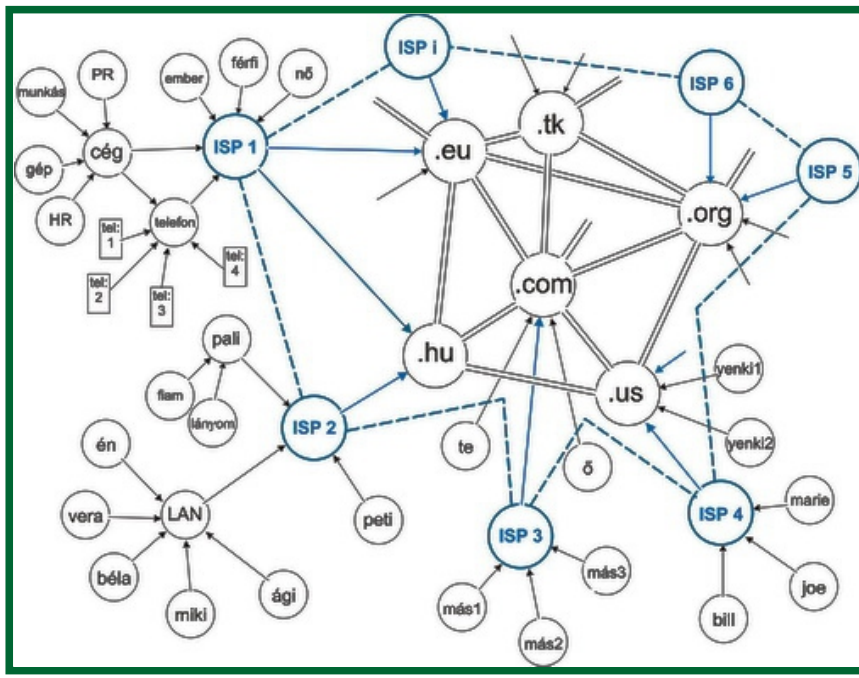
PETER G. GYARMATI

SOME WORDS ABOUT NETWORKS



TCC COMPUTER STUDIO

Peter G. Gyarmati

Some words about Networks

SOME WORDS ABOUT NETWORKS

Compiled by

Peter G. Gyarmati

TCC COMPUTER STUDIO

2011.

Some articles of this compilation originated from different sources.
Unfortunately I could not list of them.
Anyway, I have to express my thanks to all the contributors making possible this compilation.
I also have to express thanks to them in the name of all the hopeful Readers.

Imagine a world in which every single human being can freely share in the sum of all knowledge.

For additional information and updates on this book, visit

www. gyarmati.tk

www.gyarmati.dr.hu

ISBN 978-963-08-1468-3

Some words about Networks

Table of contents

1. Preface

Some time ago I retired, yes, retired from any tenure, curriculum, examination, and other everyday obligations, by so became free for thinking, reading, researching to my delight using as many forces from my remaining as I like. Truly speaking only as many as my wife let me put to such superfluous matter like thinking. She believes that this is only a needless pulling the mouse, pressing buttons, but mainly stretching in the pampering chair, living a live of ease. From a certain point of view she has some truth as I decided to make effort to my delight as a technique of a retired. Still it is a kind of job, a research for which I had no time in my earlier life or for the sake of God I forgot.

Anyhow I do make this work hoping there will other people being interest about.

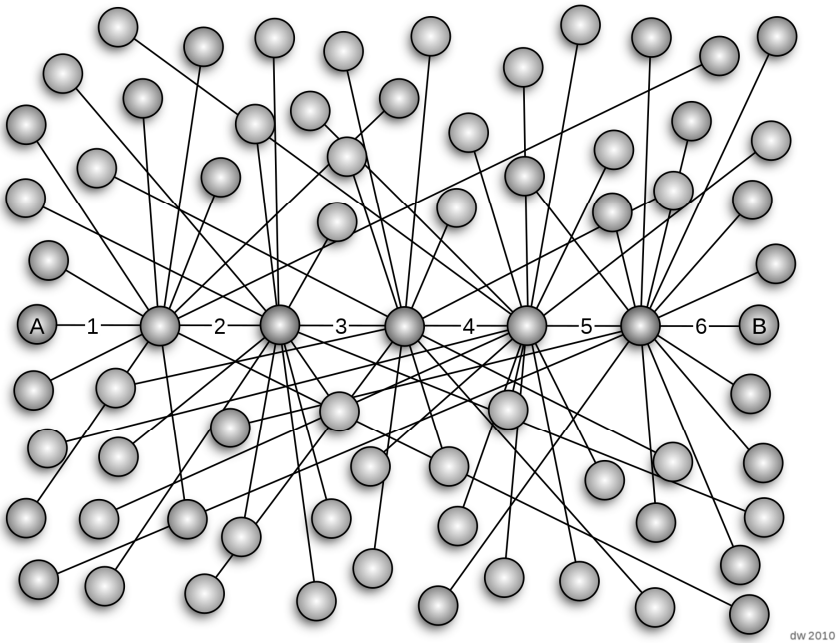
Did you dear Reader tried anytime to gather people, friends and family together to listen you, your newest discovery in your science? If yes, than you know already what a tremendous success to have one. This is how I feel now as I have, I found even more than one such community to listen to me speaking and projecting about networks, all their meaning, working, effecting to our life, and all these coming from my sitting before a computer, pulling mouse, living my ease of life and than writing all about. The other result is this little book, a kind of collected knowledge, science about the different kind of networks. It is not at all full and of course not a curriculum, but a certain way it is a guide trough the network science, understanding this new world, these new knowledge.

Now some hints how to use this book. The simplest way just read through the table of contents and the one page long first chapter. Other people could choose the more interest from the chapters. The even deeper inquirer could read trough all of them and using the reach references also.

I have to tell you again, this is a collection work, researching for the good enough and understandable texts for each topic.

I hope you will use this either obtain knowledge or use as a breviary at work.

I wish all readers turn the leaves of this book successfully.



SMALL WORLD

2. Network science

Network science is a new and emerging scientific discipline that examines the interconnections among diverse physical or engineered networks, information networks, biological networks, cognitive and semantic networks, and social networks. This field of science seeks to discover common principles, algorithms and tools that govern network behavior. The National Research Council defines Network Science as "the study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena."

The study of networks has emerged in diverse disciplines as a means of analyzing complex relational data. The earliest known paper in this field is the famous Seven Bridges of Konigsberg written by Leonhard Euler in 1736. Euler's mathematical description of vertices and edges was the foundation of graph theory, a branch of mathematics that studies the properties of pair wise relations in a network structure. The field of graph theory continued to develop and found applications in chemistry (Sylvester, 1878).

In the 1930s Jacob Moreno, a psychologist in the Gestalt tradition, arrived in the United States. He developed the sociogram and presented it to the public in April 1933 at a convention of medical scholars. Moreno claimed that "before the advent of sociometry no one knew what the interpersonal structure of a group 'precisely' looked like (Moreno, 1953). The sociogram was a representation of the social structure of a group of elementary school students. The boys were friends of boys and the girls were friends of girls with the exception of one boy who said he liked a single girl. The feeling was not reciprocated. This network representation of social structure was found so intriguing that it was printed in The New York Times (April 3, 1933, page 17). The sociogram has found many applications and has grown into the field of social network analysis.

Probabilistic theory in network science developed as an off-shoot of graph theory with Paul Erdős and Alfréd Rényi's eight famous papers on random graphs. For social networks the exponential random graph model or p* graph is a notational framework used to represent the probability space of a tie occurring in a social network. An alternate approach to network probability structures is the network probability matrix, which models the probability of edges occurring in a network, based on the historic presence

or absence of the edge in a sample of networks.

In the 1998, David Krackhardt and Kathleen Carley introduced the idea of a meta-network with the PCANS Model. They suggest that "all organizations are structured along these three domains, Individuals, Tasks, and Resources. Their paper introduced the concept that networks occur across multiple domains and that they are interrelated. This field has grown into another sub-discipline of network science called dynamic network analysis.

More recently other network science efforts have focused on mathematically describing different network topologies. Duncan Watts reconciled empirical data on networks with mathematical representation, describing the small-world network. Albert-László Barabási and Reka Albert developed the scale-free network which is a loosely defined network topology that contains hub vertices with many connections, which grow in a way to maintain a constant ratio in the number of the connections versus all other nodes. Although many networks, such as the internet, appear to maintain this aspect, other networks have long tailed distributions of nodes that only approximate scale free ratios.

Today, network science is an exciting and growing field. Scientists from many diverse fields are working together. Network science holds the promise of increasing collaboration across disciplines, by sharing data, algorithms, and software tools.

3. Network theory

Network theory is an area of computer science and network science and part of graph theory. It has application in many disciplines including particle physics, computer science, biology, economics, operations research, and sociology. Network theory concerns itself with the study of graphs as a representation of either symmetric relations or, more generally, of asymmetric relations between discrete objects. Applications of network theory include logistical networks, the World Wide Web, gene regulatory networks, metabolic networks, social networks, epistemological networks, etc. See list of network theory topics for more examples.

Network optimization

Network problems that involve finding an optimal way of doing something are studied under the name of combinatorial optimization. Examples include network flow, shortest path problem, transport problem, transshipment problem, location problem, matching problem, assignment problem, packing problem, routing problem, Critical Path Analysis and PERT (Program Evaluation & Review Technique).

Network analysis

Social network analysis

Social network analysis maps relationships between individuals in social networks.[1] Such individuals are often persons, but may be groups (including cliques and cohesive blocks), organizations, nation states, web sites, or citations between scholarly publications (scientometrics).

Network analysis, and its close cousin traffic analysis, has significant use in intelligence. By monitoring the communication patterns between the network nodes, its structure can be established. This can be used for uncovering insurgent networks of both hierarchical and leaderless nature.

Biological network analysis

With the recent explosion of publicly available high throughput biological data, the analysis of molecular networks has gained significant interest. The type of analysis in this content are closely related to social network analysis,

but often focusing on local patterns in the network. For example network motifs are small subgraphs that are over-represented in the network. Activity motifs are similar over-represented patterns in the attributes of nodes and edges in the network that are over represented given the network structure.

Link analysis

Link analysis is a subset of network analysis, exploring associations between objects. An example may be examining the addresses of suspects and victims, the telephone numbers they have dialed and financial transactions that they have partaken in during a given timeframe, and the familial relationships between these subjects as a part of police investigation.

Link analysis here provides the crucial relationships and associations between very many objects of different types that are not apparent from isolated pieces of information.

Computer-assisted or fully automatic computer-based link analysis is increasingly employed by banks and insurance agencies in fraud detection, by telecommunication operators in telecommunication network analysis, by medical sector in epidemiology and pharmacology, in law enforcement investigations, by search engines for relevance rating (and conversely by the spammers for spamdexing and by business owners for search engine optimization), and everywhere else where relationships between many objects have to be analyzed.

Web link analysis

Several Web search ranking algorithms use link-based centrality metrics, including (in order of appearance) Marchiori's Hyper Search, Google's PageRank, Kleinberg's HITS algorithm, and the TrustRank algorithm. Link analysis is also conducted in information science and communication science in order to understand and extract information from the structure of collections of web pages. For example the analysis might be of the interlinking between politicians' web sites or blogs.

Centrality measures

Information about the relative importance of nodes and edges in a graph can be obtained through centrality measures, widely used in disciplines like sociology. For example, eigenvector centrality uses the eigenvectors of the adjacency matrix to determine nodes that tend to be frequently visited.

Spread of content in networks

Content in a complex network can spread via two major methods: conserved spread and non-conserved spread.[2] In conserved spread, the total amount of content that enters a complex network remains constant as it passes through. The model of conserved spread can best be represented by a pitcher containing a fixed amount of water being poured into a series of funnels connected by tubes . Here, the pitcher represents the original source and the water is the content being spread. The funnels and connecting tubing represent the nodes and the connections between nodes, respectively. As the water passes from one funnel into another, the water disappears instantly from the funnel that was previously exposed to the water. In non-conserved spread, the amount of content changes as it enters and passes through a complex network. The model of non-conserved spread can best be represented by a continuously running faucet running through a series of funnels connected by tubes . Here, the amount of water from the original source is infinite. Also, any funnels that have been exposed to the water continue to experience the water even as it passes into successive funnels. The non-conserved model is the most suitable for explaining the transmission of most infectious diseases.

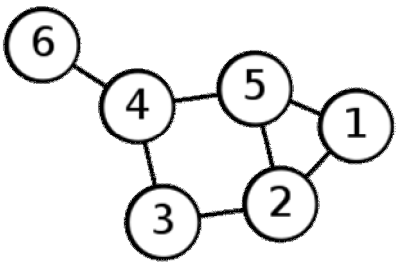
Software implementations

- Orange, a free data mining software suite, module http://orange.biolab.si/doc/modules/orngNetwork.htm
- http://pajek.imfm.si/doku.php program for (large) network analysis and visualization

References

1. Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.
2. Newman, M., Barabási, A.-L., Watts, D.J. [eds.] (2006) The Structure and Dynamics of Networks. Princeton, N.J.: Princeton University Press.

4. Graph theory



In mathematics and computer science, **graph theory** is the study of *graphs*: mathematical structures used to model pair wise relations between objects from a certain collection. A "graph" in this context refers to a collection of vertices or 'nodes' and a collection of *edges* that connect pairs of vertices. A graph may be *undirected*, meaning that there is no distinction between the two vertices associated with each edge, or its edges may be *directed* from one vertex to another; see graph (mathematics) for more detailed definitions and for other variations in the types of graphs that are commonly considered. The graphs studied in graph theory should not be confused with "graphs of functions" and other kinds of graphs.

History

The Konigsberg Bridge problem

The paper written by Leonhard Euler on the *Seven Bridges of Konigsberg* and published in 1736 is regarded as the first paper in the history of graph theory. This paper, as well as the one written by Vandermonde on the *knight problem,* carried on with the *analysis situs* initiated by Leibniz. Euler's formula relating the number of edges, vertices, and faces of a convex polyhedron was studied and generalized by Cauchy and L'Huillier, and is at the origin of topology.

More than one century after Euler's paper on the bridges of Konigsberg and while Listing introduced topology, Cayley was led by the study of particular

analytical forms arising from differential calculus to study a particular class of graphs, the *trees*. This study had many implications in theoretical chemistry. The involved techniques mainly concerned the enumeration of graphs having particular properties. Enumerative graph theory then rose from the results of Cayley and the fundamental results published by Pólya between 1935 and 1937 and the generalization of these by De Bruijn in 1959. Cayley linked his results on trees with the contemporary studies of chemical composition. The fusion of the ideas coming from mathematics with those coming from chemistry is at the origin of a part of the standard terminology of graph theory.

In particular, the term "graph" was introduced by Sylvester in a paper published in 1878 in *Nature*, where he draws an analogy between "quantic invariants" and "co-variants" of algebra and molecular diagrams.

"[...] Every invariant and co-variant thus becomes expressible by a *graph* precisely identical with a Kekuléan diagram or chemicograph. [...] I give a rule for the geometrical multiplication of graphs, *i.e.* for constructing a *graph* to the product of in- or co-variants whose separate graphs are given.

One of the most famous and productive problems of graph theory is the four color problem: "Is it true that any map drawn in the plane may have its regions colored with four colors, in such a way that any two regions having a common border have different colors?" This problem was first posed by Francis Guthrie in 1852 and its first written record is in a letter of De Morgan addressed to Hamilton the same year. Many incorrect proofs have been proposed, including those by Cayley, Kempe, and others. The study and the generalization of this problem by Tait, Heawood, Ramsey and Hadwiger led to the study of the colorings of the graphs embedded on surfaces with arbitrary genus. Tait's reformulation generated a new class of problems, the *factorization problems*, particularly studied by Petersen and Kőnig. The works of Ramsey on colorations and more specially the results obtained by Turán in 1941 was at the origin of another branch of graph theory, *extremal graph theory*.

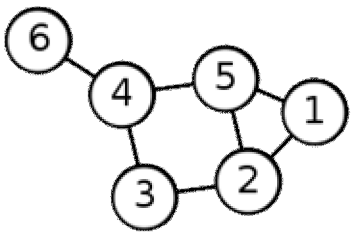
The four color problem remained unsolved for more than a century. A proof produced in 1976 by Kenneth Appel and Wolfgang Haken, which involved checking the properties of 1,936 configurations by computer, was not fully accepted at the time due to its complexity. A simpler proof considering only 633 configurations was given twenty years later by Robertson, Seymour, Sanders and Thomas.

The autonomous development of topology from 1860 and 1930 fertilized

graph theory back through the works of Jordan, Kuratowski and Whitney. Another important factor of common development of graph theory and topology came from the use of the techniques of modern algebra. The first example of such a use comes from the work of the physicist Gustav Kirchhoff, who published in 1845 his Kirchhoff's circuit laws for calculating the voltage and current in electric circuits.

The introduction of probabilistic methods in graph theory, especially in the study of Erdős and Rényi of the asymptotic probability of graph connectivity, gave rise to yet another branch, known as *random graph theory*, which has been a fruitful source of graph-theoretic results.

Vertex (graph theory)



In graph theory, a **vertex** (plural **vertices**) or **node** is the fundamental unit out of which graphs are formed: an undirected graph consists of a set of vertices and a set of edges (unordered pairs of vertices), while a directed graph consists of a set of vertices and a set of arcs (ordered pairs of vertices). From the point of view of graph theory, vertices are treated as featureless and indivisible objects, although they may have additional structure depending on the application from which the graph arises; for instance, a semantic network is a graph in which the vertices represent concepts or classes of objects.

The two vertices forming an edge are said to be its endpoints, and the edge is said to be incident to the vertices. A vertex *w* is said to be adjacent to another vertex *v* if the graph contains an edge (*v*,*w*). The neighborhood of a vertex *v* is an induced subgraph of the graph, formed by all vertices adjacent to *v*.

The degree of a vertex in a graph is the number of edges incident to it. An **isolated vertex** is a vertex with degree zero; that is, a vertex that is not an endpoint of any edge. A **leaf vertex** (also **pendant vertex**) is a vertex with degree one. In a directed graph, one can distinguish the outdegree (number of outgoing edges) from the indegree (number of incoming edges); a **source vertex** is a vertex with indegree zero, while a **sink vertex** is a vertex with outdegree zero.

A cut vertex is a vertex the removal of which would disconnect the remaining graph; a vertex separator is a collection of vertices the removal

of which would disconnect the remaining graph into small pieces. A k-vertex-connected graph is a graph in which removing fewer than *k* vertices always leaves the remaining graph connected. An independent set is a set of vertices no two of which are adjacent, and a vertex cover is a set of vertices that includes the endpoint of each edge in the graph. The vertex space of a graph is a vector space having a set of basis vectors corresponding with the graph's vertices.

A graph is vertex-transitive if it has symmetries that map any vertex to any other vertex. In the context of graph enumeration and graph isomorphism it is important to distinguish between **labeled vertices** and **unlabeled vertices**. A labeled vertex is a vertex that is associated with extra information that enables it to be distinguished from other labeled vertices; two graphs can be considered isomorphic only if the correspondence between their vertices pairs up vertices with equal labels. An unlabeled vertex is one that can be substituted for any other vertex based only on its adjacencies in the graph and not based on any additional information.

Vertices in graphs are analogous to, but not the same as, vertices of polyhedra: the skeleton of a polyhedron forms a graph, the vertices of which are the vertices of the polyhedron, but polyhedron vertices have additional structure (their geometric location) that is not assumed to be present in graph theory. The vertex figure of a vertex in a polyhedron is analogous to the neighborhood of a vertex in a graph.

In a directed graph, the forward star of a node *u* is defined as its outgoing edges. In a Graph *G* with the set of vertices *V* and the set of edges *E*, the forward star of *u* can be described as $\{(u,v) \in E\}$.

Drawing graphs

Graphs are represented graphically by drawing a dot for every vertex, and drawing an arc between two vertices if they are connected by an edge. If the graph is directed, the direction is indicated by drawing an arrow.

A graph drawing should not be confused with the graph itself (the abstract, non-visual structure) as there are several ways to structure the graph drawing. All that matters is which vertices are connected to which others by how many edges and not the exact layout. In practice it is often difficult to decide if two drawings represent the same graph. Depending on the problem domain some layouts may be better suited and easier to understand than others.

Graph-theoretic data structures

There are different ways to store graphs in a computer system. The data structure used depends on both the graph structure and the algorithm used for manipulating the graph. Theoretically one can distinguish between list and matrix structures but in concrete applications the best structure is often a combination of both. List structures are often preferred for sparse graphs as they have smaller memory requirements. Matrix structures on the other hand provide faster access for some applications but can consume huge amounts of memory.

List structures

- Incidence list

The edges are represented by an array containing pairs (tuples if directed) of vertices (that the edge connects) and possibly weight and other data. Vertices connected by an edge are said to be *adjacent*.

- Adjacency list

Much like the incidence list, each vertex has a list of which vertices it is adjacent to. This causes redundancy in an undirected graph: for example, if vertices A and B are adjacent, A's adjacency list contains B, while B's list contains A. Adjacency queries are faster, at the cost of extra storage space.

Matrix structures

- Incidence matrix

The graph is represented by a matrix of size $|V|$ (number of vertices) by $|E|$ (number of edges) where the entry [vertex, edge] contains the edge's endpoint data (simplest case: 1 - connected, 0 - not connected).

- Adjacency matrix

This is the n by n matrix A, where n is the number of vertices in the graph. If there is an edge from some vertex x to some vertex y, then the element $a_{x,y}$ is 1 (or in general the number of xy edges), otherwise it is 0. In computing, this matrix makes it easy to find subgraphs, and to reverse a directed graph.

- Laplacian matrix or Kirchhoff matrix or Admittance matrix

This is defined as $D - A$, where D is the diagonal degree matrix. It explicitly contains both adjacency information and degree information.

- Distance matrix

A symmetric n by n matrix D whose element $d_{x,y}$ is the length of a shortest

path between *x* and *y*;

if there is no such path $d_{x,y}$ = infinity,

otherwise it can be derived from powers of *A:*

$$d_{x,y} = \min\{n \mid A^n[x, y] \neq 0\}.$$

Problems in graph theory

Enumeration

There is a large literature on graphical enumeration: the problem of counting graphs meeting specified conditions. Some of this work is found in Harary and Palmer (1973).

Subgraphs, induced subgraphs, and minors

A common problem, called the subgraph isomorphism problem, is finding a fixed graph as a subgraph in a given graph. One reason to be interested in such a question is that many graph properties are *hereditary* for subgraphs, which means that a graph has the property if and only if all subgraphs have it too. Unfortunately, finding maximal subgraphs of a certain kind is often an NP-complete problem.

- Finding the largest complete graph is called the clique problem (NP-complete).
- A similar problem is finding induced subgraphs in a given graph. Again, some important graph properties are hereditary with respect to induced subgraphs, which means that a graph has a property if and only if all induced subgraphs also has it. Finding maximal induced subgraphs of a certain kind is also often NP-complete. For example,
- Finding the largest edgeless induced subgraph, or independent set, called the independent set problem (NP-complete).
- Still another such problem, the *minor containment problem*, is to find a fixed graph as a minor of a given graph. A minor or **subcontraction** of a graph is any graph obtained by taking a subgraph and contracting some (or no) edges. Many graph properties are hereditary for minors, which mean that a graph has a property if and only if all minors have it too.
- A graph is planar if it contains as a minor neither the complete bipartite graph $K_{3,3}$ (See the Three-cottage problem) nor the complete graph K_5.
- Another class of problems has to do with the extent to which various

species and generalizations of graphs are determined by their *point-deleted subgraphs*, for example:

- The reconstruction conjecture

Graph coloring

- The four-color theorem
- The strong perfect graph theorem
- The Erdős–Faber–Lovász conjecture (unsolved)
- The total coloring conjecture (unsolved)
- The list coloring conjecture (unsolved)
- The Hadwiger conjecture (graph theory) (unsolved)

Route problems

- Hamiltonian path and cycle problems
- Minimum spanning tree
- Route inspection problem (also called the "Chinese Postman Problem")
- Seven Bridges of Königsberg
- Shortest path problem
- Steiner tree
- Three-cottage problem
- Traveling salesman problem (NP-complete)

Network flow

There are numerous problems arising especially from applications that have to do with various notions of flows in networks, for example: Max flow min cut theorem

Visibility graph problems

- Museum guard problem

Covering problems

Covering problems are specific instances of subgraph-finding problems, and they tend to be closely related to the clique problem or the independent set problem.

- Set cover problem
- Vertex cover problem

Graph classes

Many problems involve characterizing the members of various classes of

graphs. Overlapping significantly with other types in this list, this type of problem includes, for instance:

- Enumerating the members of a class
- Characterizing a class in terms of forbidden substructures
- Ascertaining relationships among classes (e.g., does one property of graphs imply another)
- Finding efficient algorithms to decide membership in a class
- Finding representations for members of a class

Applications

Applications of graph theory are primarily, but not exclusively, concerned with labeled graphs and various specializations of these.

Structures that can be represented as graphs are ubiquitous, and many problems of practical interest can be represented by graphs. The link structure of a website could be represented by a directed graph: the vertices are the web pages available at the website and a directed edge from page *A* to page *B* exists if and only if *A* contains a link to *B*. A similar approach can be taken to problems in travel, biology, computer chip design, and many other fields. The development of algorithms to handle graphs is therefore of major interest in computer science. There, the transformation of graphs is often formalized and represented by graph rewrite systems. They are either directly used or properties of the rewrite systems(e.g. confluence) are studied.

A graph structure can be extended by assigning a weight to each edge of the graph. Graphs with weights, or weighted graphs, are used to represent structures in which pair wise connections have some numerical values. For example if a graph represents a road network, the weights could represent the length of each road. A digraph with weighted edges in the context of graph theory is called a network.

Networks have many uses in the practical side of graph theory, network analysis (for example, to model and analyze traffic networks). Within network analysis, the definition of the term "network" varies, and may often refer to a simple graph.

Many applications of graph theory exist in the form of network analysis. These split broadly into three categories:

1. First, analysis to determine structural properties of a network, such as the distribution of vertex degrees and the diameter of the graph. A vast

number of graph measures exist, and the production of useful ones for various domains remains an active area of research.

2. Second, analysis to find a measurable quantity within the network, for example, for a transportation network, the level of vehicular flow within any portion of it.
3. Third, analysis of dynamical properties of networks.

Graph theory is also used to study molecules in chemistry and physics. In condensed matter physics, the three dimensional structure of complicated simulated atomic structures can be studied quantitatively by gathering statistics on graph-theoretic properties related to the topology of the atoms. For example, Franzblau's shortest-path (SP) rings. In chemistry a graph makes a natural model for a molecule, where vertices represent atoms and edges bonds. This approach is especially used in computer processing of molecular structures, ranging from chemical editors to database searching.

Graph theory is also widely used in sociology as a way, for example, to measure actors' prestige or to explore diffusion mechanisms, notably through the use of social network analysis software.

Likewise, graph theory is useful in biology and conservation efforts where a vertex can represent regions where certain species exist (or habitats) and the edges represent migration paths, or movement between the regions. This information is important when looking at breeding patterns or tracking the spread of disease, parasites or how changes to the movement can affect other species.

Related topics

- Graph property
- Algebraic graph theory
- Conceptual graph
- Data structure
- Disjoint-set data structure
- Entitative graph
- Existential graph
- Graph data structure
- Graph algebras
- Graph automorphism
- Graph coloring
- Graph database
- Graph drawing
- Graph equation
- Graph rewriting
- Intersection graph
- Logical graph
- Loop
- Null graph
- Perfect graph
- Quantum graph
- Spectral graph theory
- Strongly regular graphs
- Symmetric graphs

- Tree data structure

Algorithms

- Bellman-Ford algorithm
- Dijkstra's algorithm
- Ford-Fulkerson algorithm
- Kruskal's algorithm
- Nearest neighbor algorithm
- Prim's algorithm
- Depth-first search
- Breadth-first search

Subareas

- Algebraic graph theory
- Geometric graph theory
- External graph theory
- Probabilistic graph theory
- Topological graph theory

Related areas of mathematics

- Combinatorics
- Group theory
- Knot theory
- Ramsey theory

Prominent graph theorists

- Berge, Claude
- Bollobás, Béla
- Chung, Fan
- Dirac, Gabriel Andrew
- Erdős, Paul
- Euler, Leonhard
- Faudree, Ralph
- Golumbic, Martin
- Graham, Ronald
- Harary, Frank
- Heawood, Percy John
- Kőnig, Dénes
- Lovász, László
- Nešetřil, Jaroslav
- Rényi, Alfréd
- Ringel, Gerhard
- Robertson, Neil
- Seymour, Paul
- Szemerédi, Endre
- Thomas, Robin
- Thomassen, Carsten
- Turán, Pál
- Tutte, W. T.

References

1. Biggs, N.; Lloyd, E. and Wilson, R. (1986). *Graph Theory, 1736-1936*. Oxford University Press.
2. Cauchy, A.L. (1813). "Recherche sur les polyèdres - premier mémoire". *Journal de l'Ecole Polytechnique* 9 (Cahier 16): 66–86.
3. L'Huillier, S.-A.-J. (1861). "Mémoire sur la polyèdrométrie". *Annales de Mathématiques* **3**: 169–189.
4. Cayley, A. (1875). "Ueber die Analytischen Figuren, welche in der

Mathematik Bäume genannt werden und ihre Anwendung auf die Theorie chemischer Verbindungen". *Berichte der deutschen Chemischen Gesellschaft* **8**: 1056–1059. doi:10.1002/cber.18750080252.
5. John Joseph Sylvester (1878), *Chemistry and Algebra.* Nature, volume 17, page 284. doi:10.1038/017284a0. Online version accessed on 2009-12-30.
6. Appel, K. and Haken, W. (1977). "Every planar map is four colorable. Part I. Discharging". *Illinois J. Math.* **21**: 429–490.
7. Appel, K. and Haken, W. (1977). "Every planar map is four colorable. Part II. Reducibility". *Illinois J. Math.* **21**: 491–567.
8. Robertson, N.; Sanders, D.; Seymour, P. and Thomas, R. (1997). "The four color theorem". *Journal of Combinatorial Theory Series B* **70**: 2–44. doi:10.1006/jctb.1997.1750.
9. Berge, Claude (1958), *Théorie des graphes et ses applications*, Collection Universitaire de Mathématiques, **II**, Paris: Dunod. English edition, Wiley 1961; Methuen & Co, New York 1962; Russian, Moscow 1961; Spanish, Mexico 1962; Rumanian, Bucharest 1969; Chinese, Shanghai 1963; Second printing of the 1962 first English edition, Dover, New York 2001.
10. Biggs, N.; Lloyd, E.; Wilson, R. (1986), *Graph Theory, 1736–1936*, Oxford University Press.
11. Bondy, J.A.; Murty, U.S.R. (2008), *Graph Theory*, Springer, ISBN 978-1-84628-969-9.
12. Chartrand, Gary (1985), *Introductory Graph Theory*, Dover, ISBN 0-486-24775-9.
13. Gibbons, Alan (1985), *Algorithmic Graph Theory*, Cambridge Univ. Press.
14. Golumbic, Martin (1980), *Algorithmic Graph Theory and Perfect Graphs*, Academic Press.
15. Harary, Frank (1969), *Graph Theory*, Reading, MA: Addison-Wesley.
16. Harary, Frank; Palmer, Edgar M. (1973), *Graphical Enumeration*, New York, NY: Academic Press.
17. Mahadev, N.V.R.; Peled, Uri N. (1995), *Threshold Graphs and Related Topics*, North-Holland.
18. Gallo, Giorgio; Pallotino, Stefano (December 1988). "Shortest Path Algorithms" (PDF). *Annals of Operations Research* (Netherlands: Springer) **13** (1): 1–79. doi:10.1007/BF02288320. http://www.springerlink.com/content/awn535w405321948/. Retrieved 2008-06-18.
19. Chartrand, Gary (1985). *Introductory graph theory*. New York: Dover. ISBN 0-486-24775-9.
20. Biggs, Norman; Lloyd, E. H.; Wilson, Robin J. (1986). *Graph theory, 1736-*

1936. Oxford [Oxford shire]: Clarendon Press. ISBN 0-19-853916-9.
21. Harary, Frank (1969). *Graph theory*. Reading, Mass.: Addison-Wesley Publishing. ISBN 0-201-41033-8.
22. Harary, Frank; Palmer, Edgar M. (1973). *Graphical enumeration*. New York, Academic Press. ISBN 0-12-324245-2.

