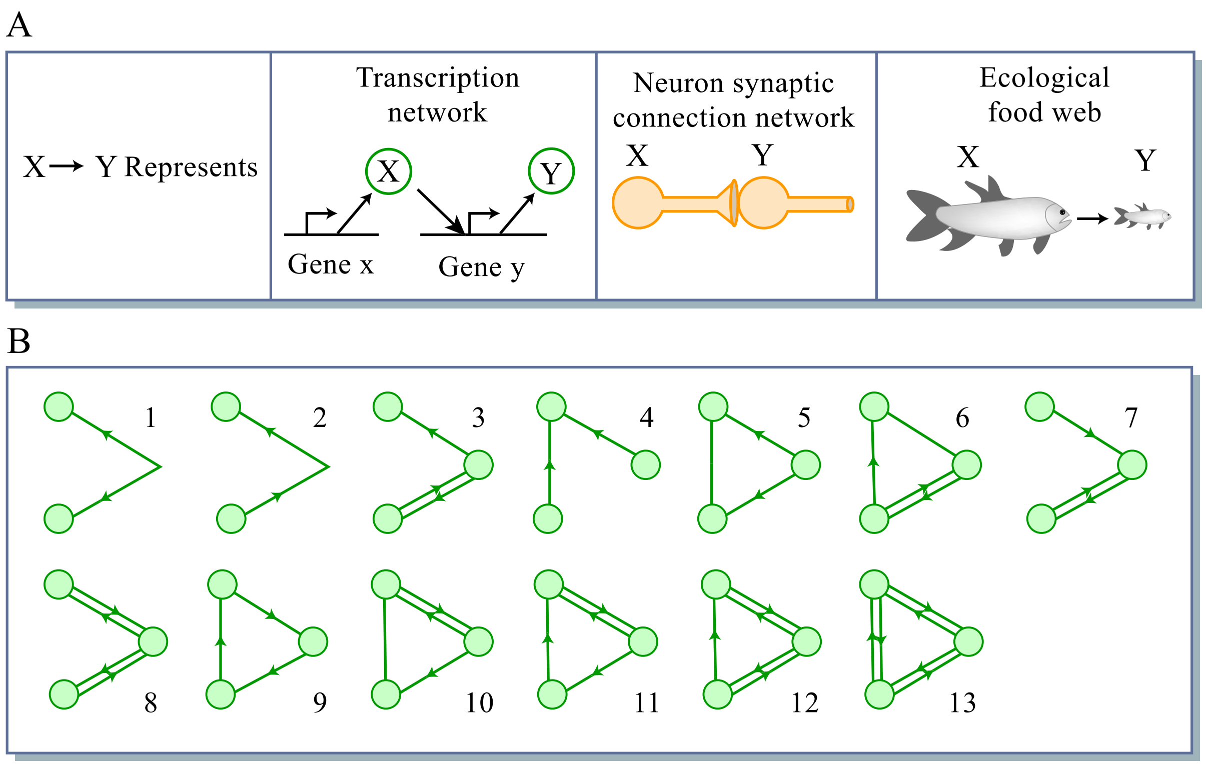
(C) Enlarged view of the substrate module of pyrimidine metabolism. The colored boxes denote the first two levels of the three levels of nested modularity suggested by the hierarchical tree.

Clustering of Metabolic Reactions



(D)
A detailed diagram of the metabolic reactions that surround and incorporate the pyrimidine metabolic module. Red-outlined boxes denote the substrates directly appearing in the reduced metabolism and the tree shown in (C).

“Motifs” in Networks

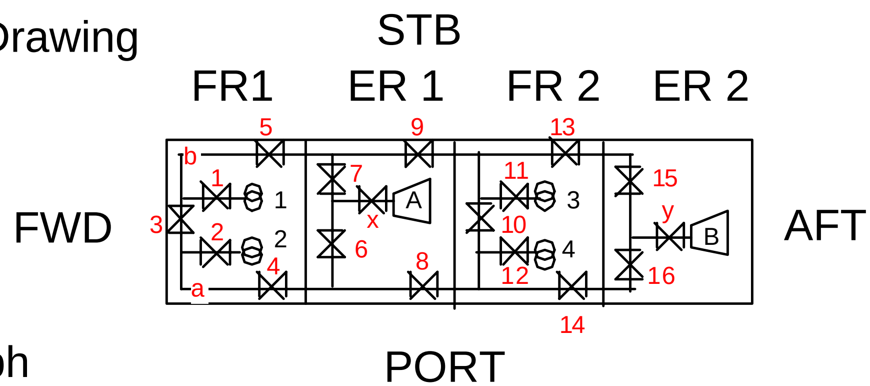


(A) Examples of interactions represented by directed edges between nodes in some of the networks used for the present study. These networks go from the scale of biomolecules (transcription factor protein X binds regulatory DNA regions of a gene to regulate the production rate of protein Y), through cells (neuron X is synaptically connected to neuron Y), to organisms (X feeds on Y). (B) All 13 types of three-node connected subgraphs.