Online references

1. Graph Theory with Applications (1976) by Bondy and Murty
2. Phase Transitions in Combinatorial Optimization Problems, Section 3: Introduction to Graphs (2006) by Hartmann and Weigt
3. Digraphs: Theory Algorithms and Applications 2007 by J. Bang-Jensen and G. Gutin
4. Graph Theory, by Reinhard Diestel
5. Weisstein, Eric W., "Graph Vertex" from MathWorld.

5. Complex network

In the context of network theory, a **complex network** is a network (graph) with non-trivial topological features—features that do not occur in simple networks such as lattices or random graphs. The study of complex networks is a young and active area of scientific research inspired largely by the empirical study of real-world networks such as computer networks and social networks.

Definition

Most social, biological, and technological networks display substantial non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. Such features include a heavy tail in the degree distribution, a high clustering coefficient, assortativity or disassortativity among vertices, community structure, and hierarchical structure. In the case of directed networks these features also include reciprocity, triad significance profile and other features. In contrast, many of the mathematical models of networks that have been studied in the past, such as lattices and random graphs, do not show these features.

Two well-known and much studied classes of complex networks are scale-free networks and small-world networks, whose discovery and definition are canonical case-studies in the field. Both are characterized by specific structural features—power-law degree distributions for the former and short path lengths and high clustering for the latter. However, as the study of complex networks has continued to grow in importance and popularity, many other aspects of network structure have attracted attention as well.

The field continues to develop at a brisk pace, and has brought together researchers from many areas including mathematics, physics, biology, computer science, sociology, epidemiology, and others. Ideas from network science have been applied to the analysis of metabolic and genetic regulatory networks, the design of robust and scalable communication networks both wired and wireless, the development of vaccination strategies for the control of disease, and a broad range of other practical issues. Research on networks has seen regular publication in some of the most visible scientific journals and vigorous funding in many countries, has been the topic of conferences in a variety of different fields, and has been

the subject of numerous books both for the lay person and for the expert.

Scale-free networks

A network is named scale-free if its degree distribution, i.e., the probability that a node selected uniformly at random has a certain number of links (degree), follows a particular mathematical function called a power law. The power law implies that the degree distribution of these networks has no characteristic scale. In contrast, network with a single well-defined scale are somewhat similar to a lattice in that every node has (roughly) the same degree.

Examples of networks with a single scale include the Erdős–Rényi random graph and hypercubes. In a network with a scale-free degree distribution, some vertices have a degree that is orders of magnitude larger than the average - these vertices are often called "hubs", although this is a bit misleading as there is no inherent threshold above which a node can be viewed as a hub. If there were, then it wouldn't be a scale-free distribution!

Interest in scale-free networks began in the late 1990s with the apparent discovery of a power-law degree distribution in many real world networks such as the World Wide Web, the network of Autonomous systems (ASs), some network of Internet routers, protein interaction networks, email networks, etc. Although many of these distributions are not unambiguously power laws, their breadth, both in degree and in domain, shows that networks exhibiting such a distribution are clearly very different from what you would expect if edges existed independently and at random (a Poisson distribution). Indeed, there are many different ways to build a network with a power-law degree distribution.

The Yule process is a canonical generative process for power laws, and has been known since 1925. However, it is known by many other names due to its frequent reinvention, e.g., The Gibrat principle by Herbert Simon, the Matthew effect, cumulative advantage and, most recently, preferential attachment by Barabási and Albert for power-law degree distributions.

Networks with a power-law degree distribution can be highly resistant to the random deletion of vertices, i.e., the vast majority of vertices remain connected together in a giant component. Such networks can also be quite sensitive to targeted attacks aimed at fracturing the network quickly. When the graph is uniformly random except for the degree distribution, these critical vertices are the ones with the highest degree, and have thus been implicated in the spread of disease (natural and artificial) in social and

communication networks, and in the spread of fads (both of which are modeled by a percolation or branching process).

Small-world networks

A network is called a small-world network by analogy with the small-world phenomenon (popularly known as six degrees of separation). The small world hypothesis, which was first described by the Hungarian writer Frigyes Karinthy in 1929, and tested experimentally by Stanley Milgram (1967), is the idea that two arbitrary people are connected by only six degrees of separation, i.e. the diameter of the corresponding graph of social connections is not much larger than six. In 1998, Duncan J. Watts and Steven Strogatz published the first small-world network model, which through a single parameter smoothly interpolates between a random graph to a lattice. Their model demonstrated that with the addition of only a small number of long-range links, a regular graph, in which the diameter is proportional to the size of the network, can be transformed into a "small world" in which the average number of edges between any two vertices is very small (mathematically, it should grow as the logarithm of the size of the network), while the clustering coefficient stays large. It is known that a wide variety of abstract graphs exhibit the small-world property, e.g., random graphs and scale-free networks. Further, real world networks such as the World Wide Web and the metabolic network also exhibit this property.

In the scientific literature on networks, there is some ambiguity associated with the term "small world." In addition to referring to the size of the diameter of the network, it can also refer to the co-occurrence of a small diameter and a high clustering coefficient. The clustering coefficient is a metric that represents the density of triangles in the network. For instance, sparse random graphs have a vanishingly small clustering coefficient while real world networks often have a coefficient significantly larger. Scientists point to this difference as suggesting that edges are correlated in real world networks.

Researchers and scientists

- Réka Albert
- Luis Amaral
- Alex Arenas
- Albert-László Barabási
- Alain Barrat
- Marc Barthelemy
- Stefano Boccaletti
- Dirk Brockmann
- Guido Caldarelli
- Roger Guimerà

- Shlomo Havlin
- Jon Kleinberg
- José Mendes
- Yamir Moreno
- Adilson E. Motter
- Mark Newman
- Sidney Redner
- Steven Strogatz
- Alessandro Vespignani
- Duncan J. Watts

References

1. Albert-László Barabási, *Linked: How Everything is Connected to Everything Else*, 2004, ISBN 0-452-28439-2
2. Alain Barrat, Marc Barthelemy, Alessandro Vespignani, *Dynamical processes in complex networks*, Cambridge University Press, 2008, ISBN 978-0-521-87950-7
3. Stefan Bornholdt (Editor) and Heinz Georg Schuster (Editor), *Handbook of Graphs and Networks: From the Genome to the Internet*, 2003, ISBN 3-527-40336-1
4. Guido Caldarelli, *Scale-Free Networks* Oxford University Press, 2007, ISBN 0-19-921151-7
5. Matthias Dehmer and Frank Emmert-Streib (Eds.), "Analysis of Complex Networks: From Biology to Linguistics", Wiley-VCH, 2009, ISBN 3-527-32345-7
6. S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks: From biological networks to the Internet and WWW*, Oxford University Press, 2003, ISBN 0-19-851590-1
7. Mark Newman, Albert-László Barabási, and Duncan J. Watts, *The Structure and Dynamics of Networks*, Princeton University Press, Princeton, 2006, ISBN 978-0-691-11357-9
8. R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A statistical physics approach*, Cambridge University Press, 2004, ISBN 0-521-82698-5
9. Duncan J. Watts, *Six Degrees: The Science of a Connected Age*, Norton & Company, 2003, ISBN 0-393-04142-5
10. Duncan J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, 2003, ISBN 0-691-11704-7
11. D. J. Watts and S. H. Strogatz., *Collective dynamics of 'small-world' networks*, Nature Vol 393 (1998) 440-442
12. S. H. Strogatz, *Exploring Complex Networks*, Nature Vol 410 (2001) 268-276
13. R. Albert and A.-L. Barabási, "Statistical mechanics of complex

networks" Reviews of Modern Physics 74, (2002) 47
14. S. N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks*, Adv. Phys. 51, 1079 (2002)
15. M. E. J. Newman, The structure and function of complex networks, SIAM Review 45, 167-256 (2003)
16. A. Barabasi and E. Bonabeau, *Scale-Free Networks*, Scientific American, (May 2003), 50-59
17. S. Boccaletti et al., *Complex Networks: Structure and Dynamics*, Phys. Rep., 424 (2006), 175-308.
18. S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, *Critical phenomena in complex networks*, Rev. Mod. Phys. 80, 1275, (2008)
19. R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, "Resilience of the Internet to random breakdown" Phys. Rev. Lett. 85, 4626 (2000).

On-line references

1. Network Science — United States Military Academy - Network Science Center http://www.netscience.usma.edu/
2. Resources in Complex Networks — University of São Paulo - Institute of Physics at São Carlos http://cyvision.if.sc.usp.br/networks/
3. Cx-Nets — Complex Networks Collaboratory http://sites.google.com/site/cxnets/
4. GNET — Group of Complex Systems & Random Networks http://www2.fis.ua.pt/grupoteorico/gteorico.htm
5. UCLA Human Complex Systems Program http://www.hcs.ucla.edu/
6. New England Complex Systems Institute http://necsi.edu/
7. Barabasi Networks Group http://www.barabasilab.com/
8. Cosin Project Codes, Papers and Data on Complex Networks http://www.cosinproject.org/
9. Complex network on arxiv.org http://xstructure.inr.ac.ru/x-bin/theme3.py?level=2&index1=127691
10. Anna Nagurney's Virtual Center for Supernetworks
11. BIOREL resource for quantitative estimation of the network bias in relation to external information http://mips.helmholtz-muenchen.de/proj/biorel/
12. Complexity Virtual Laboratory (VLAB) http://vlab.infotech.monash.edu.au/
13. French computer science research group on networks http://www.complexnetworks.fr/

6. Flow network

In graph theory, a **flow network** is a directed graph where each edge has a **capacity** and each edge receives a flow. The amount of flow on an edge cannot exceed the capacity of the edge. Often in Operations Research, a directed graph is called a **network**, the vertices are called **nodes** and the edges are called **arcs**. A flow must satisfy the restriction that the amount of flow into a node equals the amount of flow out of it, except when it is a **source**, which has more outgoing flow, or **sink**, which has more incoming flow. A network can be used to model traffic in a road system, fluids in pipes, currents in an electrical circuit, or anything similar in which something travels through a network of nodes.

Definition

Suppose $G(V,E)$ is a finite directed graph in which every edge $(u,v) \in E$ has a non-negative, real-valued capacity $c(u,v)$. If $(u,v) \notin E$, we assume that $c(u,v) = 0$. We distinguish two vertices: a source s and a sink t. A flow network is a real function $f : V \times V \to \mathbb{R}$ with the following three properties for all nodes $u$and v:

Capacity constraints: $f(u,v) \leq c(u,v)$. The flow along an edge can not exceed its capacity.

Skew symmetry: $f(u,v) = -f(v,u)$. Flow from u to v must be the opposite of the from v to u.

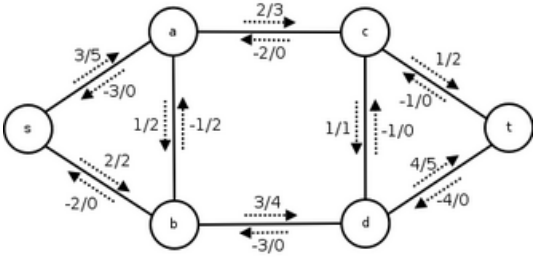
Flow conservation: $\sum_{w \in V} f(u,w) = 0$ unless $u = s$ or $w = t$. The net flow to a node is zero, except for the source, which "produces" flow, and the sink, which "consumes" flow.

Notice that $f(u,v)$ is the *net* flow from $u$to v. If the graph represents a physical network, and if there is a real capacity of, for example, 4 units from $u$to v, and a real flow of 3 units from $v$to u, we have $f(u,v) = 1$and $f(v,u) = -1$.

The **residual capacity** of an edge is $c_f(u,v) = c(u,v) - f(u,v)$. This defines a **residual network** denoted $G_f(V, E_f)$, giving the amount of *available* capacity. See that there can be an edge from u to v in the residual network, even though there is no edge from u to v in the original network. Since flows in opposite directions cancel out, *decreasing* the flow from v to u is the same as *increasing* the flow from u to v. An **augmenting path** is a path $(u_1, u_2, \ldots, u_k)$ in the residual network, where $u_1 = s$, $u_k = t$, and $c_f(u_i, u_{i+1}) > 0$. A network is at maximum flow if and only if there is no augmenting path in the residual network.

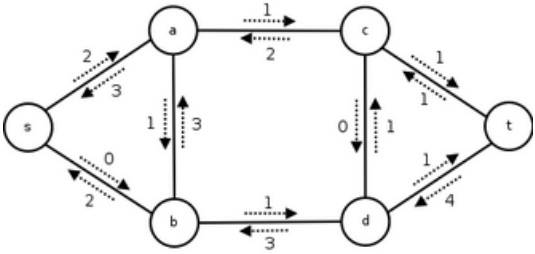
Example

A flow network showing flow and capacity.



Here you see a flow network with source labeled *s*, sink *t*, and four additional nodes. The flow and capacity is denoted *f* / *c*. Notice how the network upholds skew symmetry, capacity constraints and flow conservation. The total amount of flow from *s* to *t* is 5, which can be easily seen from the fact that the total outgoing flow from *s* is 5, which is also the incoming flow to *t*. We know that no flow appears or disappears in any of the other nodes.

Residual network for the above flow network, showing residual capacities.



Here is the residual network for the given flow. Notice how there is positive residual capacity on some edges where the original capacity is zero, for example for the edge (*d*,*c*). This flow is not a maximum flow. There is available capacity along the paths (*s*,*a*,*c*,*t*), (*s*,*a*,*b*,*d*,*t*) and (*s*,*a*,*b*,*d*,*c*,*t*), which are then the augmenting paths. The residual capacity of the first path is min(*c*(*s*,*a*) − *f*(*s*,*a*),*c*(*a*,*c*) − *f*(*a*,*c*),*c*(*c*,*t*) − *f*(*c*,*t*)) = min(5 − 3,3 − 2,2 − 1) = min(2,1,1) = 1. Notice that augmenting path (*s*,*a*,*b*,*d*,*c*,*t*) does not exist in the

original network, but you can send flow along it, and still get a legal flow.

If this is a real network, there might actually be a flow of 2 from *a* to *b*, and a flow of 1 from *b* to *a*, but we only maintain the **net** flow.

Applications

Picture a series of water pipes, fitting into a network. Each pipe is of a certain diameter, so it can only maintain a flow of a certain amount of water. Anywhere that pipes meet, the total amount of water coming into that junction must be equal to the amount going out, otherwise we would quickly run out of water, or we would have a build up of water. We have a water inlet, which is the source, and an outlet, the sink. A flow would then be one possible way for water to get from source to sink so that the total amount of water coming out of the outlet is consistent. Intuitively, the total flow of a network is the rate at which water comes out of the outlet.

Flows can pertain to people or material over transportation networks, or to electricity over electrical distribution systems. For any such physical network, the flow coming into any intermediate node needs to equal the flow going out of that node. Bollobás characterizes this constraint in terms of Kirchhoff's current law, while later authors (ie: Chartrand) mention its generalization to some conservation equation.

Flow networks also find applications in ecology: flow networks arise naturally when considering the flow of nutrients and energy between different organizations in a food web. The mathematical problems associated with such networks are quite different from those that arise in networks of fluid or traffic flow. The field of ecosystem network analysis, developed by Robert Ulanowicz and others, involves using concepts from information theory and thermodynamics to study the evolution of these networks over time..

Generalizations and specializations

The simplest and most common problem using flow networks is to find what is called the maximum flow, which provides the largest possible total flow from the source to the sink in a given graph. There are many other problems which can be solved using max flow algorithms, if they are appropriately modeled as flow networks, such as bipartite matching, the assignment problem and the transportation problem.

In a multi-commodity flow problem, you have multiple sources and sinks, and various "commodities" which are to flow from a given source to a given

sink. This could be for example various goods that are produced at various factories, and are to be delivered to various given customers through the *same* transportation network.

In a minimum cost flow problem, each edge *u*,*v* has a given cost *k*(*u*,*v*), and the cost of sending the flow *f*(*u*,*v*) across the edge is $f(u,v) \cdot k(u,v)$. The objective is to send a given amount of flow from the source to the sink, at the lowest possible price.

In a circulation problem, you have a lower bound *l*(*u*,*v*) on the edges, in addition to the upper bound *c*(*u*,*v*). Each edge also has a cost. Often, flow conservation holds for *all* nodes in a circulation problem, and there is a connection from the sink back to the source. In this way, you can dictate the total flow with *l*(*t*,*s*) and *c*(*t*,*s*). The flow *circulates* through the network, hence the name of the problem.

In a **network with gains** or **generalized network** each edge has a **gain**, a real number (not zero) such that, if the edge has gain *g*, and an amount *x* flows into the edge at its tail, then an amount *gx* flows out at the head.

References

1. Ravindra K. Ahuja, Thomas L. Magnanti, and James B. Orlin (1993). *Network Flows: Theory, Algorithms and Applications*. Prentice Hall. ISBN 0-13-617549-X.
2. Bollobás, Béla (1979). *Graph Theory: An Introductory Course*. Heidelberg: Springer-Verlag. ISBN 3-540-90399-2.
3. Chartrand, Gary & Oellermann, Ortrud R. (1993). *Applied and Algorithmic Graph Theory*. New York: McGraw-Hill. ISBN 0-07-557101-3.
4. Even, Shimon (1979). *Graph Algorithms*. Rockville, Maryland: Computer Science Press. ISBN 0-914894-21-8.
5. Gibbons, Alan (1985). *Algorithmic Graph Theory*. Cambridge: Cambridge University Press. ISBN 0-521-28881-9 ISBN 0-521-24659-8.
6. Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein (2001) [1990]. "26". *Introduction to Algorithms* (2nd edition ed.). MIT Press and McGraw-Hill. pp. 696–697. ISBN 0-262-03293-7.

7. Network diagram

A **network diagram** is a general type of diagram, which represents some kind of network. A network in general is an interconnected group or system, or a fabric or structure of fibrous elements attached to each other at regular intervals, or formally: a graph.

A network diagram is a special kind of cluster diagram, which even more general represents any cluster or small group or bunch of something, structured or not. Both the flow diagram and the tree diagram can be seen as a specific type of network diagram.

There are different types network diagrams:

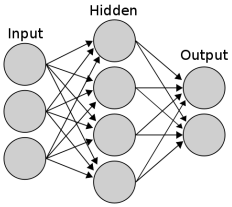
- Artificial neural network or "neural network" (NN), is a mathematical model or computational model based on biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation.
- Computer network diagram is a schematic depicting the nodes and connections amongst nodes in a computer network or, more generally, any telecommunications network.
- In project management according to Baker et al. (2003), a "network diagram is the logical representation of activities, that defines the sequence or the work of a project. It shows the path of a project, lists starting and completion dates, and names the responsibilities for each task. At a glance it explains how the work of the project goes together... A network for a simple project might consist of one or two pages, and on a larger project several network diagrams may exist.
- Project network: a general flow chart depicting the sequence in which a project's terminal elements are to be completed by showing terminal elements and their dependencies.
- PERT network (Program Evaluation and Review Technique).
- Neural network diagram: is a network or circuit of biological neurons or artificial neural networks, which are composed of artificial neurons or nodes.
- A semantic network is a network or circuit of biological neurons. The modern usage of the term often refers to artificial neural networks,

which are composed of artificial neurons or nodes.

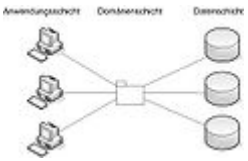
- A sociogram is a graphic representation of social links that a person has. It is a sociometric chart that plots the structure of interpersonal relations in a group situation.

This Gallery shows example drawings of network diagrams:

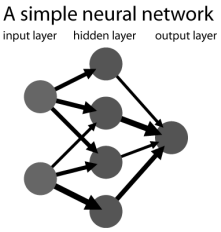
Gallery



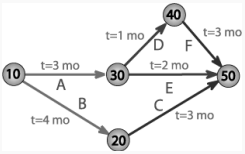
Artificial neural network



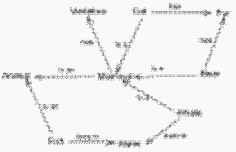
Computer network



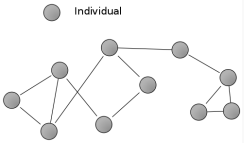
Neural network diagram



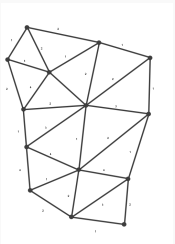
PERT diagram



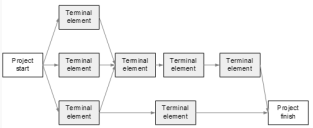
Semantic network



Sociogram



Spin network



Project network

Network topologies

In computer science the elements of a network are arranged in certain basic shapes (see figure here below):

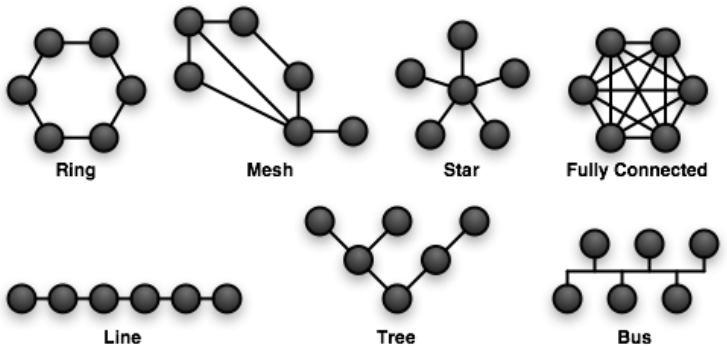


Diagram of different network topologies.

- Ring: The ring network connects each node to exactly two other nodes, forming a circular pathway for activity or signals - a ring. The interaction or data travels from node to node, with each node handling every packet.
- Mesh is a way to route data, voice and instructions between nodes. It allows for continuous connections and reconfiguration around broken or blocked paths by "hopping" from node to node until the destination is reached.
- Star: The star network consists of one central element, switch, hub or computer, which acts as a conduit to coordinate activity or transmit messages.
- Fully connected: Every node is connected to every other node.
- Line: Everything connected in a single line.
- Tree: This consists of tree-configured nodes connected to switches/concentrators, each connected to a linear bus backbone. Each hub rebroadcasts all transmissions received from any peripheral node to all peripheral nodes on the network, sometimes including the originating node. All peripheral nodes may thus communicate with all others by transmitting to, and receiving from, the central node only.

- Bus: In this network architecture a set of clients are connected via a shared communications line, called a bus.

Network theory

Network theory is an area of applied mathematics and part of graph theory. It has application in many disciplines including particle physics, computer science, biology, economics, operations research, and sociology. Network theory concerns itself with the study of graphs as a representation of either symmetric relations or, more generally, of asymmetric relations between discrete objects. Examples of which include logistical networks, the World Wide Web, gene regulatory networks, metabolic networks, social networks, epistemological networks, etc. See list of network theory topics for more examples.

Network topology

Network topology is the study of the arrangement or mapping of the elements (links, nodes, etc.) of a network, especially the physical (real) and logical (virtual) interconnections between nodes.[4]

Any particular network topology is determined only by the graphical mapping of the configuration of physical and/or logical connections between nodes. LAN Network Topology is, therefore, technically a part of graph theory. Distances between nodes, physical interconnections, transmission rates, and/or signal types may differ in two networks and yet their topologies may be identical.

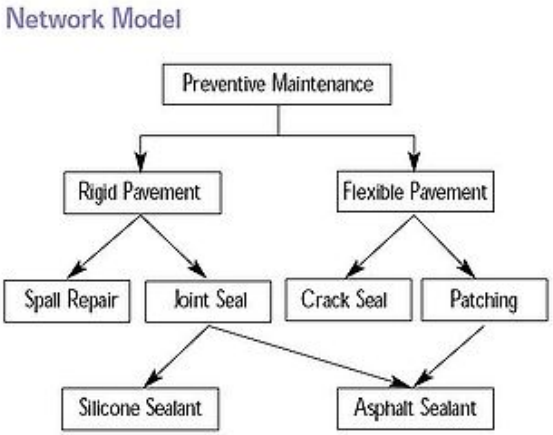
References

1. Sunny Baker, G. Michael Campbell, Kim Baker (2003). *The Complete Idiot's Guide to Project Management*. pp. 104. ISBN 0028639200.
2. Jonh F.Sowa (1987). "Semantic Networks". in Stuart C Shapiro. *Encyclopedia of Artificial Intelligence*. http://www.jfsowa.com/pubs/semnet.htm. Retrieved 2008-04-29.
3. Committee on Network Science for Future Army Applications, National Research Council (2005). *Network Science*. National Academies Press. ISBN 0309100267. http://www.nap.edu/catalog.php?record_id=11516 .
4. Groth, David Toby Skandier (2005). *'Network+ Study Guide, Fourth Edition'*. Sybex, Inc.. ISBN 0-7821-4406-3.

8. Network model (database)

The **network model** is a database model conceived as a flexible way of representing objects and their relationships. Its distinguishing feature is that the schema, viewed as a graph in which object types are nodes and relationship types are arcs, is not restricted to being a hierarchy or lattice.

The network model is a database model conceived as a flexible way of representing objects and their relationships. Its original inventor was Charles Bachman, and it was developed into a standard specification published in 1969 by the CODASYL Consortium. Where the hierarchical model structures data as a tree of records, with each record having one parent record and many children, the network model allows each record to have multiple parent and child records, forming a lattice structure.



Example of a Network Model.

The network model's original inventor was Charles Bachman, and it was developed into a standard specification published in 1969 by the CODASYL Consortium.

Overview

Where the hierarchical model structures data as a tree of records, with each record having one parent record and many children, the network model allows each record to have multiple parent and child records, forming a generalized graph structure. This property applies at two levels: the schema is a generalized graph of record types connected by relationship types (called "set types" in CODASYL), and the database itself is a generalized

graph of record occurrences connected by relationships (CODASYL "sets"). Cycles are permitted at both levels.

The chief argument in favor of the network model, in comparison to the hierarchic model, was that it allowed a more natural modeling of relationships between entities. Although the model was widely implemented and used, it failed to become dominant for two main reasons. Firstly, IBM chose to stick to the hierarchical model with semi-network extensions in their established products such as IMS and DL/I. Secondly, it was eventually displaced by the relational model, which offered a higher-level, more declarative interface. Until the early 1980s the performance benefits of the low-level navigational interfaces offered by hierarchical and network databases were persuasive for many large-scale applications, but as hardware became faster, the extra productivity and flexibility of the relational model led to the gradual obsolescence of the network model in corporate enterprise usage.

Some Well-known Network Databases

- Digital Equipment Corporation DBMS-10
- Digital Equipment Corporation DBMS-20
- Digital Equipment Corporation VAX DBMS
- Honeywell IDS (Integrated Data Store)
- IDMS (Integrated Database Management System)
- Raima Data Manager (RDM) Embedded
- RDM Server
- TurboIMAGE
- Univac DMS-1100

History

In 1969, the Conference on Data Systems Languages (CODASYL) established the first specification of the network database model. This was followed by a second publication in 1971, which became the basis for most implementations. Subsequent work continued into the early 1980s, culminating in an ISO specification, but this had little influence on products.

References

- Charles W. Bachman, *The Programmer as Navigator.* ACM Turing Award lecture, Communications of the ACM, Volume 16, Issue 11, 1973, pp. 653-658, ISSN 0001-0782, doi:10.1145/355611.362534

9. Network analysis (electrical circuits)

A network, in the context of electronics, is a collection of interconnected components. **Network analysis** is the process of finding the voltages across, and the currents through, every component in the network. There are a number of different techniques for achieving this. However, for the most part, they assume that the components of the network are all linear. The methods described in this article are only applicable to *linear* network analysis except where explicitly stated.

Definitions

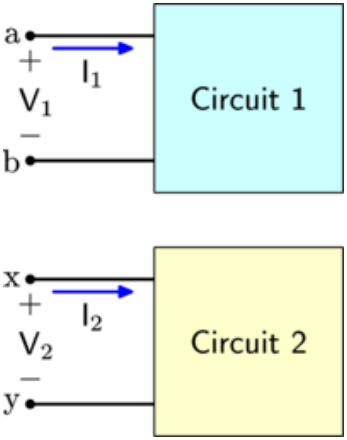
Component	A device with two or more terminals into which, or out of which, charge may flow.
Node	A point at which terminals of more than two components are joined. A conductor with a substantially zero resistance is considered to be a node for the purpose of analysis.
Branch	The component(s) joining two nodes.
Mesh	A group of branches within a network joined so as to form a complete loop.
Port	Two terminals where the current into one is identical to the current out of the other.
Circuit	A current from one terminal of a generator, through load component(s) and back into the other terminal. A circuit is, in this sense, a one-port network and is a trivial case to analyze. If there is any connection to any other circuits then a non-trivial network has been formed and at least two ports must exist. Often, "circuit" and "network" are used interchangeably, but many analysts reserve "network" to mean an idealized model consisting of ideal components.
Transfer function	The relationship of the currents and/or voltages between two ports. Most often, an input and an output port are discussed and the transfer function is described as gain or attenuation.

Component transfer function

For a two-terminal component (i.e. one-port component), the current and voltage are taken as the input and output and the transfer function will have units of impedance or admittance (it is usually a matter of arbitrary convenience whether voltage or current is considered the input). A three (or more) terminal component effectively has two (or more) ports and the transfer function cannot be expressed as a single impedance. The usual approach is to express the transfer function as a matrix of parameters. These parameters can be impedances, but there is a large number of other approaches, see two-port network.

Equivalent circuits

A useful procedure in network analysis is to simplify the network by reducing the number of components. This can be done by replacing the actual components with other notional components that have the same effect. A particular technique might directly reduce the number of components, for instance by combining impedances in series. On the other hand it might merely change the form in to one in which the components can be reduced in a later operation. For instance, one might transform a voltage generator into a current generator using Norton's theorem in order to be able to later combine the internal resistance of the generator with a parallel impedance load.



A resistive circuit is a circuit containing only resistors, ideal current sources, and ideal voltage sources. If the sources are constant (DC) sources, the result is a DC circuit. The analysis of a circuit refers to the process of solving for the voltages and currents present in the circuit. The solution principles outlined here also apply to phase analysis of AC circuits.

Two circuits are said to be **equivalent** with respect to a pair of terminals if the voltage across the terminals and current through the terminals for one network have the same relationship as the voltage and current at the terminals of the other network.

If $V_2 = V_1$ implies $I_2 = I_1$ for all (real) values of V_1, then with respect to terminals ab and xy, circuit 1 and circuit 2 are equivalent.

The above is a sufficient definition for a one-port network. For more than one port, then it must be defined that the currents and voltages between all pairs of corresponding ports must bear the same relationship. For instance, star and delta networks are effectively three port networks and hence require three simultaneous equations to fully specify their equivalence.