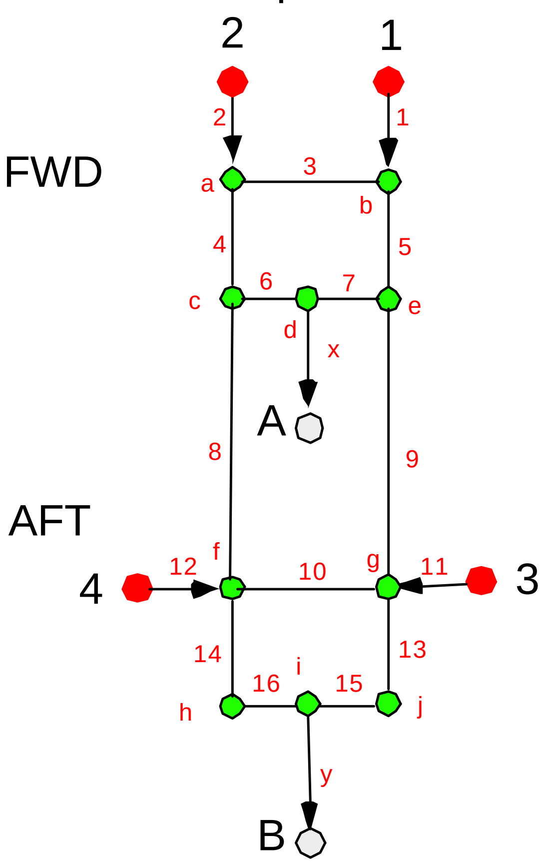
Figure by MIT OCW.