Impedances in series and in parallel

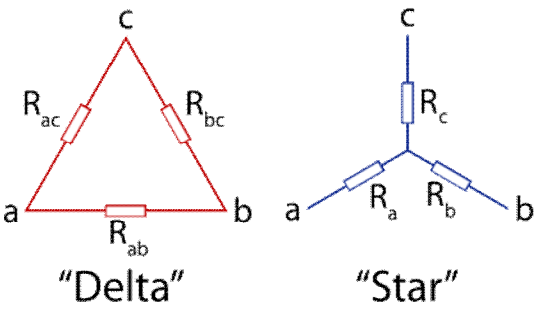
Any two terminal network of impedances can eventually be reduced to a single impedance by successive applications of impedances in series or impedances in parallel.

Impedances in series: $$Z_{\text{eq}} = Z_1 + Z_2 + \cdots + Z_n.$$

Impedances in parallel: $$\frac{1}{Z_{\text{eq}}} = \frac{1}{Z_1} + \frac{1}{Z_2} + \cdots + \frac{1}{Z_n}.$$

The above simplified for only two impedances in parallel: $$Z_{\text{eq}} = \frac{Z_1 Z_2}{Z_1 + Z_2}.$$

Delta-Star transformation



A network of impedances with more than two terminals cannot be reduced to a single impedance equivalent circuit. An n-terminal network can, at best, be reduced to n impedances. For a three terminal network, the three impedances can be expressed as a three node delta (Δ) network or a four node star (Y) network. These two networks are equivalent and the transformations between them are given below. A general network with an arbitrary number of terminals cannot be reduced to the minimum number of impedances using only series and parallel combinations. In general, Y-Δ and Δ-Y transformations must also be used. It can be shown that this is sufficient to find the minimal network for any arbitrary network with

successive applications of series, parallel, Y-Δ and Δ-Y; no more complex transformations are required.

For equivalence, the impedances between any pair of terminals must be the same for both networks, resulting in a set of three simultaneous equations. The equations below are expressed as resistances but apply equally to the general case with impedances.

Delta-to-star transformation equations

$$R_a = \frac{R_{\text{ac}} R_{\text{ab}}}{R_{\text{ac}} + R_{\text{ab}} + R_{\text{bc}}}$$

$$R_b = \frac{R_{\text{ab}} R_{\text{bc}}}{R_{\text{ac}} + R_{\text{ab}} + R_{\text{bc}}}$$

$$R_c = \frac{R_{\text{bc}} R_{\text{ac}}}{R_{\text{ac}} + R_{\text{ab}} + R_{\text{bc}}}$$

Star-to-delta transformation equations

$$R_{\text{ac}} = \frac{R_a R_b + R_b R_c + R_c R_a}{R_b}$$

$$R_{\text{ab}} = \frac{R_a R_b + R_b R_c + R_c R_a}{R_c}$$

$$R_{\text{bc}} = \frac{R_a R_b + R_b R_c + R_c R_a}{R_a}$$

General form of network node elimination

The star-to-delta and series-resistor transformations are special cases of the general resistor network node elimination algorithm. Any node connected by *N* resistors (R_1 ... R_N) to nodes 1 .. N can be replaced by $\binom{N}{2}$ resistors interconnecting the remaining *N* nodes. The resistance between any two nodes *x* and *y* is given by:

$$R_{\text{xy}} = R_x R_y \sum_{i=1}^{N} \frac{1}{R_i}$$

For a star-to-delta ($N = 3$) this reduces to:

$$R_{ab} = R_a R_b \left(\frac{1}{R_a} + \frac{1}{R_b} + \frac{1}{R_c}\right) = \frac{R_a R_b (R_a R_b + R_a R_c + R_b R_c)}{R_a R_b R_c} =$$

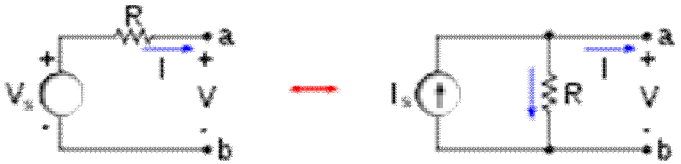
$$= \frac{R_a R_b + R_b R_c + R_c R_a}{R_c}$$

For a series reduction ($N = 2$) this reduces to:

$$R_{ab} = R_a R_b \left(\frac{1}{R_a} + \frac{1}{R_b}\right) = \frac{R_a R_b (R_a + R_b)}{R_a R_b} = R_a + R_b$$

For a dangling resistor ($N = 1$) it results in the elimination of the resistor because $R_{ab} = R_a R_b \left(\frac{1}{R_a} + \frac{1}{R_b}\right) = \frac{R_a R_b (R_a + R_b)}{R_a R_b} = R_a + R_b$.

Source transformation



A generator with an internal impedance (i.e. non-ideal generator) can be represented as either, an ideal voltage generator or an ideal current generator plus the impedance. These two forms are equivalent and the transformations are given below. If the two networks are equivalent with respect to terminals a-b, then V and I must be identical for both networks. Thus,

$$V_s = RI_s \quad \text{or} \quad I_s = \frac{V_s}{R};$$

- Norton's theorem states that any two-terminal network can be reduced to an ideal current generator and a parallel impedance.
- Thévenin's theorem states that any two-terminal network can be reduced to an ideal voltage generator plus a series impedance.

Simple networks

Some very simple networks can be analyzed without the need to apply the more systematic approaches.

Voltage division of series components

Consider n impedances that are connected in **series**. The voltage V_i across any impedance Z_i is

$$V_i = Z_i I = \left(\frac{Z_i}{Z_1 + Z_2 + \cdots + Z_n} \right) V$$

Current division of parallel components

Consider n impedances that are connected in **parallel**. The current I_i through any impedance Z_i is

$$I_i = \left(\frac{\left(\frac{1}{Z_i}\right)}{\left(\frac{1}{Z_1}\right) + \left(\frac{1}{Z_2}\right) + \cdots + \left(\frac{1}{Z_n}\right)} \right) I \qquad \text{for } i = 1,2,\ldots,n.$$

Special case: Current division of two parallel components

$$I_1 = \left(\frac{Z_2}{Z_1 + Z_2} \right) I; \quad I_2 = \left(\frac{Z_1}{Z_1 + Z_2} \right) I$$

Nodal analysis

1. Label all **nodes** in the circuit. Arbitrarily select any node as reference.

2. Define a voltage variable from every remaining node to the reference. These voltage variables must be defined as voltage rises with respect to the reference node.

3. Write a KCL equation for every node except the reference.

4. Solve the resulting system of equations.

Mesh analysis

Mesh — a loop that does not contain an inner loop.

1. Count the number of "window panes" in the circuit. Assign a mesh

current to each window pane.

2. Write a KVL equation for every mesh whose current is unknown.

3. Solve the resulting equations

Superposition

In this method, the effect of each generator in turn is calculated. All the generators other than the one being considered are removed; either short-circuited in the case of voltage generators or open circuited in the case of current generators. The total current through or the total voltage across, a particular branch is then calculated by summing all the individual currents or voltages.

There is an underlying assumption to this method that the total current or voltage is a linear superposition of its parts. The method cannot, therefore, be used if non-linear components are present. Note that mesh analysis and node analysis also implicitly use superposition so these too, are only applicable to linear circuits.

Choice of method

Choice of method is to some extent a matter of taste. If the network is particularly simple or only a specific current or voltage is required then ad-hoc application of some simple equivalent circuits may yield the answer without recourse to the more systematic methods.

- Superposition is possibly the most conceptually simple method but rapidly leads to a large number of equations and messy impedance combinations as the network becomes larger.
- Nodal analysis: The number of voltage variables, and hence simultaneous equations to solve, equals the number of nodes minus one. Every voltage source connected to the reference node reduces the number of unknowns (and equations) by one. Nodal analysis is thus best for voltage sources.
- Mesh analysis: The number of current variables, and hence simultaneous equations to solve, equals the number of meshes. Every current source in a mesh reduces the number of unknowns by one. Mesh analysis is thus best for current sources. Mesh analysis, however, cannot be used with networks which cannot be drawn as a planar network, that is, with no crossing components.

Transfer function

A transfer function expresses the relationship between an input and an output of a network. For resistive networks, this will always be a simple real number or an expression which boils down to a real number.

Resistive networks are represented by a system of simultaneous algebraic equations. However in the general case of linear networks, the network is represented by a system of simultaneous linear differential equations. In network analysis, rather than use the differential equations directly, it is usual practice to carry out a Laplace transform on them first and then express the result in terms of the Laplace parameter s, which in general is complex. This is described as working in the s-domain. Working with the equations directly would be described as working in the time (or t) domain because the results would be expressed as time varying quantities.

The Laplace transform is the mathematical method of transforming between the s-domain and the t-domain.

This approach is standard in control theory and is useful for determining stability of a system, for instance, in an amplifier with feedback.

Two terminal component transfer functions

For two terminal components the transfer function is the relationship between the current input to the device and the resulting voltage across it. The transfer function, Z(s), will thus have units of impedance - ohms. For the three passive components found in electrical networks, the transfer functions are;

Resistor $Z(s) = R$

Inductor $Z(s) = sL$

Capacitor $Z(s) = \frac{1}{sC}$

For a network to which only steady ac signals are applied, s is replaced with $j\omega$ and the more familiar values from ac network theory result.

Resistor $Z(j\omega) = R$

Inductor $Z(j\omega) = j\omega L$

Capacitor $Z(j\omega) = \frac{1}{j\omega C}$

Finally, for a network to which only steady dc is applied, s is replaced with zero and dc network theory applies.

Resistor $Z = R$

Inductor $Z = 0$

Capacitor $Z = \infty$

Two port network transfer function

Transfer functions, in general, in control theory are given the symbol H(s). Most commonly in electronics, transfer function is defined as the ratio of output voltage to input voltage and given the symbol A(s), or more commonly (because analysis is invariably done in terms of sine wave response), A(jω), so that;

$$A(j\omega) = \frac{V_o}{V_i}$$

The A standing for attenuation, or amplification, depending on context. In general, this will be a complex function of *jω*, which can be derived from an analysis of the impedances in the network and their individual transfer functions. Sometimes the analyst is only interested in the magnitude of the gain and not the phase angle. In this case the complex numbers can be eliminated from the transfer function and it might then be written as:

$$A(\omega) = \left|\frac{V_o}{V_i}\right|$$

Two port parameters

The concept of a two-port network can be useful in network analysis as a black box approach to analysis. The behavior of the two-port network in a larger network can be entirely characterized without necessarily stating anything about the internal structure. However, to do this it is necessary to have more information than just the A(jω) described above. It can be shown that four such parameters are required to fully characterize the two-port network. These could be the forward transfer function, the input impedance, the reverse transfer function (ie, the voltage appearing at the

input when a voltage is applied to the output) and the output impedance. There are many others (see the main article for a full listing), one of these expresses all four parameters as impedances. It is usual to express the four parameters as a matrix;

$$\begin{bmatrix} V_1 \\ V_0 \end{bmatrix} = \begin{bmatrix} z(j\omega)_{11} & z(j\omega)_{12} \\ z(j\omega)_{21} & z(j\omega)_{22} \end{bmatrix} \begin{bmatrix} I_1 \\ I_0 \end{bmatrix}$$

The matrix may be abbreviated to a representative element;
$[z(j\omega)]$ or just $[z]$

These concepts are capable of being extended to networks of more than two ports. However, this is rarely done in reality as in many practical cases ports are considered either purely input or purely output. If reverse direction transfer functions are ignored, a multi-port network can always be decomposed into a number of two-port networks.

Distributed components

Where a network is composed of discrete components, analysis using two-port networks is a matter of choice, not essential. The network can always alternatively be analyzed in terms of its individual component transfer functions. However, if a network contains distributed components, such as in the case of a transmission line, then it is not possible to analyze in terms of individual components since they do not exist. The most common approach to this is to model the line as a two-port network and characterize it using two-port parameters (or something equivalent to them). Another example of this technique is modeling the carriers crossing the base region in a high frequency transistor. The base region has to be modeled as distributed resistance and capacitance rather than lumped components.

Image analysis

Transmission lines and certain types of filter design use the image method to determine their transfer parameters. In this method, the behavior of an infinitely long cascade connected chain of identical networks is considered. The input and output impedances and the forward and reverse transmission functions are then calculated for this infinitely long chain. Although, the theoretical values so obtained can never be exactly realized in practice, in many cases they serve as a very good approximation for the behavior of a finite chain as long as it is not too short.

Non-linear networks

Most electronic designs are, in reality, non-linear. There is very little that does not include some semiconductor devices. These are invariably non-linear, the transfer function of an ideal semiconductor pn junction is given by the very non-linear relationship;

$$i = I_o(e^{\frac{v}{V_T}} - 1)$$

where;

- *i* and *v* are the instantaneous current and voltage.
- I_o is an arbitrary parameter called the reverse leakage current whose value depends on the construction of the device.
- V_T is a parameter proportional to temperature called the thermal voltage and equal to about 25mV at room temperature.

There are many other ways that non-linearity can appear in a network. All methods utilizing linear superposition will fail when non-linear components are present. There are several options for dealing with non-linearity depending on the type of circuit and the information the analyst wishes to obtain.

Boolean analysis of switching networks

A switching device is one where the non-linearity is utilized to produce two opposite states. CMOS devices in digital circuits, for instance, have their output connected to either the positive or the negative supply rail and are never found at anything in between except during a transient period when the device is actually switching. Here the non-linearity is designed to be extreme, and the analyst can actually take advantage of that fact. These kinds of networks can be analyzed using Boolean algebra by assigning the two states ("on"/"off", "positive"/"negative" or whatever states are being used) to the Boolean constants "0" and "1".

The transients are ignored in this analysis, along with any slight discrepancy between the actual state of the device and the nominal state assigned to a Boolean value. For instance, Boolean "1" may be assigned to the state of +5V. The output of the device may actually be +4.5V but the analyst still considers this to be Boolean "1". Device manufacturers will usually specify a range of values in their data sheets that are to be considered undefined (ie the result will be unpredictable).

The transients are not entirely uninteresting to the analyst. The maximum

rate of switching is determined by the speed of transition from one state to the other. Happily for the analyst, for many devices most of the transition occurs in the linear portion of the devices transfer function and linear analysis can be applied to obtain at least an approximate answer.

It is mathematically possible to derive Boolean algebras which have more than two states. There is not too much use found for these in electronics, although three-state devices are passingly common.

Separation of bias and signal analyses

This technique is used where the operation of the circuit is to be essentially linear, but the devices used to implement it are non-linear. A transistor amplifier is an example of this kind of network. The essence of this technique is to separate the analysis in to two parts. Firstly, the dc biases are analyzed using some non-linear method. This establishes the quiescent operating point of the circuit. Secondly, the small signal characteristics of the circuit are analyzed using linear network analysis. Examples of methods that can be used for both these stages are given below.

Graphical method of dc analysis

In a great many circuit designs, the dc bias is fed to a non-linear component via a resistor (or possibly a network of resistors). Since resistors are linear components, it is particularly easy to determine the quiescent operating point of the non-linear device from a graph of its transfer function. The method is as follows: from linear network analysis the output transfer function (that is output voltage against output current) is calculated for the network of resistor(s) and the generator driving them. This will be a straight line and can readily be superimposed on the transfer function plot of the non-linear device. The point where the lines cross is the quiescent operating point.

Perhaps the easiest practical method is to calculate the (linear) network open circuit voltage and short circuit current and plot these on the transfer function of the non-linear device. The straight line joining these two point is the transfer function of the network.

In reality, the designer of the circuit would proceed in the reverse direction to that described. Starting from a plot provided in the manufacturers data sheet for the non-linear device, the designer would choose the desired operating point and then calculate the linear component values required to achieve it.

It is still possible to use this method if the device being biased has its bias fed through another device which is itself non-linear - a diode for instance. In this case however, the plot of the network transfer function onto the device being biased would no longer be a straight line and is consequently more tedious to do.

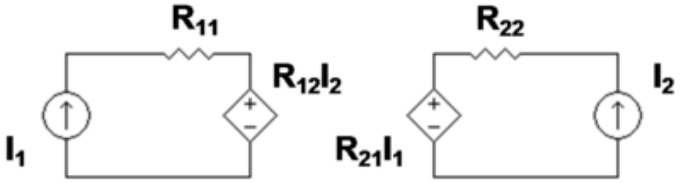
Small signal equivalent circuit

This method can be used where the deviation of the input and output signals in a network stay within a substantially linear portion of the non-linear devices transfer function, or else are so small that the curve of the transfer function can be considered linear. Under a set of these specific conditions, the non-linear device can be represented by an equivalent linear network. It must be remembered that this equivalent circuit is entirely notional and only valid for the small signal deviations. It is entirely inapplicable to the dc biasing of the device.

For a simple two-terminal device, the small signal equivalent circuit may be no more than two components. A resistance equal to the slope of the v/i curve at the operating point (called the dynamic resistance), and tangent to the curve. A generator, because this tangent will not, in general, pass through the origin. With more terminals, more complicated equivalent circuits are required.

A popular form of specifying the small signal equivalent circuit amongst transistor manufacturers is to use the two-port network parameters known as [h] parameters. These are a matrix of four parameters as with the [z] parameters but in the case of the [h] parameters they are a hybrid mixture of impedances, admittances, current gains and voltage gains. In this model the three-terminal transistor is considered to be a two-port network, one of its terminals being common to both ports. The [h] parameters are quite different depending on which terminal is chosen as the common one. The most important parameter for transistors is usually the forward current gain, h_{21}, in the common emitter configuration. This is designated h_{fe} on data sheets.

The small signal equivalent circuit in terms of two-port parameters leads to the concept of dependent generators. That is, the value of a voltage or current generator depends linearly on a voltage or current elsewhere in the circuit. For instance the [z] parameter model leads to dependent voltage generators as shown in this diagram;



[z] parameter equivalent circuit showing dependent voltage generators

There will always be dependent generators in a two-port parameter equivalent circuit. This applies to the [h] parameters as well as to the [z] and any other kind. These dependencies must be preserved when developing the equations in a larger linear network analysis.

Piecewise linear method

In this method, the transfer function of the non-linear device is broken up into regions. Each of these regions is approximated by a straight line. Thus, the transfer function will be linear up to a particular point where there will be a discontinuity. Past this point the transfer function will again be linear but with a different slope.

A well known application of this method is the approximation of the transfer function of a pn junction diode. The actual transfer function of an ideal diode has been given at the top of this (non-linear) section. However, this formula is rarely used in network analysis, a piecewise approximation being used instead. It can be seen that the diode current rapidly diminishes to $-I_o$ as the voltage falls. This current, for most purposes, is so small it can be ignored. With increasing voltage, the current increases exponentially. The diode is modeled as an open circuit up to the knee of the exponential curve, then past this point as a resistor equal to the bulk resistance of the semiconducting material.

The commonly accepted values for the transition point voltage are 0.7V for silicon devices and 0.3V for germanium devices. An even simpler model of the diode, sometimes used in switching applications, is short circuit for forward voltages and open circuit for reverse voltages.

The model of a forward biased pn junction having an approximately constant 0.7V is also a much used approximation for transistor base-emitter junction voltage in amplifier design.

The piecewise method is similar to the small signal method in that linear network analysis techniques can only be applied if the signal stays within

certain bounds. If the signal crosses a discontinuity point then the model is no longer valid for linear analysis purposes. The model does have the advantage over small signal however, in that it is equally applicable to signal and dc bias. These can therefore both be analyzed in the same operations and will be linearly super imposable.

Time-varying components

In linear analysis, the components of the network are assumed to be unchanging, but in some circuits this does not apply, such as sweep oscillators, voltage controlled amplifiers, and variable equalizers. In many circumstances the change in component value is periodic. A non-linear component excited with a periodic signal, for instance, can be represented as periodically varying *linear* component. Sidney Darlington disclosed a method of analyzing such periodic time varying circuits. He developed canonical circuit forms which are analogous to the canonical forms of Ronald Foster and Wilhelm Cauer used for analyzing linear circuits.

References

1. Belevitch V (May 1962). "Summary of the history of circuit theory". *Proceedings of the IRE* **50** (5): 849. doi:10.1109/JRPROC.1962.288301. cites "IRE Standards on Circuits: Definitions of Terms for Linear Passive Reciprocal Time Invariant Networks, 1960". *Proceedings of the IRE* **48** (9): 1609. September 1960. doi:10.1109/JRPROC.1960.287676. to justify this definition.
   Sidney Darlington Darlington S (1984). "A history of network synthesis and filter theory for circuits composed of resistors, inductors, and capacitors". *IEEE Trans. Circuits and Systems* **31** (1): 4.follows Belevitch but notes there are now also many colloquial uses of "network".
2. Nilsson, J W, Riedel, S A (2007). *Electric Circuits* (8th ed). Pearson Prentice Hall. p. 112–113. ISBN 0-13-198925-1. http://books.google.com/books?id=sxmM8RFL99wC&lpg=PA200&dq=isbn%3A0131989251&lr=&as_drrb_is=q&as_minm_is=0&as_miny_is=&as_maxm_is=0&as_maxy_is=&as_brr=0&pg=PA112#v=onepage&q=112&f=false.
3. Nilsson, J W, Riedel, S A (2007). *Electric Circuits* (8th ed.). Pearson Prentice Hall. p. 94. ISBN 0-13-198925-1. http://books.google.com/books?id=sxmM8RFL99wC&lpg=PA200&dq=isbn%3A0131989251&lr=&as_drrb_is=q&as_minm_is=0&as_miny_is=&as_maxm_is=0&as_maxy_is=&as_brr=

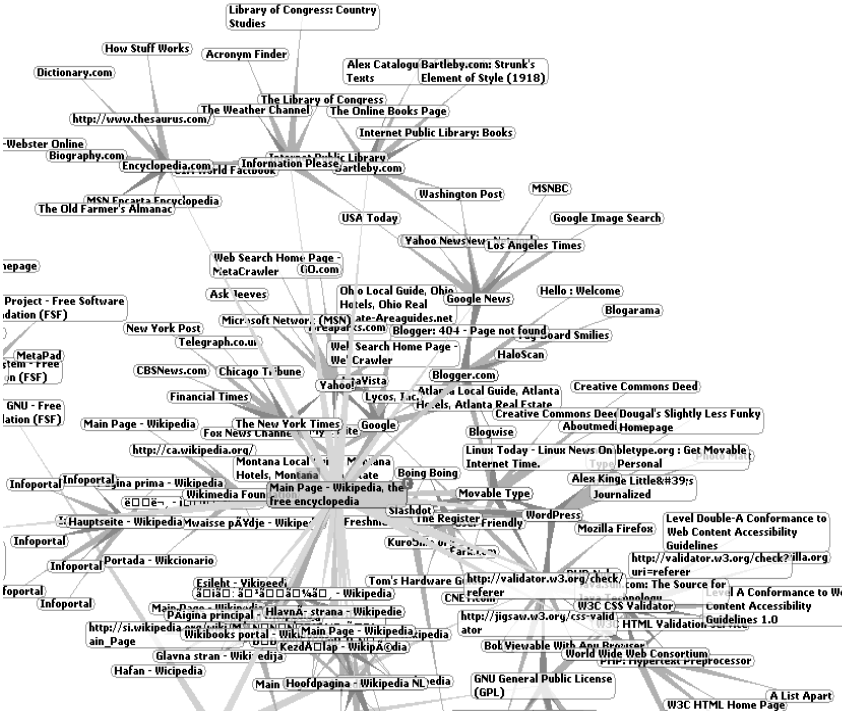
0&pg=PA94#v=onepage&q=&f=false.
4. US patent 3265973, Sidney Darlington, Irwin W. Sandberg, "*Synthesis of two-port networks having periodically time-varying elements*", granted 1966-08-09

Other places read

- Circuit Analysis Techniques
- Nodal Analysis of Op Amp Circuits
- Analysis of Resistive Circuits
- Circuit Analysis Related Laws
- Bartlett's bisection theorem
- Circuit theory
- Equivalent impedance transforms
- Kirchhoff's circuit laws
- Mesh analysis
- Millman's Theorem
- Ohm's law
- Reciprocity theorem
- Resistive circuit
- Series and parallel circuits
- Tellegen's theorem
- Two-port network
- Wye-delta transform

10. Social network

A **social network** is a social structure made of individuals (or organizations) called "nodes," which are tied (connected) by one or more specific types of interdependency, such as friendship, kinship, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge or prestige.



An example of a social network diagram.

Social network analysis views social relationships in terms of network theory consisting of *nodes* and *ties*. Nodes are the individual actors within the networks, and ties are the relationships between the actors. The resulting graph-based structures are often very complex. There can be many kinds of ties between the nodes. Research in a number of academic

fields has shown that social networks operate on many levels, from families up to the level of nations, and play a critical role in determining the way problems are solved, organizations are run, and the degree to which individuals succeed in achieving their goals.

In its simplest form, a social network is a map of all of the relevant ties between all the nodes being studied. The network can also be used to measure social capital -- the value that an individual gets from the social network. These concepts are often displayed in a social network diagram, where nodes are the points and ties are the lines.

Social network analysis

Social network analysis (related to *network theory*) has emerged as a key technique in modern sociology. It has also gained a significant following in anthropology, biology, communication studies, economics, geography, information science, organizational studies, social psychology, and sociolinguistics, and has become a popular topic of speculation and study.

People have used the idea of **"social network"** loosely for over a century to connote complex sets of relationships between members of social systems at all scales, from interpersonal to international. In 1954, J. A. Barnes started using the term systematically to denote patterns of ties, encompassing concepts traditionally used by the public and those used by social scientists: bounded groups (e.g., tribes, families) and social categories (e.g., gender, ethnicity). Scholars such as S.D. Berkowitz, Stephen Borgatti, Ronald Burt, Kathleen Carley, Martin Everett, Katherine Faust, Linton Freeman, Mark Granovetter, David Knoke, David Krackhardt, Peter Marsden, Nicholas Mullins, Anatol Rapoport, Stanley Wasserman, Barry Wellman, Douglas R. White, and Harrison White expanded the use of systematic social network analysis.

Social network analysis has now moved from being a suggestive metaphor to an analytic approach to a paradigm, with its own theoretical statements, methods, social network analysis software, and researchers. Analysts reason from whole to part; from structure to relation to individual; from behavior to attitude. They typically either study *whole networks* (also known as *complete networks*), all of the ties containing specified relations in a defined population, or *personal networks* (also known as *egocentric networks*), the ties that specified people have, such as their "personal communities". The distinction between whole/complete networks and personal/egocentric networks has depended largely on how analysts were

able to gather data. That is, for groups such as companies, schools, or membership societies, the analyst was expected to have complete information about who was in the network, all participants being both potential egos and alters. Personal/egocentric studies were typically conducted when identities of egos were known, but not their alters. These studies rely on the egos to provide information about the identities of alters and there is no expectation that the various egos or sets of alters will be tied to each other. A *snowball network* refers to the idea that the alters identified in an egocentric survey then become egos themselves and are able in turn to nominate additional alters. While there are severe logistic limits to conducting snowball network studies, a method for examining *hybrid networks* has recently been developed in which egos in complete networks can nominate alters otherwise not listed who are then available for all subsequent egos to see. [3] The hybrid network may be valuable for examining whole/complete networks that are expected to include important players beyond those who are formally identified. For example, employees of a company often work with non-company consultants who may be part of a network that cannot fully be defined prior to data collection.

Several analytic tendencies distinguish social network analysis:

There is no assumption that groups are the building blocks of society: the approach is open to studying less-bounded social systems, from nonlocal communities to links among websites.

Rather than treating individuals (persons, organizations, states) as discrete units of analysis, it focuses on how the structure of ties affects individuals and their relationships.

In contrast to analyses that assume that socialization into norms determines behavior, network analysis looks to see the extent to which the structure and composition of ties affect norms.

The shape of a social network helps determine a network's usefulness to its individuals. Smaller, tighter networks can be less useful to their members than networks with lots of loose connections (weak ties) to individuals outside the main network. More open networks, with many weak ties and social connections, are more likely to introduce new ideas and opportunities to their members than closed networks with many redundant ties. In other words, a group of friends who only do things with each other already share the same knowledge and opportunities. A group of individuals with connections to other social worlds is likely to have access to a wider range

of information. It is better for individual success to have connections to a variety of networks rather than many connections within a single network. Similarly, individuals can exercise influence or act as brokers within their social networks by bridging two networks that are not directly linked (called filling structural holes).

The power of social network analysis stems from its difference from traditional social scientific studies, which assume that it is the attributes of individual actors—whether they are friendly or unfriendly, smart or dumb, etc.—that matter. Social network analysis produces an alternate view, where the attributes of individuals are less important than their relationships and ties with other actors within the network. This approach has turned out to be useful for explaining many real-world phenomena, but leaves less room for individual agency, the ability for individuals to influence their success, because so much of it rests within the structure of their network.

Social networks have also been used to examine how organizations interact with each other, characterizing the many informal connections that link executives together, as well as associations and connections between individual employees at different organizations. For example, power within organizations often comes more from the degree to which an individual within a network is at the center of many relationships than actual job title. Social networks also play a key role in hiring, in business success, and in job performance. Networks provide ways for companies to gather information, deter competition, and collude in setting prices or policies.

History of social network analysis

A summary of the progress of social networks and social network analysis has been written by Linton Freeman.

Precursors of social networks in the late 1800s include Émile Durkheim and Ferdinand Tönnies. Tönnies argued that social groups can exist as personal and direct social ties that either link individuals who share values and belief (*gemeinschaft*) or impersonal, formal, and instrumental social links (*gesellschaft*). Durkheim gave a non-individualistic explanation of social facts arguing that social phenomena arise when interacting individuals constitute a reality that can no longer be accounted for in terms of the properties of individual actors. He distinguished between a traditional society – "mechanical solidarity" – which prevails if individual differences are minimized, and the modern society – "organic solidarity" – that

develops out of cooperation between differentiated individuals with independent roles.

Georg Simmel, writing at the turn of the twentieth century, was the first scholar to think directly in social network terms. His essays pointed to the nature of network size on interaction and to the likelihood of interaction in ramified, loosely-knit networks rather than groups (Simmel 1908/1971).

After a hiatus in the first decades of the twentieth century, three main traditions in social networks appeared. In the 1930s, J.L. Moreno pioneered the systematic recording and analysis of social interaction in small groups, especially classrooms and work groups (sociometry), while a Harvard group led by W. Lloyd Warner and Elton Mayo explored interpersonal relations at work. In 1940, A.R. Radcliffe-Brown's presidential address to British anthropologists urged the systematic study of networks. However, it took about 15 years before this call was followed-up systematically.

Social network analysis developed with the kinship studies of Elizabeth Bott in England in the 1950s and the 1950s-1960s urbanization studies of the University of Manchester group of anthropologists (centered around Max Gluckman and later J. Clyde Mitchell) investigating community networks in southern Africa, India and the United Kingdom. Concomitantly, British anthropologist S.F. Nadel codified a theory of social structure that was influential in later network analysis.

In the 1960s-1970s, a growing number of scholars worked to combine the different tracks and traditions. One large group was centered around Harrison White and his students at Harvard University: Ivan Chase, Bonnie Erickson, Harriet Friedmann, Mark Granovetter, Nancy Howell, Joel Levine, Nicholas Mullins, John Padgett, Michael Schwartz and Barry Wellman. Also important in this early group were Charles Tilly, who focused on networks in political sociology and social movements, and Stanley Milgram, who developed the "six degrees of separation" thesis.[10] White's group thought of themselves as rebelling against the reigning structural-functionalist orthodoxy of then-dominant Harvard sociologist Talcott Parsons, leading them to devalue concerns with symbols, values, norms and culture. They also were opposed to the methodological individualism espoused by another Harvard sociologist, George Homans, which was endemic among the dominant survey researchers and positivists of the time. Mark Granovetter and Barry Wellman are among the former students of White who have elaborated and popularized social network analysis.