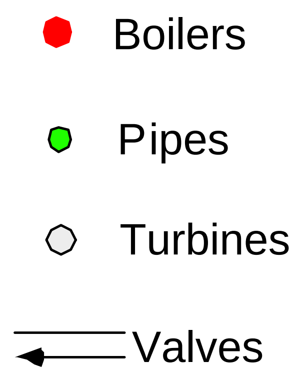
Ship Steam System

Mechanical Drawing



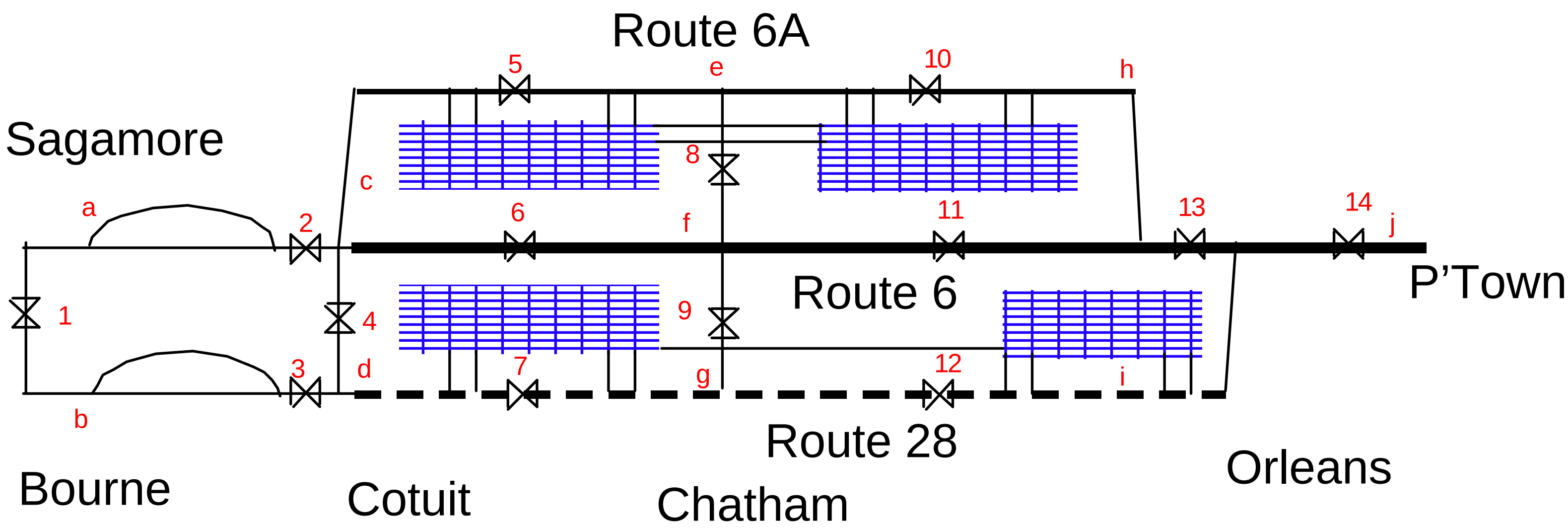
Network Graph

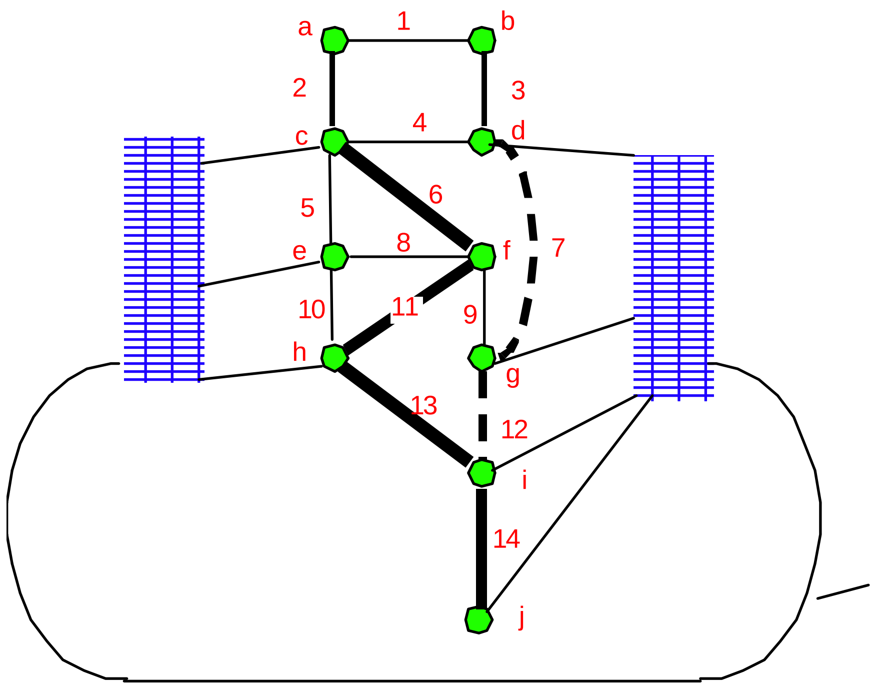




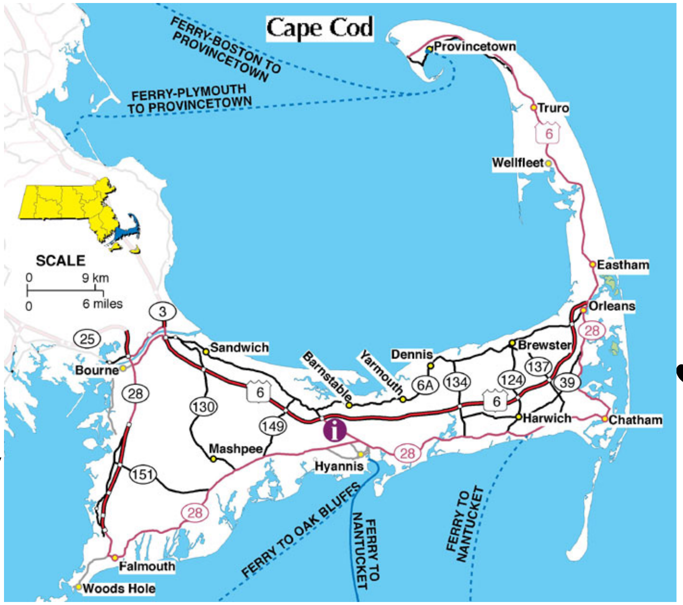
Note: The network graph is the dual of the mechanical drawing

Cape Cod Roadways



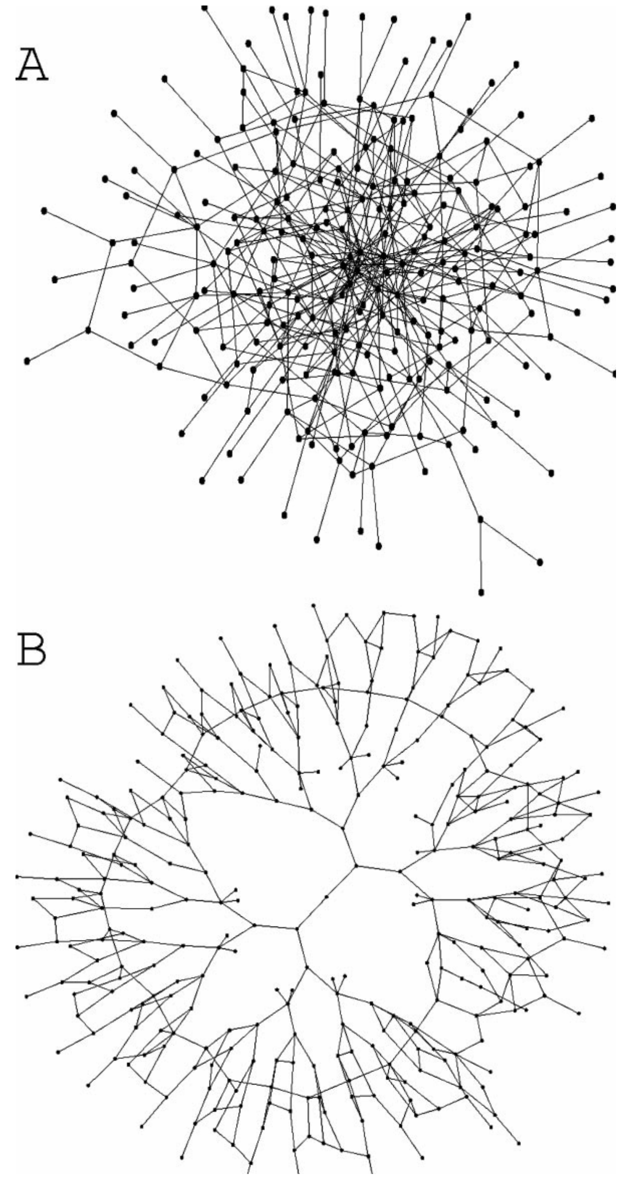


Many like this



Network Analysis of Electric Circuits

This article does not use the word planar



(a) A lowly clustered logic circuit having C=0.0013<C^{rand}=0.015 and d=4.33=d^{rand}=4.22. The graph has N=236 vertices and (k)=3.64.

(b) A highly clustered logic circuit having c=0.053>C^{rand}=0.0099 and d=5.06=d^{rand}=4.99. The graph has N=320 vertices, and (k)=3.175.

Topology of technology graphs: Small world patterns in electronic circuits,ÓRamon Ferrer i Cancho, Christiaan Janssen, and Ricard V. Sole, PHYSICAL REVIEW E, VOLUME 64 046119 Š 1 thru - 5

Possible Data Sets

- Western Power Grid
- Example at the end of Alexander *Notes on the Synthesis of Form*
- Most any road map
- Coauthors, metabolites
- Pajek web site (huge networks)

Tutorial Book

- http://faculty.ucr.edu/~hanneman/nettext/index.html
- Linked to UCINET software
  http://www.analytictech.com/ucinet/ucinet.htm

Resources, with search paths

- Google>graph theory>
  - http://people.freenet.de/Emden-Weinert/graphs.html
  - http://www.c3.lanl.gov/mega-math/gloss/graph/gr.html
  - http://www.utm.edu/departments/math/graph/
  - http://mathworld.wolfram.com/topics/GraphTheory.html
- Google>social science network analysis>
  - http://www.sfu.ca/~insna/INSNA/soft_inf.html
  - http://www.research.att.com/sw/tools/graphviz/ (found on INSNA/soft_inf, software toolkit for drawing graphs and networks)
- Google> graph theory analysis software>
  - http://www-personal.umich.edu/~mejn/courses/2004/cscs535/ (Mark Newman's course at U of M)
    - http://eclectic.ss.uci.edu/~drwhite/Anthro179a/SocialDynamics02.html > http://mlab48.itc.uci.edu/~eclectic/drwhite/Complexity/SpecialIssue.htm
  - http://directory.google.com/Top/Science/Math/Combinatorics/Software/Graph_Drawing/

More Resources

- Google> graph theory analysis software>
  – http://www.ececs.uc.edu/~berman/gnat/ (a research group)
  – http://mathforum.org/library/topics/graph_theory/?keyid=10077171&start_at=51&num_to_see=50 (The Math Forum @ Drexel)
  – http://www.ai.mit.edu/~murphyk/Bayes/bayes.html (Bayesian belief networks)
  – http://www.math.niu.edu/~rusin/known-math/index/05-XX.html (combinatorics) > http://www.math.niu.edu/~rusin/known-math/index/05CXX.html (graph theory)
  – http://www.indiana.edu/%7Ecortex/connectivity_toolbox.html (matlab toolbox)
  – http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=3941&objectType=file (another matlab toolbox)
  – http://jung.sourceforge.net (a java toolkit)
  – http://www-personal.umich.edu/~mejn/pubs.html (Mark Newman's publications)