White's was not the only group. Significant independent work was done by

scholars elsewhere: University of California Irvine social scientists interested in mathematical applications, centered around Linton Freeman, including John Boyd, Susan Freeman, Kathryn Faust, A. Kimball Romney and Douglas White; quantitative analysts at the University of Chicago, including Joseph Galaskiewicz, Wendy Griswold, Edward Laumann, Peter Marsden, Martina Morris, and John Padgett; and communication scholars at Michigan State University, including Nan Lin and Everett Rogers. A substantively-oriented University of Toronto sociology group developed in the 1970s, centered on former students of Harrison White: S.D. Berkowitz, Harriet Friedmann, Nancy Leslie Howard, Nancy Howell, Lorne Tepperman and Barry Wellman, and also including noted modeler and game theorist Anatol Rapoport.

Research

Social network analysis has been used in epidemiology to help understand how patterns of human contact aid or inhibit the spread of diseases such as HIV in a population. The evolution of social networks can sometimes be modeled by the use of agent based models, providing insight into the interplay between communication rules, rumor spreading and social structure.

SNA may also be an effective tool for mass surveillance -- for example the Total Information Awareness program was doing in-depth research on strategies to analyze social networks to determine whether or not U.S. citizens were political threats.

Diffusion of innovations theory explores social networks and their role in influencing the spread of new ideas and practices. Change agents and opinion leaders often play major roles in spurring the adoption of innovations, although factors inherent to the innovations also play a role.

Robin Dunbar has suggested that the typical size of a egocentric network is constrained to about 150 members due to possible limits in the capacity of the human communication channel. The rule arises from cross-cultural studies in sociology and especially anthropology of the maximum size of a village (in modern parlance most reasonably understood as an *ecovillage*). It is theorized in evolutionary psychology that the number may be some kind of limit of average human ability to recognize members and track emotional facts about all members of a group. However, it may be due to economics and the need to track "free riders", as it may be easier in larger groups to take advantage of the benefits of living in a community without contributing to those benefits.

Mark Granovetter found in one study that more numerous weak ties can be important in seeking information and innovation. Cliques have a tendency to have more homogeneous opinions as well as share many common traits. This homophilic tendency was the reason for the members of the cliques to be attracted together in the first place. However, being similar, each member of the clique would also know more or less what the other members knew. To find new information or insights, members of the clique will have to look beyond the clique to its other friends and acquaintances. This is what Granovetter called the "the strength of weak ties".

Guanxi is a central concept in Chinese society (and other East Asian cultures) that can be summarized as the use of personal influence. Guanxi can be studied from a social network approach.

The small world phenomenon is the hypothesis that the chain of social acquaintances required to connect one arbitrary person to another arbitrary person anywhere in the world is generally short. The concept gave rise to the famous phrase six degrees of separation after a 1967 *small world experiment* by psychologist Stanley Milgram. In Milgram's experiment, samples of US individuals were asked to reach a particular target person by passing a message along a chain of acquaintances. The average length of successful chains turned out to be about five intermediaries or six separation steps (the majority of chains in that study actually failed to complete). The methods (and ethics as well) of Milgram's experiment was later questioned by an American scholar, and some further research to replicate Milgram's findings had found that the degrees of connection needed could be higher. Academic researchers continue to explore this phenomenon as Internet-based communication technology has supplemented the phone and postal systems available during the times of Milgram. A recent electronic small world experiment at Columbia University found that about five to seven degrees of separation are sufficient for connecting any two people through e-mail.

Collaboration graphs can be used to illustrate good and bad relationships between humans. A positive edge between two nodes denotes a positive relationship (friendship, alliance, dating) and a negative edge between two nodes denotes a negative relationship (hatred, anger). Signed social network graphs can be used to predict the future evolution of the graph. In signed social networks, there is the concept of "balanced" and "unbalanced" cycles. A balanced cycle is defined as a cycle where the products of all the signs are positive. Balanced graphs represent a group of

people who are unlikely to change their opinions of the other people in the group. Unbalanced graphs represent a group of people who are very likely to change their opinions of the people in their group. For example, a group of 3 people (A, B, and C) where A and B have a positive relationship, B and C have a positive relationship, but C and A have a negative relationship is an unbalanced cycle. This group is very likely to morph into a balanced cycle, such as one where B only has a good relationship with A, and both A and B have a negative relationship with C. By using the concept of balances and unbalanced cycles, the evolution of signed social network graphs can be predicted.

One study has found that happiness tends to be correlated in social networks. When a person is happy, nearby friends have a 25 percent higher chance of being happy themselves. Furthermore, people at the center of a social network tend to become happier in the future than those at the periphery. Clusters of happy and unhappy people were discerned within the studied networks, with a reach of three degrees of separation: a person's happiness was associated with the level of happiness of their friends' friends' friends.

Some researchers have suggested that human social networks may have a genetic basis. Using a sample of twins from the National Longitudinal Study of Adolescent Health, they found that in-degree (the number of times a person is named as a friend), transitivity (the probability that two friends are friends with one another), and betweenness centrality (the number of paths in the network that pass through a given person) are all significantly heritable. Existing models of network formation cannot account for this intrinsic node variation, so the researchers propose an alternative "Attract and Introduce" model that can explain heritability and many other features of human social networks.

Application to Environmental Issues

The 1984 book The IRG Solution argued that central media and government-type hierarchical organizations could not adequately understand the environmental crisis we were manufacturing, or how to initiate adequate solutions. It argued that the widespread introduction of Information Routing Groups was required to create a social network whose overall intelligence could collectively understand the issues and devise and implement correct workable solutions and policies.

Measures in social network analysis

Betweenness

The extent to which a node lies between other nodes in the network. This measure takes into account the connectivity of the node's neighbors, giving a higher value for nodes which bridge clusters. The measure reflects the number of people who a person is connecting indirectly through their direct links.

Bridge

An edge is said to be a bridge if deleting it would cause its endpoints to lie in different components of a graph.

Centrality

This measure gives a rough indication of the social power of a node based on how well they "connect" the network. "Betweenness", "Closeness", and "Degree" are all measures of centrality.

Centralization

The difference between the number of links for each node divided by maximum possible sum of differences. A centralized network will have many of its links dispersed around one or a few nodes, while a decentralized network is one in which there is little variation between the number of links each node possesses.

Closeness

The degree an individual is near all other individuals in a network (directly or indirectly). It reflects the ability to access information through the "grapevine" of network members. Thus, closeness is the inverse of the sum of the shortest distances between each individual and every other person in the network. The shortest path may also be known as the "geodesic distance".

Clustering coefficient

A measure of the likelihood that two associates of a node are associates them. A higher clustering coefficient indicates a greater 'cliquishness'.

Cohesion

The degree to which actors are connected directly to each other by cohesive bonds. Groups are identified as ‘cliques’ if every individual is directly tied to every other individual, ‘social circles’ if there is less stringency of direct contact, which is imprecise, or as structurally cohesive blocks if precision is wanted.

Degree

The count of the number of ties to other actors in the network. See also degree (graph theory).

(Individual-level) Density
The degree a respondent's ties know one another/ proportion of ties among an individual's nominees. Network or global-level density is the proportion of ties in a network relative to the total number possible (sparse versus dense networks).

Flow betweenness centrality
The degree that a node contributes to sum of maximum flow between all pairs of nodes (not that node).

Eigenvector centrality
A measure of the importance of a node in a network. It assigns relative scores to all nodes in the network based on the principle that connections to nodes having a high score contribute more to the score of the node in question.

Local Bridge
An edge is a local bridge if its endpoints share no common neighbors. Unlike a bridge, a local bridge is contained in a cycle.

Path Length
The distances between pairs of nodes in the network. Average path-length is the average of these distances between all pairs of nodes.

Prestige
In a directed graph prestige is the term used to describe a node's centrality. "Degree Prestige", "Proximity Prestige", and "Status Prestige" are all measures of Prestige.

Radiality
Degree an individual's network reaches out into the network and provides novel information and influence.

Reach
The degree any member of a network can reach other members of the network.

Structural cohesion
The minimum number of members who, if removed from a group, would disconnect the group.

Structural equivalence
Refers to the extent to which nodes have a common set of linkages to other nodes in the system. The nodes don't need to have any ties to each

other to be structurally equivalent.

Structural hole
Static holes that can be strategically filled by connecting one or more links to link together other points. Linked to ideas of social capital: if you link to two people who are not linked you can control their communication.

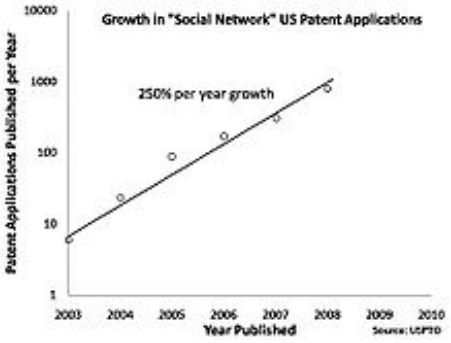
Network analytic software

Network analytic tools are used to represent the nodes (agents) and edges (relationships) in a network, and to analyze the network data. Like other software tools, the data can be saved in external files. Network analysis tools allow researchers to investigate large networks like the Internet, disease transmission, etc. These tools provide mathematical functions that can be applied to the network model.

Visual representation of social networks is important to understand the network data and convey the result of the analysis [2]. Network analysis tools are used to change the layout, colors, size and advanced properties of the network representation.

Patents

There has been rapid growth in the number of US patent applications that cover new technologies related to social networking. The number of published applications has been growing at about 250% per year over the past five years.



There are now over 2000 published applications. Only about 100 of these applications have issued as patents, however, largely due to the multi-year backlog in examination of business method patents and ethical issues connected with this patent category

References

1. Linton Freeman, *The Development of Social Network Analysis.* Vancouver: Empirical Press, 2006.
2. Wellman, Barry and S.D. Berkowitz, eds., 1988. *Social Structures: A*

Network Approach. Cambridge: Cambridge University Press.
3. Hansen, William B. and Reese, Eric L. 2009. Network Genie User Manual. Greensboro, NC: Tanglewood Research.
4. Freeman, Linton. 2006. *The Development of Social Network Analysis.* Vancouver: Empirical Pres, 2006; Wellman, Barry and S.D. Berkowitz, eds., 1988. *Social Structures: A Network Approach.* Cambridge: Cambridge University Press.
5. Scott, John. 1991. *Social Network Analysis.* London: Sage.
6. Wasserman, Stanley, and Faust, Katherine. 1994. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.
7. *The Development of Social Network Analysis* Vancouver: Empirical Press.
8. A.R. Radcliffe-Brown, "On Social Structure," *Journal of the Royal Anthropological Institute:* 70 (1940): 1-12.
9. [Nadel, SF. 1957. *The Theory of Social Structure.* London: Cohen and West.
10. The Networked Individual: A Profile of Barry Wellman.
11. Mark Granovetter, "Introduction for the French Reader," *Sociologica* 2 (2007): 1-8; Wellman, Barry. 1988. "Structural Analysis: From Method and Metaphor to Theory and Substance." Pp. 19-61 in *Social Structures: A Network Approach*, edited by Barry Wellman and S.D. Berkowitz. Cambridge: Cambridge University Press.
12. Mark Granovetter, "Introduction for the French Reader," *Sociologica* 2 (2007): 1-8; Wellman, Barry. 1988. "Structural Analysis: From Method and Metaphor to Theory and Substance." Pp. 19-61 in *Social Structures: A Network Approach*, edited by Barry Wellman and S.D. Berkowitz. Cambridge: Cambridge University Press. (See also Scott, 2000 and Freeman, 2004).
13. Barry Wellman, Wenhong Chen and Dong Weizhen. "Networking Guanxi." Pp. 221-41 in Social Connections in China: Institutions, Culture and the Changing Nature of Guanxi, edited by Thomas Gold, Douglas Guthrie and David Wank. Cambridge University Press, 2002.
14. *Could It Be A Big World After All?*: Judith Kleinfeld article.
15. Six Degrees: The Science of a Connected Age, Duncan Watts.
16. James H. Fowler and Nicholas A. Christakis. 2008. "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study." *British Medical Journal*. December 4, 2008: doi:10.1136/bmj.a2338. Media account for those who cannot retrieve the original: Happiness: It Really is Contagious Retrieved

December 5, 2008.
17. "Genes and the Friends You Make". *Wall Street Journal*. January 27, 2009. http://online.wsj.com/article/SB123302040874118079.html.
18. Fowler, J. H. (10 February 2009). "Model of Genetic Variation in Human Social Networks" (PDF). *Proceedings of the National Academy of Sciences* 106 (6): 1720–1724. doi:10.1073/pnas.0806746106. http://jhfowler.ucsd.edu/genes_and_social_networks.pdf.
19. The most comprehensive reference is: Wasserman, Stanley, & Faust, Katherine. (1994). Social Networks Analysis: Methods and Applications. Cambridge: Cambridge University Press. A short, clear basic summary is in Krebs, Valdis. (2000). "The Social Life of Routers." *Internet Protocol Journal*, 3 (December): 14-25.
20. Moody, James, and Douglas R. White (2003). "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." *American Sociological Review* 68(1):103-127. Online: (PDF file.
21. USPTO search on published patent applications such as "social network": http://en.wikipedia.org/wiki/Social_networks#cite_ref-22#cite_ref-22

Further reading

1. Barnes, J. A. "Class and Committees in a Norwegian Island Parish", *Human Relations* 7:39-58
2. Berkowitz, Stephen D. 1982. *An Introduction to Structural Analysis: The Network Approach to Social Research.* Toronto: Butterworth. ISBN 0409813621
3. Brandes, Ulrik, and Thomas Erlebach (Eds.). 2005. *Network Analysis: Methodological Foundations* Berlin, Heidelberg: Springer-Verlag.
4. Breiger, Ronald L. 2004. "The Analysis of Social Networks." Pp. 505–526 in *Handbook of Data Analysis,* edited by Melissa Hardy and Alan Bryman. London: Sage Publications. ISBN 0761966528 *Excerpts in pdf format*
5. Burt, Ronald S. (1992). *Structural Holes: The Structure of Competition.* Cambridge, MA: Harvard University Press. ISBN 067484372X
6. Casaleggio, Davide (2008). *TU SEI RETE. La Rivoluzione del business, del marketing e della politica attraverso le reti sociali.* ISBN 8890182652
7. Carrington, Peter J., John Scott and Stanley Wasserman (Eds.). 2005. *Models and Methods in Social Network Analysis.* New York: Cambridge University Press. ISBN 9780521809597
8. Christakis, Nicholas and James H. Fowler "The Spread of Obesity in a Large Social Network Over 32 Years," *New England Journal of Medicine*

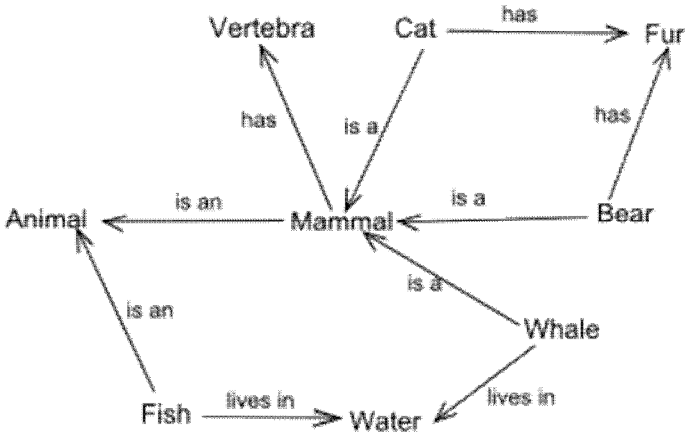
357 (4): 370-379 (26 July 2007)
9. Doreian, Patrick, Vladimir Batagelj, and Anuska Ferligoj. (2005). *Generalized Blockmodeling.* Cambridge: Cambridge University Press. ISBN 0521840856
10. Freeman, Linton C. (2004) *The Development of Social Network Analysis: A Study in the Sociology of Science.* Vancouver: Empirical Press. ISBN 1594577145
11. Hill, R. and Dunbar, R. 2002. "Social Network Size in Humans." Human Nature, Vol. 14, No. 1, pp. 53–72.
12. Jackson, Matthew O. (2003). "A Strategic Model of Social and Economic Networks". *Journal of Economic Theory* **71**: 44–74. doi:10.1006/jeth.1996.0108. pdf
13. Huisman, M. and Van Duijn, M. A. J. (2005). Software for Social Network Analysis. In P J. Carrington, J. Scott, & S. Wasserman (Editors), *Models and Methods in Social Network Analysis* (pp. 270–316). New York: Cambridge University Press. ISBN 9780521809597
14. Krebs, Valdis (2006) Social Network Analysis, A Brief Introduction. (Includes a list of recent SNA applications Web Reference.)
15. Lin, Nan, Ronald S. Burt and Karen Cook, eds. (2001). *Social Capital: Theory and Research.* New York: Aldine de Gruyter. ISBN 0202306437
16. Mullins, Nicholas. 1973. *Theories and Theory Groups in Contemporary American Sociology.* New York: Harper and Row. ISBN 0060446498
17. Müller-Prothmann, Tobias (2006): Leveraging Knowledge Communication for Innovation. Framework, Methods and Applications of Social Network Analysis in Research and Development, Frankfurt a. M. et al.: Peter Lang, ISBN 0-8204-9889-0.
18. Manski, Charles F. (2000). "Economic Analysis of Social Interactions". *Journal of Economic Perspectives* **14**: 115–36. [3] via JSTOR
19. Moody, James, and Douglas R. White (2003). "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups." *American Sociological Review* 68(1):103-127. [4]
20. Newman, Mark (2003). "The Structure and Function of Complex Networks". *SIAM Review* **56**: 167–256. doi:10.1137/S003614450342480. pdf
21. Nohria, Nitin and Robert Eccles (1992). *Networks in Organizations.* Second ed. Boston: Harvard Business Press. ISBN 0875843247
22. Nooy, Wouter d., A. Mrvar and Vladimir Batagelj. (2005). *Exploratory Social Network Analysis with Pajek.* Cambridge: Cambridge University Press. ISBN 0521841739

23. Scott, John. (2000). *Social Network Analysis: A Handbook.* 2nd Ed. Newberry Park, CA: Sage. ISBN 0761963383
24. Sethi, Arjun. (2008). *Valuation of Social Networking*
25. Tilly, Charles. (2005). *Identities, Boundaries, and Social Ties.* Boulder, CO: Paradigm press. ISBN 1594511314
26. Valente, Thomas W. (1995). *Network Models of the Diffusion of Innovations.* Cresskill, NJ: Hampton Press. ISBN 1881303217
27. Wasserman, Stanley, & Faust, Katherine. (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press. ISBN 0521382696
28. Watkins, Susan Cott. (2003). "Social Networks." Pp. 909–910 in *Encyclopedia of Population.* rev. ed. Edited by Paul George Demeny and Geoffrey McNicoll. New York: Macmillan Reference. ISBN 0028656776
29. Watts, Duncan J. (2003). *Small Worlds: The Dynamics of Networks between Order and Randomness.* Princeton: Princeton University Press. ISBN 0691117047
30. Watts, Duncan J. (2004). *Six Degrees: The Science of a Connected Age.* W. W. Norton & Company. ISBN 0393325423
31. Wellman, Barry (1998). *Networks in the Global Village: Life in Contemporary Communities.* Boulder, CO: Westview Press. ISBN 0813311500
32. Wellman, Barry. 2001. "Physical Place and Cyber-Place: Changing Portals and the Rise of Networked Individualism." *International Journal for Urban and Regional Research* 25 (2): 227-52.
33. Wellman, Barry and Berkowitz, Stephen D. (1988). *Social Structures: A Network Approach.* Cambridge: Cambridge University Press. ISBN 0521244412
34. Weng, M. (2007). *A Multimedia Social-Networking Community for Mobile Devices* Interactive Telecommunications Program, Tisch School of the Arts/ New York University
35. White, Harrison, Scott Boorman and Ronald Breiger. 1976. "Social Structure from Multiple Networks: I Blockmodels of Roles and Positions." *American Journal of Sociology* 81: 730-80.
36. The International Network for Social Network Analysis (INSNA) - professional society of social network analysts, with more than 1,000 members
37. Annual International Workshop on Social Network Mining and Analysis SNAKDD - Annual computer science workshop for interdisciplinary studies on social network mining and analysis

38. VisualComplexity.com - a visual exploration on mapping complicated and complex networks
39. Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon
40. NetLab at the University of Toronto, studies the intersection of social, communication, information and computing networks
41. Netwiki (wiki page devoted to social networks; maintained at University of North Carolina at Chapel Hill)
42. Building networks for learning- A guide to on-line resources on strengthening social networking.
43. Program on Networked Governance - Program on Networked Governance, Harvard University

11. Semantic network

A **semantic network** is a network which represents semantic relations among concepts. This is often used as a form of knowledge representation. It is a directed or undirected graph consisting of vertices, which represent concepts, and edges.



Example of a semantic network

History

"Semantic Nets" were first invented for computers by Richard H. Richens of the Cambridge Language Research Unit in 1956 as an "interlingua" for machine translation of natural languages.

They were developed by Robert F. Simmons at System Development Corporation in the early 1960s and later featured prominently in the work of Allan M. Collins and colleagues (e.g., Collins and Quillian; Collins and Loftus).

In the 1960s to 1980s the idea of a semantic link was developed within hypertext systems as the most basic unit, or edge, in a semantic network. These ideas were extremely influential, and there have been many attempts to add typed link semantics to HTML and XML.

Semantic network construction

WordNet

An example of a semantic network is WordNet, a lexical database of English. It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets. Some of the most common semantic relations defined are metonymy (A is part of B, i.e. B has A as a part of itself), homonymy (B is part of A, i.e. A has B as a part of itself), hyponymy (or *troponymy*) (A is subordinate of B; A is kind of B), hypernymy (A is superordinate of B), synonymy (A denotes the same as B) and antonymy (A denotes the opposite of B).

WordNet properties have been studied from a network theory perspective and compared to other semantic networks created from Roget's Thesaurus and word association tasks. From this perspective the three of them are a small world structure.[5]

It is also possible to represent logical descriptions using semantic networks such as the existential Graphs of Charles Sanders Peirce or the related Conceptual Graphs of John F. Sowa. These have expressive power equal to or exceeding standard first-order predicate logic. Unlike WordNet or other lexical or browsing networks, semantic networks using these representations can be used for reliable automated logical deduction. Some automated reasoners exploit the graph-theoretic features of the networks during processing.

Other examples

Other examples of semantic networks are Gellish models. Gellish English with its Gellish English dictionary, is a formal language that is defined as a network of relations between concepts and names of concepts. Gellish English is a formal subset of natural English, just as Gellish Dutch is a formal subset of Dutch, whereas multiple languages share the same concepts. Other Gellish networks consist of knowledge models and information models that are expressed in the Gellish language. A Gellish network is a network of (binary) relations between things. Each relation in the network is an expression of a fact that is classified by a relation type. Each relation type itself is a concept that is defined in the Gellish language dictionary.

Each related thing is either a concept or an individual thing that is classified by a concept. The definitions of concepts are created in the form of definition models (definition networks) that together form a Gellish Dictionary. A Gellish network can be documented in a Gellish database and is computer interpretable.

Software tools

There are also elaborate types of semantic networks connected with corresponding sets of software tools used for lexical knowledge engineering, like the Semantic Network Processing System (SNePS) of Stuart C. Shapiro or the MultiNet paradigm of Hermann Helbig, especially suited for the semantic representation of natural language expressions and used in several NLP applications.

References

1. John F. Sowa (1987). "Semantic Networks". in Stuart C Shapiro. *Encyclopedia of Artificial Intelligence*. http://www.jfsowa.com/pubs/semnet. htm. Retrieved 2008-04-29.
2. Allan M. Collins; M.R. Quillian (1969). "Retrieval time from semantic memory". *Journal of verbal learning and verbal behavior* **8** (2): 240–248. doi:10.1016/S0022-5371(69)80069-1.
3. Allan M. Collins; M. Ross Quillian (1970). "Does category size affect categorization time?". *Journal of verbal learning and verbal behavior* **9** (4): 432–438. doi:10.1016/S0022-5371(70)80084-6.
4. Allan M. Collins; Elizabeth F. Loftus (1975). "A spreading-activation theory of semantic processing". *Psychological Review* **82** (6): 407–428. doi:10.1037/0033-295X.82.6.407.
5. Steyvers, M.; Tenenbaum, J.B. (2005). "The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth". *Cognitive Science* **29** (1): 41–78. doi:10.1207/s15516709cog2901_3.
6. Allen, J. and A. Frisch (1982). "What's in a Semantic Network". In: *Proceedings of the 20th. annual meeting of ACL*, Toronto, pp. 19-27.
7. John F. Sowa, Alexander Borgida (1991). *Principles of Semantic Networks: Explorations in the Representation of Knowledge*.

12. Radio and Television networks

There are two types of radio networks currently in use around the world: the one-to-many broadcast type commonly used for public information and entertainment; and the two-way type used more commonly for public safety and public services such as police, fire, taxis, and delivery services. Following is a description of the former type of radio network although many of the same components and much of the same basic technology applies to both.

The **Broadcast** type of **radio network** is a network system which distributes programming to multiple stations simultaneously or slightly delayed, for the purpose of extending total coverage beyond the limits of a single broadcast signal. The resulting expanded audience for programming or information essentially applies the benefits of mass-production to the broadcasting enterprise. A radio network has two sales departments, one to package and sell programs to radio stations, and one to sell the audience of those programs to advertisers.

Most radio networks also produce much of their programming. Originally, radio networks owned some or all of the radio stations that broadcast the network's programming. Presently however, there are many networks that do not own any stations and only produce and/or distribute programming. Similarly station ownership does not always indicate network affiliation. A company might own stations in several different markets and purchase programming from a variety of networks.

Radio networks rose rapidly with the growth of regular broadcasting of radio to home listeners in the 1920s. This growth took various paths in different places. In Britain the BBC was developed with public funding, in the form of a broadcast receiving license, and a broadcasting monopoly in its early decades. In contrast, in the United States of America various competing commercial networks arose funded by advertising revenue. In that instance, the same corporation that owned or operated the network often manufactured and marketed the listener’s radio.

Major technical challenges to be overcome when distributing programs over long distances are maintaining signal quality and managing the number of switching/relay points in the signal chain. Early on, programs were sent to

remote stations (either owned or affiliated) by various methods, including leased telephone lines, pre-recorded gramophone records and audio tape. The world's first all-radio, non-wire line network was claimed to be the Rural Radio Network, a group of six upstate New York FM stations that began operation in June 1948. Terrestrial microwave relay, a technology later introduced to link stations, has been largely supplanted by coaxial cable, fiber, and satellite, which usually offer superior cost-benefit ratios.

Many early radio networks evolved into Television networks.

Radio network

The **Two-way** type of **radio network** shares many of the same technologies and components as the **Broadcast type radio network** but is generally set up with fixed broadcast points (transmitters) with co-located receivers and mobile receivers/transmitters or **Transceivers**. In this way both the fixed and mobile radio units can communicate with each other over broad geographic regions ranging in size from small single cities to entire states/provinces or countries. There are many ways in which multiple fixed transmit/receive sites can be interconnected to achieve the range of coverage required by the jurisdiction or authority implementing the system: conventional wireless links in numerous frequency bands, fibre-optic links, or micro-wave links. In all of these cases the signals are typically backhauled to a central switch of some type where the radio message is processed and resent (repeated) to all transmitter sites where it is required to be heard.

In contemporary two-way radio systems a concept called **trunking** is commonly used to achieve better efficiency of radio spectrum use and provide very wide ranging coverage with no switching of channels required by the mobile radio user as it roams throughout the system coverage. Trunking of two-way radio is identical to the concept used for cellular phone systems where each fixed and mobile radio is specifically identified to the system **Controller** and its operation is switched by the controller. See also the entries Two-way radio and Trunked radio system to see more detail on how various types of radios and radio systems work.

Television network

A **television network** is a distribution network for television content whereby a central operation provides programming for many television stations. Until the mid-1980s, television programming in most countries of the world was dominated by a small number of broadcast networks. Many

early television networks (e.g. the BBC, NBC or CBS) evolved from earlier radio networks.

In countries where most networks broadcast identical, centrally originated content to all their stations and where most individual transmitters therefore operate only as large "repeater stations", the terms *television network*, *television channel* and *television station* have become interchangeable in everyday language, with only professionals in TV-related occupations continuing to make a difference between them. Within the industry, a tiering is sometimes created among groups of networks based on whether their programming is simultaneously originated from a central point, and whether the network master control has the technical and administrative capability to take over the programming of their affiliates in real-time when it deems this necessary—the most common example being breaking national news events.

In North America in particular, many television channels available via cable and satellite television are branded as "networks" but are not truly networks in the sense defined above, as they are singular operations – they have no affiliates or component stations. Such channels are more precisely referred to by terms such as "specialty channels" (Canada) or "cable networks" (U.S.), although the latter term is somewhat of a misnomer, even though these channels are networked across the country by various cable and satellite systems.

A network may or may not produce all of its own programming. If not, production houses such as Warner Bros. and Sony Pictures can distribute their content to the different networks and it is common that a certain production house may have programmes on two or more rival networks. Similarly, some networks may import television programmes from other countries or use archival programming to help complement their schedules.

Regulation

FCC regulations in the United States restricted the number of television stations that could be owned by any one network, company or individual. This led to a system where most local television stations were independently owned, but received programming from the network through a franchising contract, except in a few big cities that had network owned-and-operated stations and independent stations. In the early days of television, when there were often only one or two stations broadcasting in an area, the stations were usually affiliated with several networks and were

able to choose which programs to air. Eventually, as more stations were licensed, it became common for each station to be affiliated with only one network and carry all of the "prime time" network programs. Local stations however occasionally break from regularly scheduled network programming, especially when there is breaking local news (e.g. severe weather). Moreover, when stations return to network programming from commercial breaks, the station's logo is displayed in the first few seconds before switching to the network's logo.

Another FCC regulation, the Prime Time Access Rule, restricted the number of hours of network programming that could be broadcast on the local affiliate stations. This was done to encourage the development of locally produced programs and to give local residents access to broadcast time. More often, the result included a substantial amount of syndicated programming, usually consisting of old movies, independently produced and syndicated shows, and reruns of network programs. Occasionally, these shows were presented by a local host, especially in programs that showed cartoons and short comedies intended for children. See List of local children's television series (United States).

13. Business networking

Business networking is a marketing method by which business opportunities are created through networks of like-minded business people. There are several prominent business networking organizations that create models of networking activity that, when followed, allow the business person to build new business relationship and generate business opportunities at the same time.

Many business people contend business networking is a more cost-effective method of generating new business than advertising or public relations efforts. This is because business networking is a low-cost activity that involves more personal commitment than company money.

As an example, a business network may agree to meet weekly or monthly with the purpose of exchanging business leads and referrals with fellow members. To complement this activity, members often meet outside this circle, on their own time, and build their own "one-to-one" relationship with the fellow member.

Business networking can be conducted in a local business community, or on a larger scale via the Internet. Business networking websites have grown over recent years due to the Internet's ability to connect people from all over the world.

Business networking can have a meaning also in the ICT domain, i.e. the provision of operating support to companies / organizations, and related value chains / value networks.

It refers to an activity coordination with a wider scope and a simpler implementation than pre-organized workflows or web-based

- impromptu searches for transaction counterparts (workflow is useful to coordinate activities, but it is complicated by the use of s.c. "patterns" to deviate the flow of work from a pure sequence, in order to compensate its intrinsic "linearity";

- impromptu searches for transaction counterparts on the web are useful as well, but only for non strategic supplies; both are complicated by a plethora of interfaces -- SOA / XML / web services -- needed among different organizations and even between different IT applications within the same organization).

Online business networking

Businesses are increasingly using business social networks like **Business Book** or professional business networking tools like Boardex as a means of growing their circle of business contacts and promoting themselves online. Since businesses are expanding globally, social networks make it easier to keep in touch with other contacts around the world. Specific cross-border e-commerce platforms and business partnering networks now make globalization accessible also for small and medium sized companies.

Face-to-face business networking

Professionals who wish to leverage their presentation skills with the urgency of physically being present attend general and exclusive events. Many professionals tend to prefer face-to-face networking over online based networking because the potential for higher quality relationships are possible. Many individuals also prefer face-to-face because people tend to prefer actually knowing and meeting who they intend to do business with.

General business networking

Before online networking, there was and has always been, networking face-to-face. "Schmoozing" or "rubbing elbows" are expressions used among business professionals for introducing and meeting one another, and establishing rapport.

Networked Businesses

With networking developing many businesses now have this as a core part of their strategy, those that have developed a strong network of connections suppliers and companies can be seen as "Networked Businesses" and will tend to source the business and their suppliers through the network of relationships that they have in place. Networked businesses tend to be Open, Random and Supportive - ORS whereas those relying on hierarchical, traditional managed approaches are Closed Selective and Controlling - CSC.

Business networking in the ICT domain

Companies / organizations - and related value chains / value networks - need some sort of IT support. Traditionally, it is provided by software applications, software packages /suites, ERPs and/or workflows; presently, also by different types of web-based innovations.

A truly "ICT" business networking approach rethinks - and rebuilds - the operating support from scratch, around two key business features: information contributions, to be provided by the activities involved (whether they are performed by human beings, automated tools or jointly by the two, in a coordinated way); (automated) information exchanges, to be provided by the TLC network.

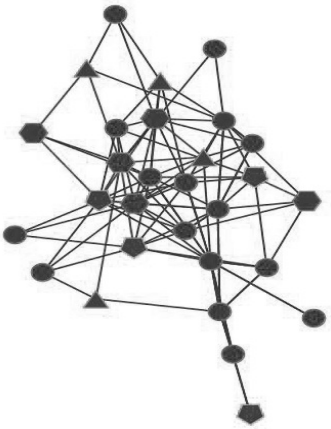
Information contributions and exchanges, in turn, need to be supported by data storage (plain or redundant, with or without automated recovery to grant service continuity) and access security (signature, encryption, authentication, decryption), which both can be provided either as add-ons or as built-in features.

14. Dynamic network analysis

Dynamic network analysis (DNA) is an emergent scientific field that brings together traditional social network analysis (SNA), link analysis (LA) and multi-agent systems (MAS) within network science and network theory. There are two aspects of this field. The first is the statistical analysis of DNA data. The second is the utilization of simulation to address issues of network dynamics. DNA networks vary from traditional social networks in that they are larger, dynamic, multi-mode, multi-plex networks, and may contain varying levels of uncertainty.

DNA statistical tools are generally optimized for large-scale networks and admit the analysis of multiple networks simultaneously in which, there are multiple types of nodes (multi-node) and multiple types of links (multi-plex). In contrast, SNA statistical tools focus on single or at most two mode data and facilitate the analysis of only one type of link at a time.

DNA statistical tools tend to provide more measures to the user, because they have measures that use data drawn from multiple networks simultaneously.



An example of a multi-entity, multi-network, dynamic network diagram

From a computer simulation perspective, nodes in DNA are like atoms in quantum theory, nodes can be, though need not be, and treated as

probabilistic. Whereas nodes in a traditional SNA model are static, nodes in a DNA model have the ability to learn. Properties change over time; nodes can adapt: A company's employees can learn new skills and increase their value to the network; or, capture one terrorist and three more are forced to improvise. Change propagates from one node to the next and so on. DNA adds the element of a network's evolution and considers the circumstances under which change is likely to occur.

Some problems that people in the DNA area work on

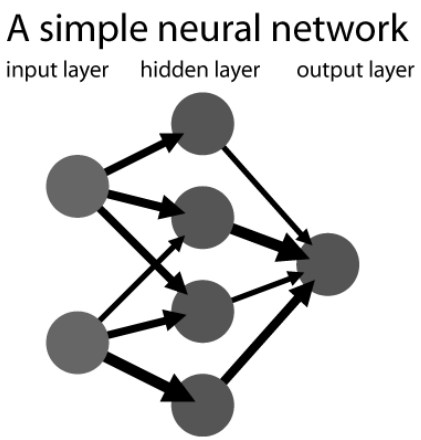
- Developing metrics and statistics to assess and identify change within and across networks.
- Developing and validating simulations to study network change, evolution, adaptation, decay... Computer simulation and organizational studies.
- Developing and validating formal models of network generation.
- Developing and testing theory of network change, evolution, adaptation, decay.
- Developing techniques to visualize network change overall or at the node or group level.
- Developing statistical techniques to see whether differences observed over time in networks are due to simply different samples from a distribution of links and nodes or changes over time in the underlying distribution of links and nodes.
- Developing control processes for networks over time.
- Developing algorithms to change distributions of links in networks over time.
- Developing algorithms to track groups in networks over time.
- Developing tools to extract or locate networks from various data sources such as texts.
- Developing statistically valid measurements on networks over time.
- Examining the robustness of network metrics under various types of missing data.
- Empirical studies of multi-mode multi-link multi-time period networks.
- Examining networks as probabilistic time-variant phenomena.
- Forecasting change in existing networks.
- Identifying trails through time given a sequence of networks.
- Identifying changes in node criticality given a sequence of networks anything else related to multi-mode multi-link multi-time period networks.

References

1. Kathleen M. Carley, 2003, "Dynamic Network Analysis" in Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers, Ronald Breiger, Kathleen Carley, and Philippa Pattison, (Eds.) Committee on Human Factors, National Research Council, National Research Council. Pp. 133–145, Washington, DC.
2. Kathleen M. Carley, 2002, "Smart Agents and Organizations of the Future" The Handbook of New Media. Edited by Leah Lievrouw and Sonia Livingstone, Ch. 12, pp. 206–220, Thousand Oaks, CA, Sage.
3. Kathleen M. Carley, Jana Diesner, Jeffrey Reminga, Maksim Tsvetovat, 2008, Toward an Interoperable Dynamic Network Analysis Toolkit, DSS Special Issue on Cyberinfrastructure for Homeland Security: Advances in Information Sharing, Data Mining, and Collaboration Systems. Decision Support Systems 43(4):1324-1347 (article 20)
4. Terrill L. Frantz, Kathleen M. Carley. 2009, Toward A Confidence Estimate For The Most-Central-Actor Finding. Academy of Management Annual Conference, Chicago, IL, USA, 7-11 August. (Awarded the Sage Publications/RM Division Best Student Paper Award)
5. Radcliffe Exploratory Seminar on Dynamic Networks http://www.eecs.harvard.edu/~parkes/RadcliffeSeminar.htm
6. Center for Computational Analysis of Social and Organizational Systems (CASOS) http://www.casos.cs.cmu.edu/

15. Neural network

Traditionally, the term **neural network** had been used to refer to a network or circuit of biological neurons. The modern usage of the term often refers to artificial neural networks, which are composed of artificial neurons or nodes. Thus the term has two distinct usages:



Biological neural networks are made up of real biological neurons that are connected or functionally related in the peripheral nervous system or the central nervous system. In the field of neuroscience, they are often identified as groups of neurons that perform a specific physiological function in laboratory analysis.

Artificial neural networks are made up of interconnecting artificial neurons (programming constructs that mimic the properties of biological neurons). Artificial neural networks may either be used to gain an understanding of biological neural networks, or for solving artificial intelligence problems without necessarily creating a model of a real biological system. The real, biological nervous system is highly complex and includes some features that may seem superfluous based on an understanding of artificial networks.

This picture is a simplified view of a feed forward artificial neural network.

This article focuses on the relationship between the two concepts; for detailed coverage of the two different concepts refer to the separate articles: Biological neural network and artificial neural network.

Overview

In general a biological neural network is composed of a group or groups of chemically connected or functionally associated neurons. A single neuron may be connected to many other neurons and the total number of neurons

and connections in a network may be extensive. Connections, called synapses, are usually formed from axons to dendrites, though dendrodendritic microcircuits and other connections are possible. Apart from the electrical signaling, there are other forms of signaling that arise from neurotransmitter diffusion, which have an effect on electrical signaling. As such, neural networks are extremely complex.

Artificial intelligence and cognitive modeling try to simulate some properties of neural networks. While similar in their techniques, the former has the aim of solving particular tasks, while the latter aims to build mathematical models of biological neural systems.

In the artificial intelligence field, artificial neural networks have been applied successfully to speech recognition, image analysis and adaptive control, in order to construct software agents (in computer and video games) or autonomous robots. Most of the currently employed artificial neural networks for artificial intelligence are based on statistical estimation, optimization and control theory.

The cognitive modeling field involves the physical or mathematical modeling of the behavior of neural systems; ranging from the individual neural level (e.g. modeling the spike response curves of neurons to a stimulus), through the neural cluster level (e.g. modeling the release and effects of dopamine in the basal ganglia) to the complete organism (e.g. behavioral modeling of the organism's response to stimuli). Artificial intelligence, cognitive modeling, and neural networks are information processing paradigms inspired by the way biological neural systems process data.

History of the neural network analogy

The concept of neural networks started in the late-1800s as an effort to describe how the human mind performed. These ideas started being applied to computational models with Turing's B-type machines and the perceptron.

In early 1950s Friedrich Hayek was one of the first to posit the idea of spontaneous order in the brain arising out of decentralized networks of simple units (neurons). In the late 1940s, Donald Hebb made one of the first hypotheses for a mechanism of neural plasticity (i.e. learning), Hebbian learning. Hebbian learning is considered to be a 'typical' unsupervised learning rule and it (and variants of it) was an early model for long term potentiation.

The perceptron is essentially a linear classifier for classifying data $x \in \mathbb{R}^n$ specified by parameters $w \in \mathbb{R}^n, b \in \mathbb{R}$ and an output function $f = w'x + b$.

Its parameters are adapted with an ad-hoc rule similar to stochastic steepest gradient descent. Because the inner product is a linear operator in the input space, the Perceptron can only perfectly classify a set of data for which different classes are linearly separable in the input space, while it often fails completely for non-separable data. While the development of the algorithm initially generated some enthusiasm, partly because of its apparent relation to biological mechanisms, the later discovery of this inadequacy caused such models to be abandoned until the introduction of non-linear models into the field.

The cognitron (1975) was an early multilayered neural network with a training algorithm. The actual structure of the network and the methods used to set the interconnection weights change from one neural strategy to another, each with its advantages and disadvantages. Networks can propagate information in one direction only, or they can bounce back and forth until self-activation at a node occurs and the network settles on a final state.

The ability for bi-directional flow of inputs between neurons/nodes was produced with the Hopfield's network (1982), and specialization of these node layers for specific purposes was introduced through the first hybrid network.

The parallel distributed processing of the mid-1980s became popular under the name connectionism.

The rediscovery of the backpropagation algorithm was probably the main reason behind the repopularisation of neural networks after the publication of "Learning Internal Representations by Error Propagation" in 1986 (Though backpropagation itself dates from 1974).

The original network utilized multiple layers of weight-sum units of the type $f = g(w'x + b)$, where g was a sigmoid function or logistic function such as used in logistic regression. Training was done by a form of stochastic steepest gradient descent. The employment of the chain rule of differentiation in deriving the appropriate parameter updates results in an algorithm that seems to 'backpropagate errors', hence the nomenclature.

However it is essentially a form of gradient descent. Determining the optimal parameters in a model of this type is not trivial, and steepest gradient descent methods cannot be relied upon to give the solution

without a good starting point. In recent times, networks with the same architecture as the backpropagation network are referred to as Multi-Layer Perceptrons. This name does not impose any limitations on the type of algorithm used for learning.

The backpropagation network generated much enthusiasm at the time and there was much controversy about whether such learning could be implemented in the brain or not, partly because a mechanism for reverse signaling was not obvious at the time, but most importantly because there was no plausible source for the 'teaching' or 'target' signal.

The brain, neural networks and computers

Neural networks, as used in artificial intelligence, have traditionally been viewed as simplified models of neural processing in the brain, even though the relation between this model and brain biological architecture is debated.

A subject of current research in theoretical neuroscience is the question surrounding the degree of complexity and the properties that individual neural elements should have to reproduce something resembling animal intelligence.

Historically, computers evolved from the von Neumann architecture, which is based on sequential processing and execution of explicit instructions. On the other hand, the origins of neural networks are based on efforts to model information processing in biological systems, which may rely largely on parallel processing as well as implicit instructions based on recognition of patterns of 'sensory' input from external sources. In other words, at its very heart a neural network is a complex statistical processor (as opposed to being tasked to sequentially process and execute).

Neural networks and artificial intelligence

A *neural network* (NN). in the case of artificial neurons called *artificial neural network* (ANN) or *simulated neural network* (SNN), is an interconnected group of natural or artificial neurons that uses a mathematical or computational model for information processing based on a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network.

In more practical terms neural networks are non-linear statistical data modeling or decision making tools. They can be used to model complex

relationships between inputs and outputs or to find patterns in data.

However, the paradigm of neural networks - i.e., *implicit*, and not *explicit* learning is stressed - seems more to correspond to some kind of *natural intelligence* than to the traditional *Artificial Intelligence*, which would stress, instead, rule-based learning.

Background

An artificial neural network involves a network of simple processing elements (artificial neurons) which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. Artificial neurons were first proposed in 1943 by Warren McCulloch, a neurophysiologist, and Walter Pitts, an MIT logician. One classical type of artificial neural network is the recurrent Hopfield net.

In a neural network model simple nodes, which can be called variously "neurons", "neurodes", "Processing Elements" (PE) or "units", are connected together to form a network of nodes — hence the term "neural network". While a neural network does not have to be adaptive *per se*, its practical use comes with algorithms designed to alter the strength (weights) of the connections in the network to produce a desired signal flow.

In modern software implementations of artificial neural networks the approach inspired by biology has more or less been abandoned for a more practical approach based on statistics and signal processing. In some of these systems, neural networks, or parts of neural networks (such as artificial neurons), are used as components in larger systems that combine both adaptive and non-adaptive elements.

The concept of a neural network appears to have first been proposed by Alan Turing in his 1948 paper "Intelligent Machinery".

Applications of natural and of artificial neural networks

The utility of artificial neural network models lies in the fact that they can be used to infer a function from observations and also to use it. This is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical.

Real life applications

The tasks to which artificial neural networks are applied tend to fall within the following broad categories:

- Function approximation, or regression analysis, including time series prediction and modeling.
- Classification, including pattern and sequence recognition, novelty detection and sequential decision making.
- Data processing, including filtering, clustering, blind signal separation and compression.

Application areas of ANNs include system identification and control (vehicle control, process control), game-playing and decision making (backgammon, chess, racing), pattern recognition (radar systems, face identification, object recognition, etc.), sequence recognition (gesture, speech, handwritten text recognition), medical diagnosis, financial applications, data mining (or knowledge discovery in databases, "KDD"), visualization and e-mail spam filtering.

- Moreover, some brain diseases, e.g. Alzheimer, are apparently, and essentially, diseases of the brain's natural NN by damaging necessary prerequisites for the functioning of the mutual interconnections between neurons and/or glia.

Neural network software

Neural network software is used to simulate, research, develop and apply artificial neural networks, biological neural networks and in some cases a wider array of adaptive systems.

Learning paradigms

There are three major learning paradigms, each corresponding to a particular abstract learning task. These are supervised learning, unsupervised learning and reinforcement learning. Usually any given type of network architecture can be employed in any of those tasks.

Supervised learning

In supervised learning, we are given a set of example pairs $(x, y), x \in X, y \in Y$ and the aim is to find a function *f* in the allowed class of functions that matches the examples. In other words, we wish to *infer* how the mapping implied by the data and the cost function is related to the mismatch between our mapping and the data.

Unsupervised learning

In unsupervised learning we are given some data *x*, and a cost function which is to be minimized which can be any function of *x* and the network's

output, *f*. The cost function is determined by the task formulation. Most applications fall within the domain of estimation problems such as statistical modeling, compression, filtering, blind source separation and clustering.

Reinforcement learning

In reinforcement learning, data *x* is usually not given, but generated by an agent's interactions with the environment. At each point in time *t*, the agent performs an action y_t and the environment generates an observation x_t and an instantaneous cost c_t, according to some (usually unknown) dynamics. The aim is to discover a *policy* for selecting actions that minimizes some measure of a long-term cost, i.e. the expected cumulative cost. The environment's dynamics and the long-term cost for each policy are usually unknown, but can be estimated. ANNs are frequently used in reinforcement learning as part of the overall algorithm. Tasks that fall within the paradigm of reinforcement learning are control problems, games and other sequential decision making tasks.

Learning algorithms

There are many algorithms for training neural networks; most of them can be viewed as a straightforward application of optimization theory and statistical estimation. They include: Back propagation by gradient descent, Rprop, BFGS, CG etc.

Evolutionary computation methods simulated annealing, expectation maximization and non-parametric methods are among other commonly used methods for training neural networks. See also machine learning.

Recent developments in this field also saw the use of particle swarm optimization and other swarm intelligence techniques used in the training of neural networks.

Neural networks and neuroscience

Theoretical and computational neuroscience is the field concerned with the theoretical analysis and computational modeling of biological neural systems. Since neural systems are intimately related to cognitive processes and behavior, the field is closely related to cognitive and behavioral modeling.

The aim of the field is to create models of biological neural systems in order to understand how biological systems work. To gain this understanding, neuroscientists strive to make a link between observed biological processes

(data), biologically plausible mechanisms for neural processing and learning (biological neural network models) and theory (statistical learning theory and information theory).

Types of models

Many models are used in the field, each defined at a different level of abstraction and trying to model different aspects of neural systems. They range from models of the short-term behavior of individual neurons, through models of how the dynamics of neural circuitry arise from interactions between individual neurons, to models of how behavior can arise from abstract neural modules that represent complete subsystems. These include models of the long-term and short-term plasticity of neural systems and its relation to learning and memory, from the individual neuron to the system level.

Current research

While initially research had been concerned mostly with the electrical characteristics of neurons, a particularly important part of the investigation in recent years has been the exploration of the role of neuromodulators such as dopamine, acetylcholine, and serotonin on behavior and learning.

Biophysical models, such as BCM theory, have been important in understanding mechanisms for synaptic plasticity, and have had applications in both computer science and neuroscience. Research is ongoing in understanding the computational algorithms used in the brain, with some recent biological evidence for radial basis networks and neural backpropagation as mechanisms for processing data.

Criticism

A common criticism of neural networks, particularly in robotics, is that they require a large diversity of training for real-world operation. Dean Pomerleau, in his research presented in the paper "Knowledge-based Training of Artificial Neural Networks for Autonomous Robot Driving," uses a neural network to train a robotic vehicle to drive on multiple types of roads (single lane, multi-lane, dirt, etc.). A large amount of his research is devoted to (1) extrapolating multiple training scenarios from a single training experience, and (2) preserving past training diversity so that the system does not become over trained (if, for example, it is presented with a series of right turns – it should not learn to always turn right). These issues

are common in neural networks that must decide from amongst a wide variety of responses.

A. K. Dewdney, a former *Scientific American* columnist, wrote in 1997, "Although neural nets do solve a few toy problems, their powers of computation are so limited that I am surprised anyone takes them seriously as a general problem-solving tool." (Dewdney, p. 82)

Arguments for Dewdney's position are that to implement large and effective software neural networks, much processing and storage resources need to be committed. While the brain has hardware tailored to the task of processing signals through a graph of neurons, simulating even a most simplified form on Von Neumann technology may compel a NN designer to fill many millions of database rows for its connections - which can lead to abusive RAM and HD necessities.

Furthermore, the designer of NN systems will often need to simulate the transmission of signals through many of these connections and their associated neurons - which must often be matched with incredible amounts of CPU processing power and time. While neural networks often yield *effective* programs, they too often do so at the cost of time and money *efficiency*.

Arguments against Dewdney's position are that neural nets have been successfully used to solve many complex and diverse tasks, ranging from autonomously flying aircraft to detecting credit card fraud.

Technology writer Roger Bridgman commented on Dewdney's statements about neural nets:

Neural networks, for instance, are in the dock not only because they have been hyped to high heaven, (what hasn't?) but also because you could create a successful net without understanding how it worked: the bunch of numbers that captures its behavior would in all probability be "an opaque, unreadable table...valueless as a scientific resource". In spite of his emphatic declaration that science is not technology, Dewdney seems here to pillory neural nets as bad science when most of those devising them are just trying to be good engineers. An unreadable table that a useful machine could read would still be well worth having.

Some other criticisms came from believers of hybrid models (combining neural networks and symbolic approaches). They advocate the intermix of these two approaches and believe that hybrid models can better capture the mechanisms of the human mind (Sun and Bookman 1994).

References

1. Roger Bridgman's defense of neural networks. http://members.fortunecity.com/templarseries/popper.html
2. Arbib, Michael A. (Ed.) (1995). *The Handbook of Brain Theory and Neural Networks*.
3. Alspector, U.S. Patent 4,874,963 "*Neuromorphic learning networks*". October 17, 1989.
4. Agre, Philip E. (1997). *Computation and Human Experience*. Cambridge Univ. Press. ISBN 0-521-38603-9., p. 80
5. Yaneer Bar-Yam (2003). *Dynamics of Complex Systems, Chapter 2*.
6. Yaneer Bar-Yam (2003). *Dynamics of Complex Systems, Chapter 3*.
7. Yaneer Bar-Yam (2005). *Making Things Work*. See chapter 3.
8. Bertsekas, Dimitri P. (1999). *Nonlinear Programming*.
9. Bertsekas, Dimitri P. & Tsitsiklis, John N. (1996). *Neuro-dynamic Programming*.
10. Bhadeshia H. K. D. H. (1992). "Neural Networks in Materials Science". *ISIJ International* **39**: 966–979. doi:10.2355/isijinternational.39.966.
11. Boyd, Stephen & Vandenberghe, Lieven (2004). *Convex Optimization*.
12. Dewdney, A. K. (1997). *Yes, We Have No Neutrons: An Eye-Opening Tour through the Twists and Turns of Bad Science*. Wiley, 192 pp. See chapter 5.
13. Egmont-Petersen, M., de Ridder, D., Handels, H. (2002). "Image processing with neural networks - a review". *Pattern Recognition* 35 (10): 2279–2301. doi:10.1016/S0031-3203(01)00178-9.
14. Fukushima, K. (1975). "Cognitron: A Self-Organizing Multilayered Neural Network". Biological Cybernetics 20: 121–136. doi:10.1007/BF00342633.
15. Frank, Michael J. (2005). "Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Non-medicated Parkinsonism". Journal of Cognitive Neuroscience 17: 51–72. doi:10.1162/0898929052880093.
16. Gardner, E.J., & Derrida, B. (1988). "Optimal storage properties of neural network models". Journal of Physics a 21: 271–284. doi:10.1088/0305-4470/21/1/031.
17. Krauth, W., & Mezard, M. (1989). "Storage capacity of memory with binary couplings". Journal de Physique 50: 3057–3066. doi:10.1051/jphys: 0198900500200305700.
18. Maass, W., & Markram, H. (2002). "On the computational power of

recurrent circuits of spiking neurons". Journal of Computer and System Sciences 69(4): 593–616.

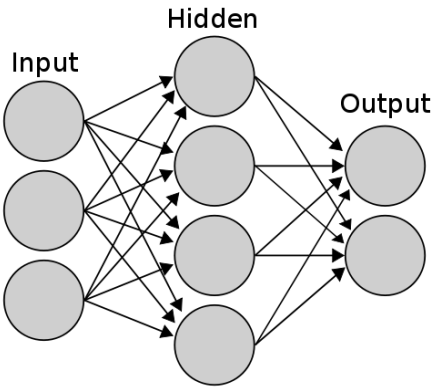
19. MacKay, David (2003). Information Theory, Inference, and Learning Algorithms.
20. Mandic, D. & Chambers, J. (2001). Recurrent Neural Networks for Prediction: Architectures, Learning algorithms and Stability. Wiley.
21. Minsky, M. & Papert, S. (1969). An Introduction to Computational Geometry. MIT Press.
22. Muller, P. & Insua, D.R. (1995). "Issues in Bayesian Analysis of Neural Network Models". Neural Computation 10: 571–592.
23. Reilly, D.L., Cooper, L.N. & Elbaum, C. (1982). "A Neural Model for Category Learning". Biological Cybernetics 45: 35–41. doi:10.1007/BF00387211.
24. Rosenblatt, F. (1962). Principles of Neurodynamics. Spartan Books.
25. Sun, R. & Bookman,L. (eds.) (1994.). Computational Architectures Integrating Neural and Symbolic Processes.. Kluwer Academic Publishers, Needham, MA..
26. Sutton, Richard S. & Barto, Andrew G. (1998). Reinforcement Learning : An introduction.
27. Van den Bergh, F. Engelbrecht, AP. Cooperative Learning in Neural Networks using Particle Swarm Optimizers. CIRG 2000.
28. Wilkes, A.L. & Wade, N.J. (1997). "Bain on Neural Networks". Brain and Cognition 33: 295–305. doi:10.1006/brcg.1997.0869.
29. Wasserman, P.D. (1989). Neural computing theory and practice. Van Nostrand Reinhold.
30. Jeffrey T. Spooner, Manfredi Maggiore, and Kevin M. Passino: Stable Adaptive Control and Estimation for Nonlinear Systems: Neural and Fuzzy Approximator Techniques, John Wiley and Sons, NY, 2002.
31. Peter Dayan, L.F. Abbott. *Theoretical Neuroscience*. MIT Press.
32. Wulfram Gerstner, Werner Kistler. *Spiking Neuron Models: Single Neurons, Populations, Plasticity*. Cambridge University Press.
33. Steeb, W-H (2008). *The Nonlinear Workbook: Chaos, Fractals, Neural Networks, Genetic Algorithms, Gene Expression Programming, Support Vector Machine, Wavelets, Hidden Markov Models, Fuzzy Logic with C++, Java and SymbolicC++ Programs: 4th edition*. World Scientific Publishing. ISBN 981-281-852-9.

Some on-line references

1. International Neural Network Society (INNS)
2. European Neural Network Society (ENNS)
3. Japanese Neural Network Society (JNNS)
4. IEEE Computational Intelligence Society (IEEE CIS)
5. A Brief Introduction to Neural Networks (D. Kriesel) - Illustrated, bilingual manuscript about artificial neural networks; Topics so far: Perceptrons, Backpropagation, Radial Basis Functions, Recurrent Neural Networks, Self Organizing Maps, Hopfield Networks.
6. Flood: An open source neural networks C++ library
7. LearnArtificialNeuralNetworks - Robot control and neural networks
8. NPatternRecognizer - A fast machine learning algorithm library written in C#. It contains SVM, neural networks, Bayes, boost, k-nearest neighbor, decision tree, ..., etc.
9. Review of Neural Networks in Materials Science
10. Artificial Neural Networks Tutorial in three languages (Univ. Politécnica de Madrid)
11. Introduction to Neural Networks and Knowledge Modeling
12. Introduction to Artificial Neural Networks
13. In Situ Adaptive Tabulation: - A neural network alternative.
14. Another introduction to ANN
15. Prediction with neural networks - includes Java applet for online experimenting with prediction of a function
16. Next Generation of Neural Networks - Google Tech Talks
17. Perceptual Learning - Artificial Perceptual Neural Network used for machine learning to play Chess
18. European Centre for Soft Computing
19. Performance of Neural Networks
20. Neural Networks and Information

16. Artificial neural network

An **artificial neural network (ANN)**, usually called "neural network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks. It consists of an interconnected group of artificial neurons and processes information using a connectionist approach to computation. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase. Neural networks are non-linear statistical data modeling tools. They can be used to model complex relationships between inputs and outputs or to find patterns in data.



A neural network is an interconnected group of nodes, akin to the vast network of neurons in the human brain.

Background

There is no precise agreed-upon definition among researchers as to what a neural network is, but most would agree that it involves a network of simple processing elements (neurons), which can exhibit complex global behavior, determined by the connections between the processing elements and element parameters. The original inspiration for the technique came from examination of the central nervous system and the neurons (and their axons, dendrites and synapses) which constitute one of its most significant information processing elements (see neuroscience). In a neural network model, simple nodes, called variously "neurons", "neurodes", "PEs" ("processing elements") or "units", are connected together to form a network of nodes — hence the term "neural network". While a neural network does not have to be adaptive per se, its practical use comes with algorithms designed to alter the strength (weights) of the connections in the network to produce a desired signal flow.

These networks are also similar to the biological neural networks in the sense that functions are performed collectively and in parallel by the units, rather than there being a clear delineation of subtasks to which various units are assigned (see also connectionism). Currently, the term Artificial Neural Network (ANN) tends to refer mostly to neural network models employed in statistics, cognitive psychology and artificial intelligence. Neural network models designed with emulation of the central nervous system (CNS) in mind are a subject of theoretical neuroscience (computational neuroscience).

In modern software implementations of artificial neural networks the approach inspired by biology has for the most part been abandoned for a more practical approach based on statistics and signal processing. In some of these systems, neural networks or parts of neural networks (such as artificial neurons) are used as components in larger systems that combine both adaptive and non-adaptive elements. While the more general approach of such adaptive systems is more suitable for real-world problem solving, it has far less to do with the traditional artificial intelligence connectionist models. What they do have in common, however, is the principle of non-linear, distributed, parallel and local processing and adaptation.

Models

Neural network models in artificial intelligence are usually referred to as artificial neural networks (ANNs); these are essentially simple mathematical models defining a function $f : X \to Y$. Each type of ANN model corresponds to a *class* of such functions.

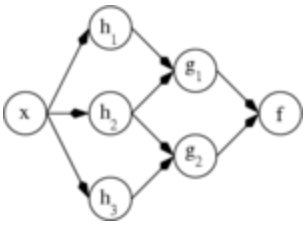
The network in artificial neural network

The word *network* in the term 'artificial neural network' arises because the function $f(x)$ is defined as a composition of other functions $g_i(x)$, which can further be defined as a composition of other functions. This can be conveniently represented as a network structure, with arrows depicting the dependencies between variables. A widely used type of composition is the *nonlinear weighted sum*, where $f(x) = K\left(\sum_i w_i g_i(x)\right)$, where K (commonly referred to as the activation function[1]) is some predefined function, such as the hyperbolic tangent. It will be convenient for the following to refer to a collection of functions g_i as simply a vector $g = (g_1, g_2, \ldots, g_n)$.

ANN dependency graph

This figure depicts such a decomposition of *f*, with dependencies between variables indicated by arrows. These can be interpreted in two ways.

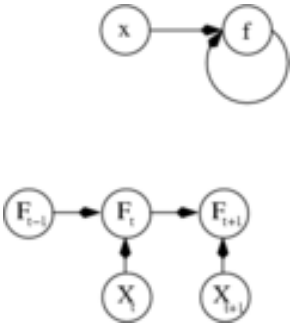
The first view is the functional view: the input *x* is transformed into a 3-dimensional vector *h*, which is then transformed into a 2-dimensional vector *g*, which is finally transformed into *f*. This view is most commonly encountered in the context of optimization.