More

- http://nicomedia.math.upatras.gr/courses/mnets/index_en.html (Univ of Patras course with readings)
- http://web.mit.edu/knh/www/11531.html (another MIT course)
- http://en.wikipedia.org/wiki/Small_world_phenomenon#The_scale-free_network_model

Prominent Network Researchers

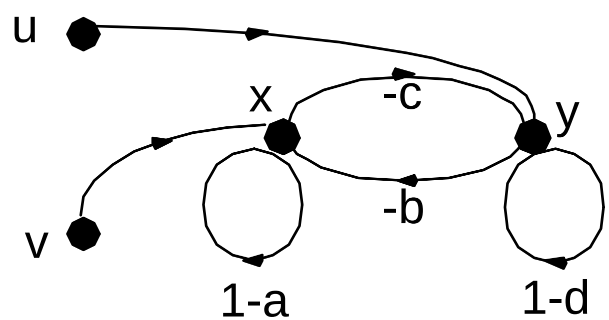
- Barabasi http://www.nd.edu/~alb/
- Doyle http://www.cds.caltech.edu/~doyle/
  - http://www.cds.caltech.edu/~doyle/CmplxNets/
  - http://hot.caltech.edu/
- Newman http://www-personal.umich.edu/~mejn/
- Strogatz http://tam.cornell.edu/Strogatz.html
- Watts http://smallworld.columbia.edu/watts.html

Backup

Graphs and Matrices

- A graph can be converted to a matrix and vice versa: node-node, node-arc, arc-arc
- A graph where an arc links only 2 nodes is equivalent to a 2 dimensional matrix
- A graph where an arc links *n* nodes is equivalent to an *n* dimensional matrix
- When form or structural patterns must be observed, matrices may be the better representation, especially if there are very many nodes and arcs
- MATLAB is applicable

Link Between Networks and Linear Algebra



$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

"Solve" for *x* and *y* given *u* and *v* by putting voltages on the *u* and *v* terminals and reading the voltages on the *x* and *y* terminals

"Scale Free" and "Scaling"

- Scale free refers to distributions of degrees in the degree distribution that are highly varied rather than being bunched around a particular value
- Scale free also means that there is no absolute value in the distribution, but rather only a ratio

In this stochastic context, a random variable X or its corresponding distribution function F is said to follow a *power law* or is *scaling* with index $\alpha > 0$ if, as $x \to \infty$,

$$P[X > x] = 1 - F(x) \approx cx^{-\alpha}, \quad (2)$$

for some constant $0 < c < \infty$ and a *tail index* $\alpha > 0$. Here, we write $f(x) \approx g(x)$ as $x \to \infty$ if $f(x)/g(x) \to 1$ as $x \to \infty$. For $1 < \alpha < 2$, F has infinite variance but finite mean, and for $0 < \alpha \leq 1$, F has not only infinite variance but also infinite mean. In general, all moments of F of order $\beta \geq \alpha$ are infinite. Since relationship (2) implies $\log(P[X > x]) \approx \log(c) - \alpha \log(x)$, doubly logarithmic plots of x versus $1 - F(x)$ yield straight lines of slope $-\alpha$, at least for large x. Well-known examples of power law

For a graph with n vertices, let $d_i = \deg(i)$ denote the degree of node i, $1 \leq i \leq n$, and call $D = \{d_1, d_2, \ldots, d_n\}$ the *degree sequence* of the graph, again assumed without loss of generality always to be ordered $d_1 \geq d_2 \geq \ldots \geq d_n$. We will say a graph has *scaling degree sequence* D (or *D is scaling*) if for all $1 \leq k \leq n_s \leq n$, D satisfies a *power law size-rank relationship* of the form $k\, d_k^{\alpha} = c$, where $c > 0$ and $\alpha > 0$ are constants, and where n_s determines the range of scaling [61]. Since this definition is simply a graph-specific version of (1) that allows for deviations from the power law relationship for nodes with low connectivity, we again recognize that doubly logarithmic plots of d_k versus k yield straight lines of slope $-\alpha$, at least for large d_k values.

Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications*

Whitney

The Aura of Scale Free

- From ~1998 to 2004 there was a frenzy of publication in which one system after another was "discovered" to be scale free
- It was claimed that these systems must have some underlying common elements or principles
- The reality may be less exciting: there are many systems with high variability but they have not gotten much attention before.