The second view is the probabilistic view: the random variable $F = f(G)$ depends upon the random variable $G = g(H)$, which depends upon $H = h(X)$, which depends upon the random variable *X*. This view is most commonly encountered in the context of graphical models.

The two views are largely equivalent. In either case, for this particular network architecture, the components of individual layers are independent of each other (e.g., the components of *g* are independent of each other given their input *h*). This naturally enables a degree of parallelism in the implementation.

Recurrent ANN dependency graph



Networks such as the previous one are commonly called feedforward, because their graph is a directed acyclic graph. Networks with cycles are commonly called recurrent. Such networks are commonly depicted in the manner shown at the top of the figure, where *f* is shown as being dependent upon itself. However, there is an implied temporal dependence which is not shown.

Learning

What has attracted the most interest in neural networks is the possibility of *learning*. Given a specific *task* to solve, and a *class* of functions *F*, learning means using a set of *observations* to find $f^* \in F$ which solves the task in some *optimal* sense.

This entails defining a cost function $C : F \rightarrow \mathbb{R}$ such that, for the optimal solution f^*, $C(f^*) \leq C(f) \forall f \in F$ (i.e., no solution has a cost less than the cost of the optimal solution).

The cost function *C* is an important concept in learning, as it is a measure of how far away a particular solution is from an optimal solution to the problem to be solved. Learning algorithms search through the solution space to find a function that has the smallest possible cost.

For applications where the solution is dependent on some data, the cost must necessarily be a *function of the observations*; otherwise we would not be modeling anything related to the data. It is frequently defined as a statistic to which only approximations can be made. As a simple example consider the problem of finding the model f which minimizes $C = E\left[(f(x) - y)^2\right]$, for data pairs (*x*,*y*) drawn from some distribution $\mathcal{D}$. In practical situations we would only have *N* samples from $\mathcal{D}$and thus, for the above example, we would only minimize $\hat{C} = \frac{1}{N}\sum_{i=1}^{N}(f(x_i) - y_i)^2$. Thus, the cost is minimized over a sample of the data rather than the entire data set.

When $N \rightarrow \infty$some form of online machine learning must be used, where the cost is partially minimized as each new example is seen. While online machine learning is often used when $\mathcal{D}$is fixed, it is most useful in the case where the distribution changes slowly over time. In neural network methods, some form of online machine learning is frequently used for finite datasets.

Choosing a cost function

While it is possible to define some arbitrary, ad hoc cost function, frequently a particular cost will be used, either because it has desirable properties (such as convexity) or because it arises naturally from a particular formulation of the problem (e.g., in a probabilistic formulation the posterior probability of the model can be used as an inverse cost). Ultimately, the cost function will depend on the task we wish to perform. The three main categories of learning tasks are overviewed below.

Learning paradigms

There are three major learning paradigms, each corresponding to a

particular abstract learning task. These are supervised learning, unsupervised learning and reinforcement learning. Usually any given type of network architecture can be employed in any of those tasks.

Supervised learning

In supervised learning, we are given a set of example pairs $(x, y), x \in X, y \in Y$ and the aim is to find a function $f : X \to Y$ in the allowed class of functions that matches the examples. In other words, we wish to *infer* the mapping implied by the data; the cost function is related to the mismatch between our mapping and the data and it implicitly contains prior knowledge about the problem domain.

A commonly used cost is the mean-squared error which tries to minimize the average squared error between the network's output, f(x), and the target value y over all the example pairs. When one tries to minimize this cost using gradient descent for the class of neural networks called Multi-Layer Perceptrons, one obtains the common and well-known backpropagation algorithm for training neural networks.

Tasks that fall within the paradigm of supervised learning are pattern recognition (also known as classification) and regression (also known as function approximation). The supervised learning paradigm is also applicable to sequential data (e.g., for speech and gesture recognition). This can be thought of as learning with a "teacher," in the form of a function that provides continuous feedback on the quality of solutions obtained thus far.

Unsupervised learning

In unsupervised learning we are given some data *x* and the cost function to be minimized, that can be any function of the data *x* and the network's output, *f*.

The cost function is dependent on the task (what we are trying to model) and our *a priori* assumptions (the implicit properties of our model, its parameters and the observed variables).

As a trivial example, consider the model $f(x) = a$, where a is a constant and the cost $C = E[(x - f(x))^2]$. Minimizing this cost will give us a value of a that is equal to the mean of the data. The cost function can be much more complicated. Its form depends on the application: for example, in compression it could be related to the mutual information between x and y, whereas in statistical modeling, it could be related to the posterior

probability of the model given the data. (Note that in both of those examples those quantities would be maximized rather than minimized).

Tasks that fall within the paradigm of unsupervised learning are in general estimation problems; the applications include clustering, the estimation of statistical distributions, compression and filtering.

Reinforcement learning

In reinforcement learning, data *x* are usually not given, but generated by an agent's interactions with the environment. At each point in time *t*, the agent performs an action y_t and the environment generates an observation x_t and an instantaneous cost c_t, according to some (usually unknown) dynamics. The aim is to discover a *policy* for selecting actions that minimizes some measure of a long-term cost; i.e., the expected cumulative cost. The environment's dynamics and the long-term cost for each policy are usually unknown, but can be estimated.

More formally, the environment is modeled as a Markov decision process (MDP) with states $s_1, ..., s_n \in S$ and actions $a_1, ..., a_m \in A$ with the following probability distributions: the instantaneous cost distribution $P(c_t \mid s_t)$, the observation distribution $P(x_t \mid s_t)$ and the transition $P(s_{t+1} \mid s_t, a_t)$, while a policy is defined as conditional distribution over actions given the observations. Taken together, the two define a Markov chain (MC). The aim is to discover the policy that minimizes the cost; i.e., the MC for which the cost is minimal.

ANNs are frequently used in reinforcement learning as part of the overall algorithm.

Tasks that fall within the paradigm of reinforcement learning are control problems, games and other sequential decision making tasks.

Learning algorithms

Training a neural network model essentially means selecting one model from the set of allowed models (or, in a Bayesian framework, determining a distribution over the set of allowed models) that minimizes the cost criterion. There are numerous algorithms available for training neural network models; most of them can be viewed as a straightforward application of optimization theory and statistical estimation.

Most of the algorithms used in training artificial neural networks employ some form of gradient descent. This is done by simply taking the derivative

of the cost function with respect to the network parameters and then changing those parameters in a gradient-related direction.

Evolutionary methods, simulated annealing, expectation-maximization and non-parametric methods are some commonly used methods for training neural networks. See also machine learning.

Temporal perceptual learning relies on finding temporal relationships in sensory signal streams. In an environment, statistically salient temporal correlations can be found by monitoring the arrival times of sensory signals. This is done by the perceptual network.

Employing artificial neural networks

Perhaps the greatest advantage of ANNs is their ability to be used as an arbitrary function approximation mechanism which 'learns' from observed data. However, using them is not so straightforward and a relatively good understanding of the underlying theory is essential.

- Choice of model: This will depend on the data representation and the application. Overly complex models tend to lead to problems with learning.
- Learning algorithm: There are numerous tradeoffs between learning algorithms. Almost any algorithm will work well with the *correct hyperparameters* for training on a particular fixed dataset. However selecting and tuning an algorithm for training on unseen data requires a significant amount of experimentation.
- Robustness: If the model, cost function and learning algorithm are selected appropriately the resulting ANN can be extremely robust.

With the correct implementation ANNs can be used naturally in online learning and large dataset applications. Their simple implementation and the existence of mostly local dependencies exhibited in the structure allows for fast, parallel implementations in hardware.

Applications

The utility of artificial neural network models lies in the fact that they can be used to infer a function from observations. This is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical.

Real life applications

The tasks to which artificial neural networks are applied tend to fall within the following broad categories:

- Function approximation, or regression analysis, including time series prediction, fitness approximation and modeling.
- Classification, including pattern and sequence recognition, novelty detection and sequential decision making.
- Data processing, including filtering, clustering, blind source separation and compression.
- Robotics, including directing manipulators, Computer numerical control.

Application areas include system identification and control (vehicle control, process control), quantum chemistry,[2] game-playing and decision making (backgammon, chess, racing), pattern recognition (radar systems, face identification, object recognition and more), sequence recognition (gesture, speech, handwritten text recognition), medical diagnosis, financial applications (automated trading systems), data mining (or knowledge discovery in databases, "KDD"), visualization and e-mail spam filtering.

Neural network software

Neural network software is used to simulate, research, develop and apply artificial neural networks, biological neural networks and in some cases a wider array of adaptive systems.

Types of neural networks

Feedforward neural network

The feedforward neural network was the first and arguably simplest type of artificial neural network devised. In this network, the information moves in only one direction, forward, from the input nodes, through the hidden nodes (if any) and to the output nodes. There are no cycles or loops in the network.

Radial basis function (RBF) network

Radial Basis Functions are powerful techniques for interpolation in multidimensional space. A RBF is a function which has built into a distance criterion with respect to a center. Radial basis functions have been applied

in the area of neural networks where they may be used as a replacement for the sigmoidal hidden layer transfer characteristic in Multi-Layer Perceptrons. RBF networks have two layers of processing: In the first, input is mapped onto each RBF in the 'hidden' layer. The RBF chosen is usually a Gaussian. In regression problems the output layer is then a linear combination of hidden layer values representing mean predicted output. The interpretation of this output layer value is the same as a regression model in statistics. In classification problems the output layer is typically a sigmoid function of a linear combination of hidden layer values, representing a posterior probability.

Performance in both cases is often improved by shrinkage techniques, known as ridge regression in classical statistics and known to correspond to a prior belief in small parameter values (and therefore smooth output functions) in a Bayesian framework.

RBF networks have the advantage of not suffering from local minima in the same way as Multi-Layer Perceptrons. This is because the only parameters that are adjusted in the learning process are the linear mapping from hidden layer to output layer. Linearity ensures that the error surface is quadratic and therefore has a single easily found minimum. In regression problems this can be found in one matrix operation. In classification problems the fixed non-linearity introduced by the sigmoid output function is most efficiently dealt with using iteratively re-weighted least squares.

RBF networks have the disadvantage of requiring good coverage of the input space by radial basis functions. RBF centers are determined with reference to the distribution of the input data, but without reference to the prediction task. As a result, representational resources may be wasted on areas of the input space that are irrelevant to the learning task. A common solution is to associate each data point with its own centre, although this can make the linear system to be solved in the final layer rather large, and requires shrinkage techniques to avoid over fitting.

Associating each input datum with an RBF leads naturally to kernel methods such as Support Vector Machines and Gaussian Processes (the RBF is the kernel function). All three approaches use a non-linear kernel function to project the input data into a space where the learning problem can be solved using a linear model. Like Gaussian Processes, and unlike SVMs, RBF networks are typically trained in a Maximum Likelihood framework by maximizing the probability (minimizing the error) of the data under the model. SVMs take a different approach to avoiding over fitting by

maximizing instead a margin. RBF networks are outperformed in most classification applications by SVMs. In regression applications they can be competitive when the dimensionality of the input space is relatively small.

Kohonen self-organizing network

The self-organizing map (SOM) invented by Teuvo Kohonen performs a form of unsupervised learning. A set of artificial neurons learn to map points in an input space to coordinates in an output space. The input space can have different dimensions and topology from the output space, and the SOM will attempt to preserve these.

Recurrent network

Contrary to feedforward networks, recurrent neural networks (RNs) are models with bi-directional data flow. While a feedforward network propagates data linearly from input to output, RNs also propagate data from later processing stages to earlier stages.

Simple recurrent network

A *simple recurrent network* (SRN) is a variation on the Multi-Layer Perceptron, sometimes called an "Elman network" due to its invention by Jeff Elman. A three-layer network is used, with the addition of a set of "context units" in the input layer. There are connections from the middle (hidden) layer to these context units fixed with a weight of one. At each time step, the input is propagated in a standard feed-forward fashion, and then a learning rule (usually back-propagation) is applied. The fixed back connections result in the context units always maintaining a copy of the previous values of the hidden units (since they propagate over the connections before the learning rule is applied). Thus the network can maintain a sort of state, allowing it to perform such tasks as sequence-prediction that are beyond the power of a standard Multi-Layer Perceptron.

In a *fully recurrent network*, every neuron receives inputs from every other neuron in the network. These networks are not arranged in layers. Usually only a subset of the neurons receive external inputs in addition to the inputs from all the other neurons, and another disjunct subset of neurons report their output externally as well as sending it to all the neurons. These distinctive inputs and outputs perform the function of the input and output layers of a feed-forward or simple recurrent network, and also join all the other neurons in the recurrent processing.

Hopfield network

The Hopfield network is a recurrent neural network in which all connections are symmetric. Invented by John Hopfield in 1982, this network guarantees that its dynamics will converge. If the connections are trained using Hebbian learning then the Hopfield network can perform as robust content-addressable (or associative) memory, resistant to connection alteration.

Echo state network

The echo state network (ESN) is a recurrent neural network with a sparsely connected random hidden layer. The weights of output neurons are the only part of the network that can change and be learned. ESN are good to (re)produce temporal patterns.

Long short term memory network

The Long short term memory is an artificial neural net structure that unlike traditional RNNs doesn't have the problem of vanishing gradients. It can therefore use long delays and can handle signals that have a mix of low and high frequency components.

Stochastic neural networks

A stochastic neural network differs from a typical neural network because it introduces random variations into the network. In a probabilistic view of neural networks, such random variations can be viewed as a form of statistical sampling, such as Monte Carlo sampling.

Boltzmann machine

The Boltzmann machine can be thought of as a noisy Hopfield network. Invented by Geoff Hinton and Terry Sejnowski in 1985, the Boltzmann machine is important because it is one of the first neural networks to demonstrate learning of latent variables (hidden units). Boltzmann machine learning was at first slow to simulate, but the contrastive divergence algorithm of Geoff Hinton (circa 2000) allows models such as Boltzmann machines and *products of experts* to be trained much faster.

Modular neural networks

Biological studies have shown that the human brain functions not as a single massive network, but as a collection of small networks. This realization gave birth to the concept of modular neural networks, in which

several small networks cooperate or compete to solve problems.

Committee of machines

A committee of machines (CoM) is a collection of different neural networks that together "vote" on a given example. This generally gives a much better result compared to other neural network models. Because neural networks suffer from local minima, starting with the same architecture and training but using different initial random weights often gives vastly different networks. A CoM tends to stabilize the result.

The CoM is similar to the general machine learning *bagging* method, except that the necessary variety of machines in the committee is obtained by training from different random starting weights rather than training on different randomly selected subsets of the training data.

Associative neural network (ASNN)

The ASNN is an extension of the *committee of machines* that goes beyond a simple/weighted average of different models. ASNN represents a combination of an ensemble of feed-forward neural networks and the k-nearest neighbor technique (kNN). It uses the correlation between ensemble responses as a measure of **distance** amid the analyzed cases for the kNN. This corrects the bias of the neural network ensemble. An associative neural network has a memory that can coincide with the training set. If new data become available, the network instantly improves its predictive ability and provides data approximation (self-learn the data) without a need to retrain the ensemble.

Another important feature of ASNN is the possibility to interpret neural network results by analysis of correlations between data cases in the space of models. The method is demonstrated at www.vcclab.org, where you can either use it online or download it.

Physical neural network

A physical neural network includes electrically adjustable resistance material to simulate artificial synapses. Examples include the ADALINE neural network developed by Bernard Widrow in the 1960's and the memristor based neural network developed by Greg Snider of HP Labs in 2008.

Holographic associative memory

Holographic associative memory represents a family of analog, correlation-based, associative, stimulus-response memories, where information is mapped onto the phase orientation of complex numbers operating.

Instantaneously trained networks

Instantaneously trained neural networks (ITNNs) were inspired by the phenomenon of short-term learning that seems to occur instantaneously. In these networks the weights of the hidden and the output layers are mapped directly from the training vector data. Ordinarily, they work on binary data, but versions for continuous data that require small additional processing are also available.

Spiking neural networks

Spiking neural networks (SNNs) are models which explicitly take into account the timing of inputs. The network input and output are usually represented as series of spikes (delta function or more complex shapes). SNNs have an advantage of being able to process information in the time domain (signals that vary over time). They are often implemented as recurrent networks. SNNs are also a form of pulse computer.

Spiking neural networks with axonal conduction delays exhibit polychronization, and hence could have a very large memory capacity.

Networks of spiking neurons — and the temporal correlations of neural assemblies in such networks — have been used to model figure/ground separation and region linking in the visual system (see, for example, Reitboeck et al.in Haken and Stadler: Synergetics of the Brain. Berlin, 1989).

In June 2005 IBM announced construction of a Blue Gene supercomputer dedicated to the simulation of a large recurrent spiking neural network.

Gerstner and Kistler have a freely available online textbook on Spiking Neuron Models.

Dynamic neural networks

Dynamic neural networks not only deal with nonlinear multivariate behaviour, but also include (learning of) time-dependent behaviour such as various transient phenomena and delay effects.

Cascading neural networks

Cascade-Correlation is an architecture and supervised learning algorithm developed by Scott Fahlman and Christian Lebiere. Instead of just adjusting the weights in a network of fixed topology, Cascade-Correlation begins with a minimal network, then automatically trains and adds new hidden units one by one, creating a multi-layer structure. Once a new hidden unit has been added to the network, its input-side weights are frozen. This unit then becomes a permanent feature-detector in the network, available for producing outputs or for creating other, more complex feature detectors. The Cascade-Correlation architecture has several advantages over existing algorithms: it learns very quickly, the network determines its own size and topology, it retains the structures it has built even if the training set changes, and it requires no back-propagation of error signals through the connections of the network. See: Cascade correlation algorithm.

Neuro-fuzzy networks

A neuro-fuzzy network is a fuzzy inference system in the body of an artificial neural network. Depending on the *FIS* type, there are several layers that simulate the processes involved in a *fuzzy inference* like fuzzification, inference, aggregation and defuzzification. Embedding an *FIS* in a general structure of an *ANN* has the benefit of using available *ANN* training methods to find the parameters of a fuzzy system.

Compositional pattern-producing networks

Compositional pattern-producing networks (CPPNs) are a variation of ANNs which differ in their set of activation functions and how they are applied. While typical ANNs often contain only sigmoid functions (and sometimes Gaussian functions), CPPNs can include both types of functions and many others. Furthermore, unlike typical ANNs, CPPNs are applied across the entire space of possible inputs so that they can represent a complete image. Since they are compositions of functions, CPPNs in effect encode images at infinite resolution and can be sampled for a particular display at whatever resolution is optimal.

One-shot associative memory

This type of network can add new patterns without the need for re-training. It is done by creating a specific memory structure, which assigns each new pattern to an orthogonal plane using adjacently connected hierarchical

arrays. The network offers real-time pattern recognition and high scalability; it however requires parallel processing and is thus best suited for platforms such as Wireless sensor networks (WSN), Grid computing, and GPGPUs.

Theoretical properties

Computational power

The multi-layer perceptron (MLP) is a universal function approximator, as proven by the Cybenko theorem. However, the proof is not constructive regarding the number of neurons required or the settings of the weights.

Work by Hava Siegelmann and Eduardo D. Sontag has provided a proof that a specific recurrent architecture with rational valued weights (as opposed to the commonly used floating point approximations) has the full power of a Universal Turing Machine[6] using a finite number of neurons and standard linear connections. They have further shown that the use of irrational values for weights results in a machine with super-Turing power.

Capacity

Artificial neural network models have a property called 'capacity', which roughly corresponds to their ability to model any given function. It is related to the amount of information that can be stored in the network and to the notion of complexity.

Convergence

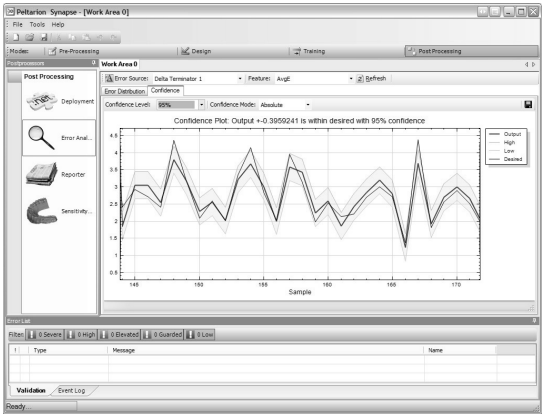
Nothing can be said in general about convergence since it depends on a number of factors. Firstly, there may exist many local minima. This depends on the cost function and the model. Secondly, the optimization method used might not be guaranteed to converge when far away from a local minimum. Thirdly, for a very large amount of data or parameters, some methods become impractical. In general, it has been found that theoretical guarantees regarding convergence are an unreliable guide to practical application.

Generalization and statistics

In applications where the goal is to create a system that generalizes well in unseen examples, the problem of overtraining has emerged.

Confidence analysis of a neural network

This arises in overcomplex or overspecified systems when the capacity of the network significantly exceeds the needed free parameters. There are two schools of thought for avoiding this problem: The first is to use cross-validation and similar techniques to check for the presence of overtraining and optimally select hyper-parameters such as to minimize the generalization error. The second is to use some form of *regularization*. This is a concept that emerges naturally in a probabilistic (Bayesian) framework, where the regularization can be performed by selecting a larger prior probability over simpler models; but also in statistical learning theory, where the goal is to minimize over two quantities: the 'empirical risk' and the 'structural risk', which roughly correspond to the error over the training set and the predicted error in unseen data due to over fitting.



Supervised neural networks that use an MSE cost function can use formal statistical methods to determine the confidence of the trained model. The MSE on a validation set can be used as an estimate for variance. This value can then be used to calculate the confidence interval of the output of the network, assuming a normal distribution. A confidence analysis made this way is statistically valid as long as the output probability distribution stays the same and the network is not modified.

By assigning a softmax activation function on the output layer of the neural network (or a softmax component in a component-based neural network) for categorical target variables, the outputs can be interpreted as posterior probabilities. This is very useful in classification as it gives a certainty measure on classifications.

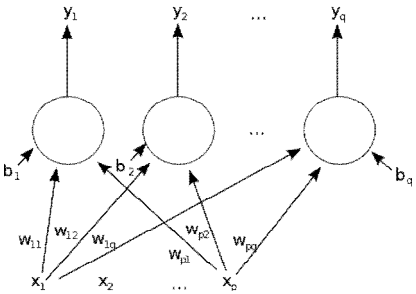
The softmax activation function $y_i = \frac{e^{x_i}}{\sum_{j=1}^{c} e^{x_j}}$ is:

Dynamic properties

Various techniques originally developed for studying disordered magnetic

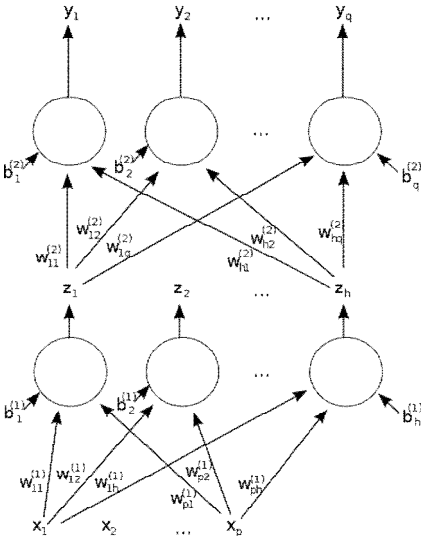
systems (i.e., the spin glass) have been successfully applied to simple neural network architectures, such as the Hopfield network. Influential work by E. Gardner and B. Derrida has revealed many interesting properties about perceptrons with real-valued synaptic weights, while later work by W. Krauth and M. Mezard has extended these principles to binary-valued synapses.

A single-layer feedforward artificial neural network.



Arrows originating from x_2 are omitted for clarity. There are p inputs to this network and q outputs. There is no activation function (or equivalently, the activation function is $g(x) = x$). In this system, the value of the qth output, y_q would be calculated as $y_q = \sum(x_i * w_{iq})$

A two-layer feedforward artificial neural network.



References

1. "The Machine Learning Dictionary". http://www.cse.unsw.edu.au/~billw/mldict.html#activnfn.
2. Roman M. Balabin, Ekaterina I. Lomakina (2009). "Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies". *J. Chem. Phys.* 131 (7): 074104. doi:10.1063/1.3206326.
3. Izhikevich EM (February 2006). "Polychronization: computation with spikes". *Neural Comput* 18 (2): 245–82. doi:10.1162/089976606775093882. PMID 16378515 16378515.
4. "IBM Research IBM and EPFL Join Forces to Uncover the Secrets of Cognitive Intelligence". http://domino.research.cbm.com/comm/pr.nsf/pages/news.20050606_CognitiveIntelligence.html.
5. B.B. Nasution, A.I. Khan, A Hierarchical Graph Neuron Scheme for Real-Time Pattern Recognition, IEEE Transactions on Neural Networks, vol 19(2), 212-229, Feb. 2008
6. Siegelmann, H.T.; Sontag, E.D. (1991). "Turing computability with neural nets". *Appl. Math. Lett.* **4** (6): 77–80. http://www.math.rutgers.edu/~sontag/FTP_DIR/aml-turing.pdf.
7. Performance comparison of neural network algorithms on UCI data sets: http://tunedit.org/results?d=UCI/&a=neural%20rbf%20perceptron
8. A close view to Artificial Neural Networks Algorithms: http://www.learnartificialneuralnetworks.com/
9. Neural Networks at the Open Directory Project: http://en.wikipedia.org/wiki/Open_Directory_Project
10. A Brief Introduction to Neural Networks (D. Kriesel) - Illustrated, bilingual manuscript about artificial neural networks; Topics so far: Perceptrons, Backpropagation, Radial Basis Functions, Recurrent Neural Networks, Self Organizing Maps, Hopfield Networks.
11. Dedicated issue of *Philosophical Transactions B* on Neural Networks and Perception. Some articles are freely available.
12. Bar-Yam, Yaneer (2003). *Dynamics of Complex Systems, Chapter 2, 3.*
13. Bar-Yam, Yaneer (2005). *Making Things Work.* Please see Chapter 3
14. Bhadeshia H. K. D. H. (1999). "Neural Networks in Materials Science". *ISIJ International* **39**: 966–979. doi:10.2355/isijinternational.39.966.
15. Bhagat, P.M. (2005) *Pattern Recognition in Industry*, Elsevier. ISBN 0-08-044538-1
16. Bishop, C.M. (1995) *Neural Networks for Pattern Recognition*, Oxford: Oxford University Press. ISBN 0-19-853849-9 (hardback) or ISBN 0-19-

853864-2 (paperback)
17. Cybenko, G.V. (1989). Approximation by Superpositions of a Sigmoidal function, *Mathematics of Control, Signals and Systems*, Vol. 2 pp. 303–314. electronic version
18. Duda, R.O., Hart, P.E., Stork, D.G. (2001) *Pattern classification (2nd edition)*, Wiley, ISBN 0-471-05669-3
19. Egmont-Petersen, M., de Ridder, D., Handels, H. (2002). "Image processing with neural networks - a review". *Pattern Recognition* **35** (10): 2279–2301. doi:10.1016/S0031-3203(01)00178-9+.
20. Gurney, K. (1997) *An Introduction to Neural Networks* London: Routledge. ISBN 1-85728-673-1 (hardback) or ISBN 1-85728-503-4 (paperback)
21. Haykin, S. (1999) *Neural Networks: A Comprehensive Foundation*, Prentice Hall, ISBN 0-13-273350-1
22. Fahlman, S, Lebiere, C (1991). *The Cascade-Correlation Learning Architecture*, created for National Science Foundation, Contract Number EET-8716324, and Defense Advanced Research Projects Agency (DOD), ARPA Order No. 4976 under Contract F33615-87-C-1499. electronic version
23. Hertz, J., Palmer, R.G., Krogh. A.S. (1990) *Introduction to the theory of neural computation*, Perseus Books. ISBN 0-201-51560-1
24. Lawrence, J. (1994) *Introduction to Neural Networks*, California Scientific Software Press. ISBN 1-883157-00-5
25. Masters, T. (1994) *Signal and Image Processing with Neural Networks*, J. Wiley & Sons, Inc. ISBN 0-471-04963-8
26. Ness, Erik. 2005. SPIDA-Web. *Conservation in Practice* 6(1):35-36. On the use of artificial neural networks in species taxonomy.
27. Ripley, Brian D. (1996) *Pattern Recognition and Neural Networks*, Cambridge
28. Siegelmann, H.T. and Sontag, E.D. (1994). Analog computation via neural networks, *Theoretical Computer Science*, v. 131, no. 2, pp. 331–360. electronic version
29. Sergios Theodoridis, Konstantinos Koutroumbas (2009) "Pattern Recognition" , 4th Edition, Academic Press, ISBN: 978-1-59749-272-0.
30. Smith, Murray (1993) *Neural Networks for Statistical Modeling*, Van Nostrand Reinhold, ISBN 0-442-01310-8
31. Wasserman, Philip (1993) *Advanced Methods in Neural Computing*, Van Nostrand Reinhold, ISBN 0-442-00461-3

17. Perceptron

The **perceptron** is a type of artificial neural network invented in 1957 at the Cornell Aeronautical Laboratory by Frank Rosenblatt. It can be seen as the simplest kind of feedforward neural network: a linear classifier.

Definition

The Perceptron is a binary classifier that maps its input *x* (a real-valued vector) to an output value *f*(*x*) (a single binary value) across the matrix.

$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{else} \end{cases}$$

where *w* is a vector of real-valued weights and $w \cdot x$ is the dot product (which computes a weighted sum). *b* is the 'bias', a constant term that does not depend on any input value.

The value of *f*(*x*) (0 or 1) is used to classify *x* as either a positive or a negative instance, in the case of a binary classification problem. The bias can be thought of as offsetting the activation function, or giving the output neuron a "base" level of activity. If *b* is negative, then the weighted combination of inputs must produce a positive value greater than | *b* | in order to push the classifier neuron over the 0 threshold. Spatially, the bias alters the position (though not the orientation) of the decision boundary.

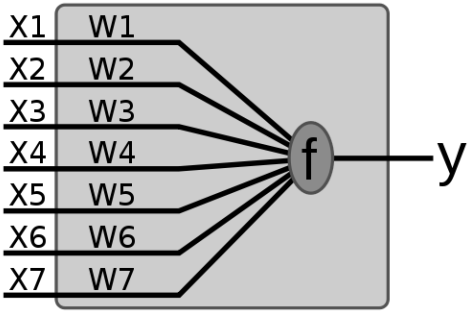
Since the inputs are fed directly to the output unit via the weighted connections, the perceptron can be considered the simplest kind of feed-forward neural network.

Learning algorithm

The learning algorithm is the same across all neurons, therefore everything that follows is applied to a single neuron in isolation. We first define some variables:

- *x*(*j*) denotes the j-th item in the n-dimensional input vector
- *w*(*j*) denotes the j-th item in the weight vector
- *f*(*x*) denotes the output from the neuron when presented with input *x*
- α is a constant where $0 < \alpha \leq 1$ (learning rate)

Further, assume for convenience that the bias term *b* is zero. This is not a restriction since an extra dimension *n* + 1 can be added to the input vectors x with *x*(*n* + 1) = 1, in which case *w*(*n* + 1) replaces the bias term.



The appropriate weights are applied to the inputs, and the resulting weighted sum passed to a function which produces the output y.

Learning is modeled as the weight vector being updated for multiple iterations over all training examples. Let $D_m = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ denote a training set of *m* training examples, where x_i is the input vector to the perceptron and y_i is the desired output value of the perceptron for that input vector.

Each iteration the weight vector is updated as follows:

For each (*x*,*y*) pair in $D_m = \{(x_1, y_1), \ldots, (x_m, y_m)\}$

$$w_{t+1}(j) = w_t(j) + \alpha(y - f(x))x(j) \quad (j = 1, \ldots, n)$$

Note that this means that a change in the weight vector will only take place for a given training example (*x*,*y*) if its output *f*(*x*) is different from the desired output *y*.

The initialization of *w* is usually performed simply by setting *w*(*j*): = 0 for all elements *w*(*j*).

Separability and Convergence

The training set D_m is said to be linearly separable if there exists a positive constant γ and a weight vector *w* such that $y_i \cdot (\langle w, x_i \rangle + b) > \gamma$ for all *i*. That is, if we say that *w* is the weight vector to the perceptron, then the output of the perceptron, $\langle w, x_i \rangle + b$, multiplied by the desired output of the perceptron, y_i, must be greater than the positive constant, γ, for all input-vector/output-value pairs (x_i,y_i) in D_m.

Novikoff (1962) proved that the perceptron algorithm converges after a finite number of iterations if the data set is linearly separable. The idea of the proof is that the weight vector is always adjusted by a bounded amount in a direction that it has a negative dot product with, and thus can be

bounded above by $O(\sqrt{t})$where t is the number of changes to the weight vector, but it can also be bounded below by $O(t)$ because if there exists an (unknown) satisfactory weight vector, then every change makes progress in this (unknown) direction by a positive amount that depends only on the input vector. This can be used to show that the number of mistakes (changes to the weight vector, i.e. t) is bounded by $(2R / \gamma)^2$ where R is the maximum norm of an input vector. However, if the training set is not linearly separable, the above online algorithm will not converge.

Note that the decision boundary of a perceptron is invariant with respect to scaling of the weight vector, i.e. a perceptron trained with initial weight vector w and learning rate α is an identical estimator to a perceptron trained with initial weight vector w / α and learning rate 1. Thus, since the initial weights become irrelevant with increasing number of iterations, the learning rate does not matter in the case of the perceptron and is usually just set to one.

Variants

The pocket algorithm with ratchet (Gallant, 1990) solves the stability problem of perceptron learning by keeping the best solution seen so far "in its pocket". The pocket algorithm then returns the solution in the pocket, rather than the last solution.

The α-perceptron further utilized a preprocessing layer of fixed random weights, with threshold output units. This enabled the perceptron to classify analogue patterns, by projecting them into a binary space. In fact, for a projection space of sufficiently high dimension, patterns can become linearly separable.

As an example, consider the case of having to classify data into two classes. Here is a small such data set, consisting of two points coming from two Gaussian distributions.

A linear classifier can only separate things with a hyperplane, so it's not possible to classify all the examples perfectly. On the other hand, we may project the data into a large number of dimensions. In this case a random matrix was used to project the data linearly to a 1000-dimensional space; then each resulting data point was transformed through the hyperbolic tangent function. A linear classifier can then separate the data, as shown in the third figure. However the data may still not be completely separable in this space, in which the perceptron algorithm would not converge. In the example shown, stochastic steepest gradient descent was used to adapt

the parameters.

Furthermore, by adding nonlinear layers between the input and output, one can separate all data and indeed, with enough training data, model any well-defined function to arbitrary precision. This model is a generalization known as a multilayer perceptron.

It should be kept in mind, however, that the best classifier is not necessarily that which classifies all the training data perfectly. Indeed, if we had the prior constraint that the data come from equi-variant Gaussian distributions, the linear separation in the input space is optimal.

Other training algorithms for linear classifiers are possible: see, e.g., support vector machine and logistic regression.

Multiclass perceptron

Like most other techniques for training linear classifiers, the perceptron generalizes naturally to multiclass classification. Here, the input *x* and the output *y* are drawn from arbitrary sets. A feature representation function *f*(*x*,*y*) maps each possible input/output pair to a finite-dimensional real-valued feature vector. As before, the feature vector is multiplied by a weight vector *w*, but now the resulting score is used to choose among many possible outputs:

$\hat{y} = \operatorname{argmax}_y f(x, y) \cdot w$

Learning again iterates over the examples, predicting an output for each, leaving the weights unchanged when the predicted output matches the target, and changing them when it does not. The update becomes:

$w_{t+1} = w_t + f(x, y) - f(x, \hat{y})$

This multiclass formulation reduces to the original perceptron when *x* is a real-valued vector, *y* is chosen from {0,1}, and *f*(*x*,*y*) = *yx*.

For certain problems, input/output representations and features can be chosen so that $\operatorname{argmax}_y f(x, y) \cdot w$ can be found efficiently even though *y* is chosen from a very large or even infinite set.

In recent years, perceptron training has become popular in the field of natural language processing for such tasks as part-of-speech tagging and syntactic parsing (Collins, 2002).

History

Although the perceptron initially seemed promising, it was eventually

proved that perceptrons could not be trained to recognize many classes of patterns. This led to the field of neural network research stagnating for many years, before it was recognized that a feedforward neural network with two or more layers (also called a multilayer perceptron) had far greater processing power than perceptrons with one layer (also called a single layer perceptron). Single layer perceptrons are only capable of learning linearly separable patterns; in 1969 a famous book entitled ***Perceptrons*** by Marvin Minsky and Seymour Papert showed that it was impossible for these classes of network to learn an XOR function. They conjectured (incorrectly) that a similar result would hold for a perceptron with three or more layers. Three years later Stephen Grossberg published a series of papers introducing networks capable of modeling differential, contrast-enhancing and XOR functions. (The papers were published in 1972 and 1973, see e.g.: Grossberg, Contour enhancement, short-term memory, and constancies in reverberating neural networks. Studies in Applied Mathematics, 52 (1973), 213-257, online [1]). Nevertheless the often-cited Minsky/Papert text caused a significant decline in interest and funding of neural network research. It took ten more years until neural network research experienced a resurgence in the 1980s. This text was reprinted in 1987 as "Perceptrons - Expanded Edition" where some errors in the original text are shown and corrected.

More recently, interest in the perceptron learning algorithm has increased again after Freund and Schapire (1998) presented a voted formulation of the original algorithm (attaining large margin) and suggested that one can apply the kernel trick to it.

References

1. Rosenblatt, Frank (1958), The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain, Cornell Aeronautical Laboratory, Psychological Review, v65, No. 6, pp. 386-408. doi:10.1037/h0042519.
2. Minsky M. L. and Papert S. A. 1969. *Perceptrons*. Cambridge, MA: MIT Press.
3. Freund, Y. and Schapire, R. E. 1998. Large margin classification using the perceptron algorithm. In Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT' 98). ACM Press.
4. Freund, Y. and Schapire, R. E. 1999. Large margin classification using the perceptron algorithm. In Machine Learning 37(3):277-296, 1999.
5. Gallant, S. I. (1990). Perceptron-based learning algorithms. IEEE

Transactions on Neural Networks, vol. 1, no. 2, pp. 179-191.
6. Novikoff, A. B. (1962). On convergence proofs on perceptrons. Symposium on the Mathematical Theory of Automata, 12, 615-622. Polytechnic Institute of Brooklyn.
7. Widrow, B., Lehr, M.A., "30 years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation," *Proc. IEEE*, vol 78, no 9, pp. 1415-1442, (1990).
8. Collins, M. 2002. Discriminative training methods for hidden Markov models: Theory and experiments with the perceptron algorithm in Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP '02)
9. Yin, Hongfeng (1996), Perceptron-Based Algorithms and Analysis, Spectrum Library, Concordia University, Canada
10. Sergei Alderman-ANN.rtf
11. Chapter 3 Weighted networks - the perceptron and chapter 4 Perceptron learning of *Neural Networks - A Systematic Introduction* by Raúl Rojas (ISBN 978-3540605058)
12. Pithy explanation of the update rule by Charles Elkan
13. C# implementation of a perceptron: http://dynamicnotions.blogspot.com/2008/09/single-layer-perceptron.html
14. Mathematics of perceptrons: http://users.ics.tkk.fi/ahonkela/dippa/node41.html

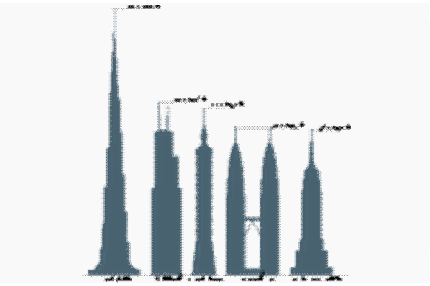
18. Cluster diagram

A **Cluster diagram** or *clustering diagram* is a general type of diagram, which represents some kind of cluster. A cluster in general is a group or bunch of several discrete items that are close to each other.

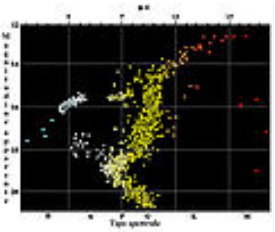
The cluster diagram figures a cluster, such as a network diagram figures a network, a flow diagram a process or movement of objects, and a tree diagram an abstract tree. But all these diagrams can be considered interconnected: A network diagram can be seen as a special orderly arranged kind of cluster diagram. A cluster diagram is a mesh kind of network diagram. A flow diagram can be seen as a line type of network diagram, and a tree diagram a tree type of network diagram.

Types of cluster diagrams

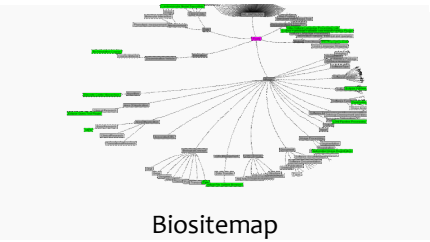
Specific types of cluster diagrams are:



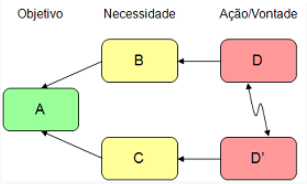
Comparison of sky scraper



Astronomic cluster of the Messier 3 globular cluster

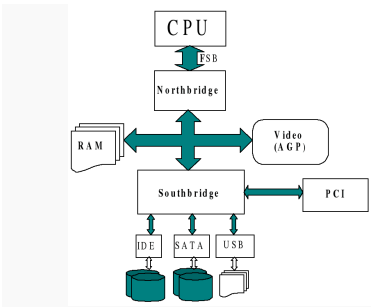


Biositemap

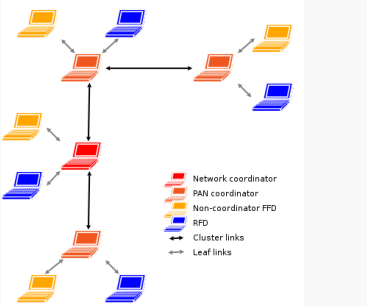


Cluster chart in brainstorming

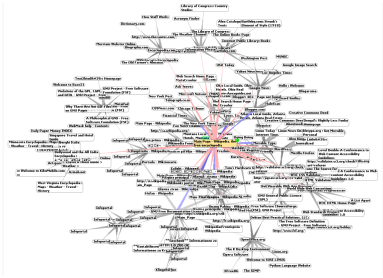
- In architecture a comparison diagram is sometimes called a cluster diagram.
- In astronomy diagrams of star cluster, galaxy groups and clusters or globular cluster.
- In brainstorming a cluster diagrams is also called *cloud diagram*. They can be considered "are types of non-linear graphic organizer that can help to systematize the generation of ideas based upon a central topic. Using this type of diagram... can more easily brainstorm a theme, associate about an idea, or explore a new subject". Also, the term cluster diagrams are sometimes used as synonym of mind maps".



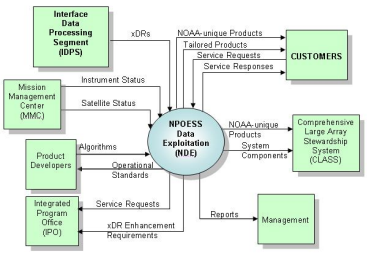
Computer architecture of a PC



Computer Network



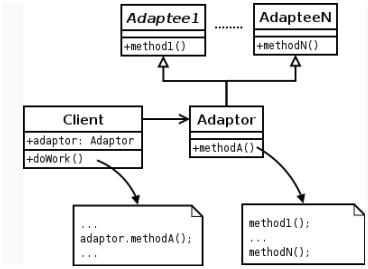
Internet



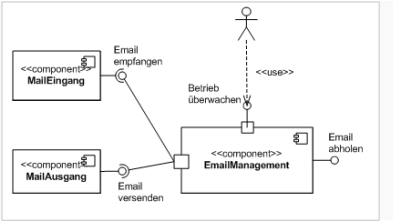
System context

- In computer science more complex diagrams of computer networks, computer architecture, file systems and internet can be considered cluster diagrams.

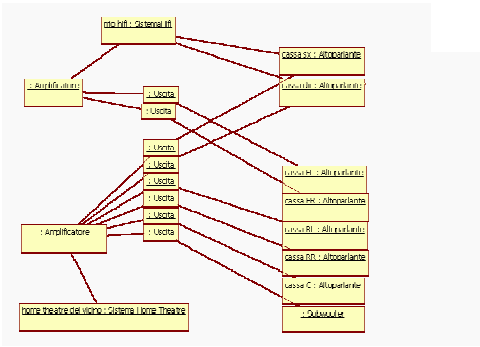
- In information visualization specific visual representation of large-scale collections of non-numerical information are sometimes drawn in the shape of a cluster diagram.
- In quantum field theory for example, according to Crawford (1998), the called *coupled cluster diagram* is a "simple diagrammatic formalism popularized by Kucharski and Bartlett [in 1986] by which one may construct the coupled cluster energy and amplitude equations far more quickly than by direct application of Wick's theorem".



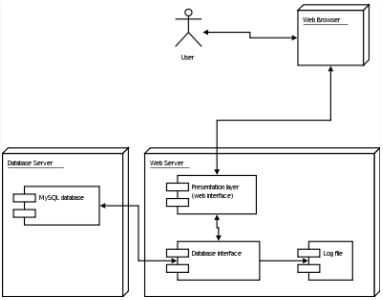
UML Class



UML Component



UML Composite structure



UML Deployment

- In the Unified Modeling Language (UML) all structure diagrams can be considered cluster diagrams. These structure diagrams emphasize what things must be in the system being modeled. UML encounters here the Class diagram, Component diagram, Composite structure diagram, Deployment diagram, Object diagram, and the Package diagram.

References

1. Lemma cluster at Wiktionary. Retrieved Sept 18, 2008.
2. Illustration called City of London Skyscraper Cluster Diagram at skyscrapernews.com. Retrieved 18 september 2008. Comment: This illustration depicts a "comparison diagram", but yet is called a "cluster diagram".
3. Cluster/Cloud Diagrams at enchantedlearning.com. 2003-2009. Accessed Nov 17, 2009.
4. Cluster diagrams are another way to mind map by starting with the keywords first, at www.brainstorming-that-works.com. Retrieved 18 September 2008.
5. T. Daniel Crawford (1998). "An Introduction to Coupled Cluster Diagrams". In: *Reviews in computational chemistry*. Kenny B. Lipkowitz, Donald B. Boyd eds. (2000) Vol 14. p.77. (Retrieved 18 september 2008).
6. Lee E. Brasseur (2003). *Visualizing technical information: a cultural critique*. Amityville, N.Y: Baywood Pub. ISBN 0-89503-240-6.
7. M. Dale and J. Moon (1988). "Statistical tests on two characteristics of the shapes of cluster diagrams". in: *Journal of Classification*, 1988, vol. 5, issue 1, pages 21–38.
8. Robert E. Horn (1999). *Visual Language: Global Communication for the 21st Century*. MacroVU Press.
9. Cluster diagram at cap.nsw.edu.au
10. City of London Skyscraper Cluster Diagram: http://www.skyscrapernews.com/diagram_city.php.

19. Scale-free network

A **scale-free network** is a network whose degree distribution follows a power law, at least asymptotically. That is, the fraction $P(k)$ of nodes in the network having k connections to other nodes goes for large values of k as $P(k) \sim k^{-\gamma}$ where γ is a constant whose value is typically in the range $2<\gamma<3$, although occasionally it may lie outside these bounds.

Scale-free networks are noteworthy because many empirically observed networks appear to be scale-free, including the protein networks, citation networks, and some social networks.

Highlights

- Scale-free networks show a power law degree distribution like many real networks.
- The mechanism of preferential attachment has been proposed as an underlying generative model to explain power law degree distributions in some networks.
- It has also been demonstrated that scale-free topologies in networks of *fixed* sizes can arise as a result of Dual Phase Evolution.

History

In studies of the networks of citations between scientific papers, Derek de Solla Price showed in 1965 that the number of links to papers—i.e., the number of citations they receive—had a heavy-tailed distribution following a Pareto distribution or power law, and thus that the citation network was scale-free. He did not however use the term "scale-free network" (which was not coined until some decades later). In a later paper in 1976, Price also proposed a mechanism to explain the occurrence of power laws in citation networks, which he called "cumulative advantage" but which is today more commonly known under the name preferential attachment.

Recent interest in scale-free networks started in 1999 with work by Albert-László Barabási and colleagues at the University of Notre Dame who mapped the topology of a portion of the Web (Barabási and Albert 1999), finding that some nodes, which they called "hubs", had many more connections than others and that the network as a whole had a power-law distribution of the number of links connecting to a node.

After finding that a few other networks, including some social and biological networks, also had heavy-tailed degree distributions, Barabási and collaborators coined the term "scale-free network" to describe the class of networks that exhibit a power-law degree distribution. Soon after, Amaral et al. showed that most of the real-world networks can be classified into two large categories according to the decay of P(k) for large k. Caroline S. Wagner (2008) demonstrated that scientific collaboration at the global level falls into scale free network structures along a power law form.

Barabási and Albert proposed a mechanism to explain the appearance of the power-law distribution, which they called "preferential attachment" and which is essentially the same as that proposed by Price. Analytic solutions for this mechanism (also similar to the solution of Price) were presented in 2000 by Dorogovtsev, Mendes and Samukhin and independently by Krapivsky, Redner, and Leyvraz, and later rigorously proved by mathematician Béla Bollobás. Notably, however, this mechanism only produces a specific subset of networks in the scale-free class, and many alternative mechanisms have been discovered since.

Although the scientific community is still debating the usefulness of the scale-free term in reference to networks, Li et al. (2005) recently offered a potentially more precise "scale-free metric". Briefly, let g be a graph with edge-set ε, and let the degree (number of edges) at a vertex i be d_i. Define

$$s(g) = \sum_{(i,j)\in\epsilon} d_i d_j.$$

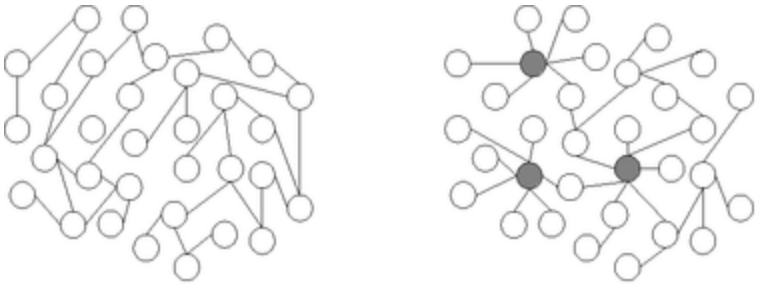
This is maximized when high-degree nodes are connected to other high-degree nodes. Now define

$$S(g) = \frac{s(g)}{s_{max}}$$

where s_{max} is the maximum value of $s(h)$ for h in the set of all graphs with an identical degree distribution to g. This gives a metric between 0 and 1, such that graphs with low $S(g)$ are "scale-rich", and graphs with $S(g)$ close to 1 are "scale-free". This definition captures the notion of self-similarity implied in the name "scale-free".

Characteristics and examples

Random network (a) and scale-free network (b). In the scale-free network, the larger hubs are highlighted.



(a) Random network (b) Scale-free network

As with all systems characterized by a power law distribution, the most notable characteristic in a scale-free network is the relative commonness of vertices with a degree that greatly exceeds the average. The highest-degree nodes are often called "hubs", and are thought to serve specific purposes in their networks, although this depends greatly on the domain.

The power law distribution highly influences the network topology. It turns out that the major hubs are closely followed by smaller ones. These, in turn, are followed by other nodes with an even smaller degree and so on. This hierarchy allows for fault tolerant behavior in the face of random failures: since the vast majority of nodes are those with small degree, the likelihood that a hub would be affected is almost negligible. Even if such event occurs, the network will not lose its connectedness, which is guaranteed by the remaining hubs. On the other hand, if a few major hubs are removed from the network, it simply falls apart and is turned into a set of rather isolated graphs. Thus hubs are both the strength of scale-free networks and their Achilles' heel.

Another important characteristic of scale-free networks is the clustering coefficient distribution, which decreases as the node degree increases. This distribution also follows a power law. That means that the low-degree nodes belong to very dense sub-graphs and those sub-graphs are connected to each other through hubs. Consider a social network in which nodes are people and links are acquaintance relationships between people. It is easy to see that people tend to form communities, i.e., small groups in which everyone knows everyone (one can think of such community as a complete graph). In addition, the members of a community also have a few acquaintance relationships to people outside that community. Some people, however, are so related to other people (e.g., celebrities,

politicians) that they are connected to a large number of communities. Those people may be considered the hubs responsible for making such networks small-world networks.

At present, the more specific characteristics of scale-free networks can only be discussed in either the context of the generative mechanism used to create them, or the context of a particular real-world network thought to be scale-free. For instance, networks generated by preferential attachment typically place the high-degree vertices in the middle of the network, connecting them together to form a core, with progressively lower-degree nodes making up the regions between the core and the periphery. Many interesting results are known for this subclass of scale-free networks. For instance, the random removal of even a large fraction of vertices impacts the overall connectedness of the network very little, while targeted attacks destroys the connectedness very quickly. Other scale-free networks, which place the high-degree vertices at the periphery, do not exhibit these properties; notably, the structure of the Internet is more like this latter kind of network than the kind built by preferential attachment. Indeed, many of the results about scale-free networks have been claimed to apply to the Internet, but are disputed by Internet researchers and engineers.

As with most disordered networks, such as the small world network model, the average distance between two vertices in the network is very small relative to a highly ordered network such as a lattice. The clustering coefficient of scale-free networks can vary significantly depending on other topological details, and there are now generative mechanisms that allow one to create such networks that have a high density of triangles.

It is interesting that Cohen and Havlin proved that uncorrelated power-law graphs having $2 < \gamma < 3$ will also have ultra small diameter $d \sim \ln \ln N$. So from the practical point of view, the diameter of a growing scale-free network might be considered almost constant.

Although many real-world networks are thought to be scale-free, the evidence remains inconclusive, primarily because the generative mechanisms proposed have not been rigorously validated against the real-world data. As such, it is too early to rule out alternative hypotheses. A few examples of networks claimed to be scale-free include:

- Some social networks, including collaboration networks. An example that has been studied extensively is the collaboration of movie actors in films.
- Protein-Protein interaction networks.

- Networks of sexual partners in humans, which affects the dispersal of sexually transmitted diseases.
- Many kinds of computer networks, including the World Wide Web.
- Semantic networks.

Generative models

These scale-free networks do not arise by chance alone. Erdős and Rényi (1960) studied a model of growth for graphs in which, at each step, two nodes are chosen uniformly at random and a link is inserted between them. The properties of these random graphs are not consistent with the properties observed in scale-free networks, and therefore a model for this growth process is needed.

The scale-free properties of the Web have been studied, and its distribution of links is very close to a power law, because there are a few Web sites with huge numbers of links, which benefit from a good placement in search engines and an established presence on the Web. Those sites are the ones that attract more of the new links. This has been called the winner takes all phenomenon.

The most widely known generative model for a subset of scale-free networks is Barabási and Albert's (1999) rich get richer generative model in which each new Web page creates links to existing Web pages with a probability distribution which is not uniform, but proportional to the current in-degree of Web pages. This model was originally discovered by Derek J. de Solla Price in 1965 under the term **cumulative advantage**, but did not reach popularity until Barabási rediscovered the results under its current name (BA Model). According to this process, a page with many in-links will attract more in-links than a regular page. This generates a power-law but the resulting graph differs from the actual Web graph in other properties such as the presence of small tightly connected communities. More general models and networks characteristics have been proposed and studied (for a review see the book by Dorogovtsev and Mendes).

A different generative model is the **copy** model studied by Kumar et al. (2000), in which new nodes choose an existent node at random and copy a fraction of the links of the existent node. This also generates a power law.

However, if we look at communities of interests in a specific topic, discarding the major hubs of the Web, the distribution of links is no longer a power law but resembles more a log-normal distribution, as observed by

Pennock et al. (2002) in the communities of the home pages of universities, public companies, newspapers and scientists. Based on these observations, they propose a generative model that mixes preferential attachment with a baseline probability of gaining a link.

The growth of the networks (adding new nodes) is not a necessary condition for creating a scale-free topology. For instance, it has been shown [2] that Dual Phase Evolution can produce scale-free topologies in networks of a fixed size. Dangalchev (2004) gives examples of generating static scale-free networks. Another possibility (Caldarelli et al. 2002) is to consider the structure as static and draw a link between vertices according to a particular property of the two vertices involved. Once specified the statistical distribution for these vertices properties (fitnesses), it turns out that in some circumstances also static networks develop scale-free properties.

Recently, Manev and Manev (Med. Hypotheses, 2005) proposed that small world networks may be operative in adult brain neurogenesis. Adult neurogenesis has been observed in mammalian brains, including those of humans, but a question remains: how do new neurons become functional in the adult brain? It is proposed that the random addition of only a few new neurons functions as a maintenance system for the brain's "small-world" networks. Randomly added to an orderly network, new links enhance signal propagation speed and synchronizability. Newly generated neurons are ideally suited to become such links: they are immature, form more new connections compared to mature ones, and their number but not their precise location may be maintained by continuous proliferation and dying off. Similarly, it is envisaged that the treatment of brain pathologies by cell transplantation would also create new random links in small-world networks and that even a small number of successfully incorporated new neurons may be functionally important.

References

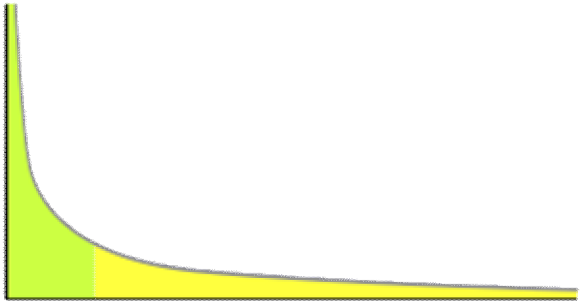
1. Albert R. and Barabási A.-L., "Statistical mechanics of complex networks", Rev. Mod. Phys. 74, 47–97 (2002).
2. Amaral, LAN, Scala, A., Barthelemy, M., Stanley, HE., "Classes of behavior of small-world networks". Proceedings of the National Academy of Sciences (USA) 97:11149-11152 (2000).

3. Barabási, Albert-László *Linked: How Everything is Connected to Everything Else*, 2004 ISBN 0-452-28439-2
4. Barabási, Albert-László, Bonabeau, Eric, "Scale-Free Networks". *Scientific American*, 288:50-59, May 2003.
5. Barabási, Albert-László and Albert, Réka. "Emergence of scaling in random networks". *Science*, 286:509-512, October 15, 1999.
6. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D-U., "Complex Networks: Structure and Dynamics", Physics Reports, 424:175-308, 2006.
7. Dan Braha and Yaneer Bar-Yam, "Topology of Large-Scale Engineering Problem-Solving Networks" Phys. Rev. E 69, 016113 (2004)
8. Caldarelli G. "Scale-Free Networks" Oxford University Press, Oxford (2007).
9. Caldarelli G., Capocci A., De Los Rios P., Muñoz M.A., "Scale-free networks from varying vertex intrinsic fitness" Physical Review Letters 89, 258702 (2002).
10. Dangalchev, Ch., *Generation models for scale-free networks*, Physica A, 338,(2004)
11. Dorogovtsev, S.N. and Mendes, J.F.F., *Evolution of Networks: from biological networks to the Internet and WWW*, Oxford University Press, 2003, ISBN 0-19-851590-1
12. Dorogovtsev, S.N. and Mendes, J.F.F. and Samukhin, A.N., "Structure of Growing Networks: Exact Solution of the Barabási--Albert's Model", Phys. Rev. Lett. 85, 4633 (2000)
13. Dorogovtsev, S.N. and Mendes, J.F.F., Evolution of networks, Advances in Physics 51, 1079-1187 (2002)
14. Erdős, P. and Rényi, A. *Random graphs.* Publication of the Mathematical Institute of the Hungarian Academy of Science, 5, pages 17–61, 1960.
15. Faloutsos, M., Faloutsos, P. and Faloutsos, C., *On power-law relationships of the internet topology* Comp. Comm. rev. 29, 251, 1999.
16. Li, L., Alderson, D., Tanaka, R., Doyle, J.C., Willinger, W., Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications (Extended Version). Internet Mathematics, 2005.
17. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., and Upfal, E.: Stochastic models for the web graph. In Proceedings of the 41st Annual Symposium on Foundations of Computer Science (FOCS), pages 57–65, Redondo Beach, CA, USA. IEEE CS Press, 2000.
18. Manev R., Manev H.; Med. Hypotheses 2005;64(1):114-7 [1]
19. Matlis, Jan. Scale-Free Networks. ComputerWorld. November 4, 2002.

20. Newman, Mark E. J. The structure and function of complex networks, 2003.
21. Pastor-Satorras, R.,Vespignani, A.,"Evolution and Structure of the Internet: A Statistical Physics Approach",Cambridge University Press, 2004, ISBN 0521826985
22. Pennock, D. M., Flake, G. W., Lawrence, S., Glover, E. J., and Giles, C. L.: Winners don't take all: Characterizing the competition for links on the web. Proceedings of the National Academy of Sciences, 99(8): 5207-5211 5207-5211, 2002.
23. Robb, John. Scale-Free Networks and Terrorism, 2004.
24. Keller, E.F. Revisiting "scale-free" networks, BioEssays 27(10) 1060-1068, 2005.
25. R. N. Onody and P. A. de Castro, *Complex Network Study of Brazilian Soccer Player*, Phys. Rev. E 70, 037103 (2004)
26. Reuven Cohen, Shlomo Havlin, "Scale-Free Networks are Ultrasmall" Phys. Rev. Lett. 90, 058701 (2003)
27. Wagner, Caroline S. The New Invisible College: Science for Development. Washington, D.C.: Brookings Press. (2008)
28. ^ Albert R. and Barabási A.-L., "Statistical mechanics of complex networks", Rev. Mod. Phys. 74, 47–97 (2002).
29. ^ Paperin, Greg; Green, David G.; Leishman, Tania G. (2008). "Dual Phase Evolution and Self-Organisation in Networks". *Proceedings of The Seventh International Conference on Simulated Evolution And Learning (SEAL'08)*. pp. 575–584. ISBN 978-3-540-89693-7. http://www.paperin.org/publications/2008/Dual_Phase_Evolution_and_Self-Organisation_in_Networks.
30. ^ Steyvers, Mark; Joshua B. Tenenbaum (2005). "The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth". *Cognitive Science* **29** (1): 41–78. doi:10.1207/s15516709cog2901_3. http://www.leaonline.com/doi/abs/10.1207/s15516709cog2901_3.

20. Power law

An example power law graph, being used to demonstrate ranking of popularity. To the right is the long tail, to the left are the few that dominate (also known as the 80-20 rule).



A **power law** is a special kind of mathematical relationship between two quantities. When the number or frequency of an object or event varies as a power of some attribute of that object (e.g., its size), the number or frequency is said to follow a power law.

For instance, the number of cities having a certain population size is found to vary as a power of the size of the population, and hence follows a power law.

Power laws govern a wide variety of natural and man-made phenomena, including frequencies of words in most languages, frequencies of family names, sizes of craters on the moon and of solar flares, the sizes of power outages, earthquakes, and wars, the popularity of books and music, and many other quantities.

Technical definition

A power law is any polynomial relationship that exhibits the property of **scale invariance.** The most common power laws relate two variables and have the form

$$f(x) = ax^k + o(x^k),$$

where *a* and *k* are constants, and $o(x^k)$ is an asymptotically small function of x^k. Here, *k* is typically called the *scaling exponent*, where the word "scaling" denotes the fact that a power-law function satisfies $f(cx) \propto f(x)$where *c* is a constant. Thus, a rescaling of the function's argument changes the constant of proportionality but preserves the shape of the function itself. This point becomes clearer if we take the logarithm of both sides:

$$\log\left(f(x)\right) = k \log x + \log a.$$

Notice that this expression has the form of a linear relationship with slope *k*. Rescaling the argument produces a linear shift of the function up or down but leaves both the basic form and the slope *k* unchanged.

Power-law relations characterize a staggering number of naturally occurring phenomena, and this is one of the principal reasons why they have attracted such wide interest. For instance, inverse-square laws, such as gravitation and the Coulomb force, are power laws, as are many common mathematical formulae such as the quadratic law of area of the circle. However much of the recent interest in power laws comes from the study of probability distributions: it's now known that the distributions of a wide variety of quantities seem to follow the power-law form, at least in their upper tail (large events).

The behavior of these large events connects these quantities to the study of theory of large deviations (also called extreme value theory), which considers the frequency of extremely rare events like stock market crashes and large natural disasters. It is primarily in the study of statistical distributions that the name "power law" is used; in other areas the power-law functional form is more often referred to simply as a polynomial form or polynomial function.

Scientific interest in power law relations stems partly from the ease with which certain general classes of mechanisms generate them. The demonstration of a power-law relation in some data can point to specific kinds of mechanisms that might underlie the natural phenomenon in question, and can indicate a deep connection with other, seemingly unrelated systems (see the reference by Simon and the subsection on universality below). The ubiquity of power-law relations in physics is partly due to dimensional constraints, while in complex systems, power laws are often thought to be signatures of hierarchy or of specific stochastic processes.

A few notable examples of power laws are the *Gutenberg-Richter law* for earthquake sizes, *Pareto's law* of income distribution, structural *self-*

similarity of fractals, and *scaling laws* in biological systems. Research on the origins of power-law relations, and efforts to observe and validate them in the real world, is an active topic of research in many fields of science, including physics, computer science, linguistics, geophysics, sociology, economics and more.

Properties of power laws

Scale invariance

The main property of power laws that makes them interesting is their scale invariance. Given a relation $f(x) = ax^k$, scaling the argument x by a constant factor causes only a proportionate scaling of the function itself. That is,

$$f(cx) = a(cx)^k = c^k f(x) \propto f(x).$$

That is, scaling by a constant simply multiplies the original power-law relation by the constant c^k. Thus, it follows that all power laws with a particular scaling exponent are equivalent up to constant factors, since each is simply a scaled version of the others. This behavior is what produces the linear relationship when both logarithms are taken of both $f(x)$ and x, and the straight-line on the log-log plot is often called the *signature* of a power law. Notably, however, with real data, such straightness is necessary, but not a sufficient condition for the data following a power-law relation. In fact, there are many ways to generate finite amounts of data that mimic this signature behavior, but, in their asymptotic limit, are not true power laws. Thus, accurately fitting and validating power-law models is an active area of research in statistics.

Universality

The equivalence of power laws with a particular scaling exponent can have a deeper origin in the dynamical processes that generate the power-law relation. In physics, for example, phase transitions in thermodynamic systems are associated with the emergence of power-law distributions of certain quantities, whose exponents are referred to as the critical exponents of the system. Diverse systems with the same critical exponents — that is, which display identical scaling behavior as they approach criticality — can be shown, via renormalization group theory, to share the same fundamental dynamics. For instance, the behavior of water and CO_2 at their boiling points fall in the same universality class because they have identical critical exponents. In fact, almost all material phase transitions are

described by a small set of universality classes. Similar observations have been made, though not as comprehensively, for various self-organized critical systems, where the critical point of the system is an attractor. Formally, this sharing of dynamics is referred to as universality, and systems with precisely the same critical exponents are said to belong to the same universality class.

Power-law functions

The general power-law function follows the polynomial form given above, and is a ubiquitous form throughout mathematics and science. Notably, however, not all polynomial functions are power laws because not all polynomials exhibit the property of scale invariance. Typically, power-law functions are polynomials in a single variable, and are explicitly used to model the scaling behavior of natural processes. For instance, allometric scaling laws for the relation of biological variables are some of the best known power-law functions in nature. In this context, the $o(x^k)$ term is most typically replaced by a deviation term ε, which can represent uncertainty in the observed values (perhaps measurement or sampling errors) or provide a simple way for observations to deviate from the no power-law function (perhaps for stochastic reasons):

$$y = ax^k + \varepsilon.$$

Examples of power law functions

- The Stevens' power law of psychophysics
- The Stefan–Boltzmann law
- The Ramberg-Osgood stress-strain relationship
- The Inverse-square laws of Newtonian gravity and Electrostatics
- Electrostatic potential and Gravitational potential
- Model of van der Waals force
- Force and potential in Simple harmonic motion
- Kepler's third law
- The Initial mass function
- Gamma correction relating light intensity with voltage
- Kleiber's law relating animal metabolism to size, and allometric laws in general
- Behaviour near second-order phase transitions involving critical exponents
- Proposed form of experience curve effects
- The differential energy spectrum of cosmic-ray nuclei

- Square-cube law (ratio of surface area to volume)
- Constructal law
- Fractals
- The Pareto principle also called the "80-20 rule"
- Zipf's Law in corpus analysis and population distributions amongst others, where frequency of an item or event is inversely proportional to its frequency rank (i.e. the second most frequent item/event occurring half as often the most frequent item and so on).
- Weight vs. length models in fish

Power-law distributions

A power-law distribution is any that, in the most general sense, has the form

$$p(x) \propto L(x)x^{-\alpha}$$

where $\alpha > 1$, and $L(x)$ is a **slowly varying function**, which is any function that satisfies $\lim_{x\to\infty} L(tx)/L(x) = 1$ with t constant. This property of $L(x)$ follows directly from the requirement that $p(x)$ be asymptotically scale invariant; thus, the form of $L(x)$ only controls the shape and finite extent of the lower tail. For instance, if $L(x)$ is the constant function, then we have a power-law that holds for all values of x. In many cases, it is convenient to assume a lower bound x_{min} from which the law holds. Combining these two cases, and where x is a continuous variable, the power law has the form

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha},$$

where the pre-factor to $x^{-\alpha}$ is the normalizing constant. We can now consider several properties of this distribution. For instance, its moments are given by

$$\langle x^m \rangle = \int_{x_{\min}}^{\infty} x^m p(x)\,\mathrm{d}x = \frac{\alpha - 1}{\alpha - 1 - m} x_{\min}^m$$

which is only well defined for $m < \alpha - 1$. That is, all moments $m \geq \alpha - 1$ diverge: when $\alpha < 2$, the average and all higher-order moments are infinite; when $2 < \alpha < 3$, the mean exists, but the variance and higher-order moments are infinite, etc. For finite-size samples drawn from such distribution, this behavior implies that the central moment estimators (like the mean and the variance) for diverging moments will never converge - as more data is accumulated, they continue to grow.

Another kind of power-law distribution, which does not satisfy the general form above, is the power law with an exponential cutoff

$$p(x) \propto L(x)x^{-\alpha}\mathrm{e}^{-\lambda x}.$$

In this distribution, the exponential decay term $e^{-\lambda x}$ eventually overwhelms the power-law behavior at very large values of *x*. This distribution does not scale and is thus not asymptotically a power law; however, it does approximately scale over a finite region before the cutoff. (Note that the pure form above is a subset of this family, with λ = 0.) This distribution is a common alternative to the asymptotic power-law distribution because it naturally captures finite-size effects. For instance, although the Gutenberg–Richter law is commonly cited as an example of a power-law distribution, the distribution of earthquake magnitudes cannot scale as a power law in the limit $x \longrightarrow \infty$because there is a finite amount of energy in the Earth's crust and thus there must be some maximum size to an earthquake. As the scaling behavior approaches this size, it must taper off.

Plotting power-law distributions

In general, power-law distributions are plotted on double logarithmic axes, which emphasizes the upper tail region. The most convenient way to do this is via the (complementary) cumulative distribution, $P(x) = \Pr(X > x)$,

$$P(x) = \Pr(X > x) = C\int_x^\infty p(X)\,\mathrm{d}X = \frac{\alpha-1}{x_{\min}^{-\alpha+1}}\int_x^\infty X^{-\alpha}\,\mathrm{d}X = \left(\frac{x}{x_{\min}}\right)^{-\alpha+1}.$$

Note that the cumulative distribution (cdf) is also a power-law function, but with a smaller scaling exponent. For data, an equivalent form of the cdf is the rank-frequency approach, in which we first sort the *n* observed values in ascending order, and plot them against the vector

$$\left[1, \frac{n-1}{n}, \frac{n-2}{n}, \ldots, \frac{1}{n}\right].$$

Although it can be convenient to log-bin the data, or otherwise smooth the probability density (mass) function directly, these methods introduce an implicit bias in the representation of the data, and thus should be avoided. The cdf, on the other hand, introduces no bias in the data and preserves the linear signature on doubly logarithmic axes.

Estimating the exponent from empirical data

There are many ways of estimating the value of the scaling exponent for a power-law tail, however not all of them yield unbiased and consistent

answers. The most reliable techniques are often based on the method of maximum likelihood. Alternative methods are often based on making a linear regression on either the log-log probability, the log-log cumulative distribution function, or on log-binned data, but these approaches should be avoided as they can all lead to highly biased estimates of the scaling exponent.

For real-valued data, we fit a power-law distribution of the form

$$p(x) = \frac{\alpha - 1}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-\alpha}$$

to the data $x \geq x_{\min}$. Given a choice for x_{min}, a simple derivation by this method yields the estimator equation

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}} \right]^{-1}$$

where $\{x_i\}$ are the n data points $x_i \geq x_{\min}$. (For a more detailed derivation, see Hall or Newman below.) This estimator exhibits a small finite sample-size bias of order $O(n^{-1})$, which is small when $n > 100$. Further, the uncertainty in the estimation can be derived from the maximum likelihood argument, and has the form $\sigma = \frac{\alpha - 1}{\sqrt{n}}$. This estimator is equivalent to the popular Hill estimator from quantitative finance and extreme value theory.

For a set of n integer-valued data points $\{x_i\}$, again where each $x_i \geq x_{\min}$, the maximum likelihood exponent is the solution to the transcendental equation

$$\frac{\zeta'(\hat{\alpha}, x_{\min})}{\zeta(\hat{\alpha}, x_{\min})} = -\frac{1}{n} \sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}$$

where $\zeta(\alpha, x_{min})$ is the incomplete zeta function. The uncertainty in this estimate follows the same formula as for the continuous equation. However, the two equations for $\hat{\alpha}$ are not equivalent, and the continuous version should not be applied to discrete data, nor vice versa.

Further, both of these estimators require the choice of x_{min}. For functions with a non-trivial $L(x)$ function, choosing x_{min} too small produces a significant bias in $\hat{\alpha}$, while choosing it too large increases the uncertainty in $\hat{\alpha}$, and reduces the statistical power of our model. In general, the best choice of x_{min} depends strongly on the particular form of the lower tail, represented by $L(x)$ above.

More about these methods, and the conditions under which they can be used, can be found in the Clauset et al. reference below. Further, this comprehensive review article provides usable code (Matlab and R) for estimation and testing routines for power-law distributions.

Examples of power-law distributions

- Pareto distribution (continuous)
- Zeta distribution (discrete)
- Yule–Simon distribution (discrete)
- Student's t-distribution (continuous), of which the Cauchy distribution is a special case
- Zipf's law and its generalization, the Zipf-Mandelbrot law (discrete)
- Lotka's law
- The scale-free network model
- Bibliograms
- Gutenberg–Richter law of earthquake magnitudes
- Horton's laws describing river systems
- Richardson's Law for the severity of violent conflicts (wars and terrorism)
- population of cities
- numbers of religious adherents
- net worth of individuals
- frequency of words in a text
- Pink noise
- 90-9-1 principle on wikis

A great many power-law distributions have been conjectured in recent years. For instance, power laws are thought to characterize the behavior of the upper tails for the popularity of websites, number of species per genus, the popularity of given names, the size of financial returns, and many others. However, much debate remains as to which of these tails are actually power-law distributed and which are not. For instance, it is commonly accepted now that the famous Gutenberg–Richter law decays more rapidly than a pure power-law tail because of a finite exponential cutoff in the upper tail.

Validating power laws

Although power-law relations are attractive for many theoretical reasons, demonstrating that data do indeed follow a power-law relation requires more than simply fitting such a model to the data. In general, many

alternative functional forms can appear to follow a power-law form for some extent. Thus, the preferred method for validation of power-law relations is by testing many orthogonal predictions of a particular generative mechanism against data, and not simply fitting a power-law relation to a particular kind of data. As such, the validation of power-law claims remains a very active field of research in many areas of modern science.

References

1. Simon, H. A. (1955). "On a Class of Skew Distribution Functions". *Biometrika* **42**: 425–440. doi:10.2307/2333389. http://links.jstor.org/sici?sici=00063444%28195512%2942%3A3%2F4%3C425%3AOACOSD%3E2.0.CO%3B2-M.
2. Hall, P. (1982). "On Some Simple Estimates of an Exponent of Regular Variation". *J. of the Royal Statistical Society, Series B (Meth)* 44 (1): 37–42. http://links.jstor.org/sici?sici=00359246(1982)44%3A1%3C37%3AOSSEOA%3E2.0.CO%3B2-4.
3. Mitzenmacher, M. (2003). "A brief history of generative models for power law and lognormal distributions". *Internet Mathematics* **1**: 226–251. http://www.internetmathematics.org/volumes/1/2/pp226_251.pdf.
4. Newman, M. E. J. (2005). "Power laws, Pareto distributions and Zipf's law". *Contemporary Physics* 46: 323–351. doi:10.1080/00107510500052444. http://www.journalsonline.tandf.co.uk/openurl.asp?genre=article&doi=10.1080/00107510500052444.
5. Clauset, A., Shalizi, C. R. and Newman, M. E. J. (2009). "Power-law distributions in empirical data". *SIAM Review* **51**: 661–703. doi:10.1137/070710111. http://arxiv.org/abs/0706.1062.
6. *Ubiquity* Mark Buchanan(2000) Wiedenfield & Nicholson ISBN 0297 643762
7. Zipf's law
8. Power laws, Pareto distributions and Zipf's law
9. Zipf, Power-laws, and Pareto - a ranking tutorial
10. Gutenberg-Richter Law
11. Stream Morphometry and Horton's Laws

12. Clay Shirky on Institutions & Collaboration: Power law in relation to the internet-based social networks
13. Clay Shirky on Power Laws, Weblogs, and Inequality
14. "How the Finance Gurus Get Risk All Wrong" by Benoit Mandelbrot & Nassim Nicholas Taleb. *Fortune*, July 11, 2005.
15. "Million-dollar Murray": power-law distributions in homelessness and other social problems; by Malcolm Gladwell. *The New Yorker*, February 13, 2006.
16. Benoit Mandelbrot & Richard Hudson: The Misbehaviour of Markets (2004)
17. Philip Ball: Critical Mass: How one thing leads to another (2005)
18. *Tyranny of the Power Law* from The Econophysics Blog
19. *So You Think You Have a Power Law — Well Isn't That Special?* from Three-Toed Sloth, the blog of Cosma Shalizi, Professor of Statistics at Carnegie-Mellon University.

21. Pareto principle

The **Pareto principle** (also known as the **80-20 rule**, the **law of the vital few,** and the **principle of factor sparsity**) states that, for many events, roughly 80% of the effects come from 20% of the causes. Business management thinker Joseph M. Juran suggested the principle and named it after Italian economist Vilfredo Pareto, who observed in 1906 that 80% of the land in Italy was owned by 20% of the population; he developed the principle by observing that 20% of the pea pods in his garden contained 80% of the peas. It is a common rule of thumb in business; e.g., "80% of your sales come from 20% of your clients." Mathematically, where something is shared among a sufficiently large set of participants, there must be a number *k* between 50 and 100 such that *k*% is taken by (100 – *k*)% of the participants. *k* may vary from 50 (in the case of equal distribution) to nearly 100 (when a tiny number of participants account for almost all of the resource). There is nothing special about the number 80% mathematically, but many real systems have *k* somewhere around this region of intermediate imbalance in distribution.

The Pareto principle is only tangentially related to Pareto efficiency, which was also introduced by the same economist. Pareto developed both concepts in the context of the distribution of income and wealth among the population.

In economics

The original observation was in connection with income and wealth. Pareto noticed that 80% of Italy's wealth was owned by 20% of the population. He then carried out surveys on a variety of other countries and found to his surprise that a similar distribution applied.

Because of the scale-invariant nature of the power law relationship, the relationship applies also to subsets of the income range. Even if we take the ten wealthiest individuals in the world, we see that the top three (Warren Buffett, Carlos Slim Helú, and Bill Gates) own as much as the next seven put together.

A chart that gave the inequality a very visible and comprehensible form, the so-called 'champagne glass' effect, was contained in the 1992 United

Nations Development Program Report, which showed the distribution of global income to be very uneven, with the richest 20% of the world's population controlling 82.7% of the world's income.

Distribution of world GDP in 1989

Quintile of population	Income
Richest 20%	82.70%
Second 20%	11.75%
Third 20%	2.30%
Fourth 20%	1.85%
Poorest 20%	1.40%

The Pareto Principle has also been used to attribute the widening economic inequality in the USA to 'skill-biased technical change' – i.e. income growth accrues to those with the education and skills required to take advantage of new technology and globalization. However, Nobel Prize winner in Economics Paul Krugman in the *New York Times* dismissed this "80-20 fallacy" as being cited "not because it's true, but because it's comforting." He asserts that the benefits of economic growth over the last 30 years have largely been concentrated in the top 1%, rather than the top 20%.

In software

In computer science and engineering control theory such as for electromechanical energy converters, the Pareto principle can be applied to optimization efforts.

Microsoft also noted that by fixing the top 20% of the most reported bugs, 80% of the errors and crashes would be eliminated.

In computer graphics the Pareto principle is used for ray-tracing: 80% of rays intersect 20% of geometry.

Other applications

In the systems science discipline, Epstein and Axtell created an agent-based simulation model called Sugarscape, from a decentralized modeling approach, based on individual behavior rules defined for each agent in the economy. Wealth distribution and Pareto's 80/20 Principle became

emergent in their results, which suggests that the principle is a natural phenomenon.

The Pareto Principle also applies to a variety of more mundane matters: one might guess approximately that we wear our 20% most favored clothes about 80% of the time, perhaps we spend 80% of the time with 20% of our acquaintances, etc.

The Pareto principle has many applications in quality control. It is the basis for the Pareto chart, one of the key tools used in total quality control and six sigma. The Pareto principle serves as a baseline for ABC-analysis and XYZ-analysis, widely used in logistics and procurement for the purpose of optimizing stock of goods, as well as costs of keeping and replenishing that stock.

The Pareto principle was a prominent part of the 2007 bestseller *The 4-Hour Workweek* by Tim Ferriss. Ferriss recommended focusing one's attention on those 20% that contribute to 80% of the income. More notably, he also recommends firing those 20% of customers who take up the majority of one's time and cause the most trouble.

In human developmental biology the principle is reflected in the gestation period where the embryonic period constitutes 20% of the whole, with the foetal development taking up the rest of the time.

In health care in the United States, it has been found that 20% of patients use 80% of health care resources.

Mathematical notes

The idea has rule-of-thumb application in many places, but it is commonly misused. For example, it is a misuse to state that a solution to a problem "fits the 80-20 rule" just because it fits 80% of the cases; it must be implied that this solution requires only 20% of the resources needed to solve all cases. Additionally, it is a misuse of the 80-20 rule to interpret data with a small number of categories or observations.

Mathematically, where something is shared among a sufficiently large set of participants, there will always be a number k between 50 and 100 such that k% is taken by (100 – k)% of the participants; however, k may vary from 50 in the case of equal distribution (e.g. exactly 50% of the people take 50% of the resources) to nearly 100 in the case of a tiny number of participants taking almost all of the resources. There is nothing special about the number 80,

but many systems will have *k* somewhere around this region of intermediate imbalance in distribution.

This is a special case of the wider phenomenon of Pareto distributions. If the parameters in the Pareto distribution are suitably chosen, then one would have not only 80% of effects coming from 20% of causes, but also 80% of that top 80% of effects coming from 20% of that top 20% of causes, and so on (80% of 80% is 64%; 20% of 20% is 4%, so this implies a "64-4" law; and a similarly implies a "51.2-0.8" law).

80-20 is only shorthand for the general principle at work. In individual cases, the distribution could just as well be, say, 80-10 or 80-30. (There is no need for the two numbers to add up to 100%, as they are measures of different things, e.g., 'number of customers' vs 'amount spent'). The classic 80-20 distribution occurs when the gradient of the line is –1 when plotted on log-log axes of equal scaling. Pareto rules are not mutually exclusive. Indeed, the 0-0 and 100-100 rules always hold. Adding up to 100 leads to a nice symmetry. For example, if 80% of effects come from the top 20% of sources, then the remaining 20% of effects come from the lower 80% of sources. This is called the "joint ratio", and can be used to measure the degree of imbalance: a joint ratio of 96:4 is very imbalanced, 80:20 is significantly imbalanced (Gini index: 60%), 70:30 is moderately imbalanced (Gini index: 40%), and 55:45 is just slightly imbalanced.

The Pareto Principle is an illustration of a "power law" relationship, which also occurs in phenomena such as brush-fires and earthquakes because it is self-similar over a wide range of magnitudes; it produces outcomes completely different from Gaussian distribution phenomena. This fact explains the frequent breakdowns of sophisticated financial instruments, which are modeled on the assumption that a Gaussian relationship is appropriate to - for example - stock movement sizes.

Equality measures

Gini coefficient and Hoover index

Using the "*A* : *B*" notation (for example, 0.8:0.2) and with *A* + *B* = 1, inequality measures like the Gini index *and* the Hoover index can be computed. In this case both are the same.

$$H = G = |2A - 1| = |2B - 1|$$
$$A : B = \left(\tfrac{1+H}{2}\right) : \left(\tfrac{1-H}{2}\right)$$

Theil index

The Theil index is an entropy measure used to quantify inequities. The measure is 0 for 50:50 distributions and reaches 1 at a Pareto distribution of 82:18. Higher inequities yield Theil indices above 1.

$$T_T = T_L = T_s = 2H \operatorname{arctanh}(H) .$$

References

1. Bookstein, Abraham (1990). "Informetric distributions, part I: Unified overview". *Journal of the American Society for Information Science* **41**: 368–375. doi:10.1002/(SICI)1097-4571(199007)41:5<368::AID-ASI8>3.0.CO;2-C.
2. Klass, O. S.; Biham, O.; Levy, M.; Malcai, O.; Soloman, S. (2006). "The Forbes 400 and the Pareto wealth distribution". *Economics Letters* **90** (2): 290–295. doi:10.1016/j.econlet.2005.08.020.
3. Koch, R. (2001). *The 80/20 Principle: The Secret of Achieving More with Less*. London: Nicholas Brealey Publishing. http://www.scribd.com/doc/3664882/ The-8020-Principle-The-Secret-to-Success-by-Achieving-More-with-Less.
4. Koch, R. (2004). *Living the 80/20 Way: Work Less, Worry Less, Succeed More, Enjoy More*. London: Nicholas Brealey Publishing. ISBN 1857883314.
5. Reed, W. J. (2001). "The Pareto, Zipf and other power laws". *Economics Letters* **74** (1): 15–19. doi:10.1016/S0165-1765(01)00524-9.
6. Rosen, K. T.; Resnick, M. (1980). "The size distribution of cities: an examination of the Pareto law and primacy". *Journal of Urban Economics* **8**: 165–186.
7. Rushton, A.; Oxley, J.; Croucher, P. (2000), *The handbook of logistics and distribution management*, London: Kogan Page, ISBN 978-0749433659.
8. The Pareto principle has several name variations, including: Pareto's Law, the 80/20 rule, the 80:20 rule, and 80 20 rule.
9. Bunkley, Nick (March 3, 2008), "Joseph Juran, 103, Pioneer in Quality Control, Dies", *New York Times*, http://www.nytimes.com/2008/03/03/business/03juran.html.
10. *What is 80/20 Rule, Pareto's Law, Pareto Principle* http://www.80-20presentationrule.com/whatisrule.html.
11. Pareto, Vilfredo; Page, Alfred N. (1971), *Translation of Manuale di economia politica ("Manual of political economy")*, A.M. Kelley, ISBN 9780678008812.

12. *The Forbes top 100 billionaire rich-list*, This is Money, http://www.thisismoney.co.uk/news/article.html?in_article_id=418243&in_page_id=3.
13. Gorostiaga, Xabier (January 27, 1995), "World has become a 'champagne glass' globalization will fill it fuller for a wealthy few", *National Catholic Reporter*.
14. United Nations Development Program (1992), *1992 Human Development Report*, New York: Oxford University Press.
15. *Human Development Report 1992, Chapter 3*, http://hdr.undp.org/en/reports/global/hdr1992/chapters/, retrieved 2007-07-08.
16. Krugman, Paul (February 27, 2006). "Graduates versus Oligarchs". *New York Times*: pp. A19. http://select.nytimes.com/2006/02/27/opinion/27krugman.html?_r=1.
17. Gen, M.; Cheng, R. (2002), *Genetic Algorithms and Engineering Optimization*, New York: Wiley.
18. Rooney, Paula (October 3, 2002), *Microsoft's CEO: 80-20 Rule Applies To Bugs, Not Just Features*, ChannelWeb, http://www.crn.com/security/18821726.
19. Slusallek, Philipp, *Ray Tracing Dynamic Scenes*, http://graphics.cs.uni-sb.de/new/fileadmin/cguds/courses/ws0809/cg/slides/CG04-RT-III.pdf.
20. Epstein, Joshua; Axtell, Robert (1996). *Growing Artificial Societies: Social Science from the Bottom-Up*. MIT Press. pp. 208. ISBN 0-262-55025-3. http://books.google.com/books?id=xXvelSs2caQC.
21. Rushton, Oxley & Croucher (2000), pp. 107–108.
22. Ferris, Tim (2007), *The 4-Hour Workweek*, Crown Publishing.
23. http://www.projo.com/opinion/contributors/content/CT_weinberg27_07-27-09_HQF0P1E_v15.3f89889.html
24. Taleb, Nassim (2007). *The Black Swan*. pp. 228–252, 274–285.
25. On Line Calculator: Inequality

22. Natural monopoly

In economics, a **natural monopoly** occurs when, due to the economies of scale of a particular industry, the maximum efficiency of production and distribution is realized through a single supplier, but in some cases inefficiency may take place.

Natural monopolies arise where the largest supplier in an industry, often the first supplier in a market, has an overwhelming cost advantage over other actual or potential competitors. This tends to be the case in industries where capital costs predominate, creating economies of scale which are large in relation to the size of the market, and hence high barriers to entry; examples include public utilities such as water services and electricity.

It is very expensive to build transmission networks (water/gas pipelines, electricity and telephone lines), therefore it is unlikely that a potential competitor would be willing to make the capital investment needed to even enter the monopolist's market.

It may also depend on control of a particular natural resource. Companies that grow to take advantage of economies of scale often run into problems of bureaucracy; these factors interact to produce an "ideal" size for a company, at which the company's average cost of production is minimized. If that ideal size is large enough to supply the whole market, then that market is a natural monopoly.

Some free-market-oriented economists argue that natural monopolies exist only in theory, and not in practice, or that they exist only as transient states.

Explanation

All industries have costs associated with entering them. Often, a large portion of these costs is required for investment. Larger industries, like utilities, require enormous initial investment. This barrier to entry reduces the number of possible entrants into the industry regardless of the earning of the corporations within.

Natural monopolies arise where the largest supplier in an industry, often the first supplier in a market, has an overwhelming cost advantage over other actual or potential competitors; this tends to be the case in industries where fixed costs predominate, creating economies of scale which are large

in relation to the size of the market - examples include water services and electricity. It is very expensive to build transmission networks (water/gas pipelines, electricity and telephone lines), therefore it is unlikely that a potential competitor would be willing to make the capital investment needed to even enter the monopolist's market.

Companies that grow to take advantage of economies of scale often run into problems of bureaucracy; these factors interact to produce an "ideal" size for a company, at which the company's average cost of production is minimized. If that ideal size is large enough to supply the whole market, then that market is a natural monopoly.

A further discussion and understanding requires more microeconomics:

Two different types of cost are important in microeconomics: **marginal cost**, and **fixed cost**. The marginal cost is the cost to the company of serving one more customer. In an industry where a natural monopoly does not exist, the vast majority of industries, the marginal cost decreases with economies of scale, then increases as the company has growing pains (overworking its employees, bureaucracy, inefficiencies, etc.).

Along with this, the average cost of its products will decrease and then increase again. A natural monopoly has a very different cost structure. A natural monopoly has a high fixed cost for a product that does not depend on output, but its marginal cost of producing one more good is roughly constant, and small.

A firm with high fixed costs will require a large number of customers in order to retrieve a meaningful return on their initial investment. This is where economies of scale become important. Since each firm has large initial costs, as the firm gains market share and increases its output the fixed cost (what they initially invested) is divided among a larger number of customers. Therefore, in industries with large initial investment requirements, average total cost declines as output increases over a much larger range of output levels.

Once a natural monopoly has been established because of the large initial cost and that, according to the rule of economies of scale, the larger corporation (to a point) has lower average cost and therefore a huge advantage. With this knowledge, no firms attempt to enter the industry and an oligopoly or monopoly develops.

Industries with a natural monopoly

Utilities are often natural monopolies. In industries with a standardized product and economies of scale, a natural monopoly will often arise. In the case of electricity, all companies provide the same product, the infrastructure required is immense, and the cost of adding one more customer is negligible, up to a point. Adding one more customer may increase the company's revenue and lowers the average cost of providing for the company's customer base. So long as the average cost of serving customers is decreasing, the larger firm will more efficiently serve the entire customer base. Of course, this might be circumvented by differentiating the product, making it no longer a pure commodity. For example, firms may gain customers who will pay more by selling "green" power, or non-polluting power, or locally-produced power.

Historical example

Such a process happened in the water industry in nineteenth century Britain. Up until the mid-nineteenth century, Parliament discouraged municipal involvement in water supply; in 1851, private companies had 60% of the market.

Competition amongst the companies in larger industrial towns lowered profit margins, as companies were less able to charge a sufficient price for installation of networks in new areas. In areas with direct competition (with two sets of mains), usually at the edge of companies' territories, profit margins were lowest of all.

Such situations resulted in higher costs and lower efficiency, as two networks, neither used to capacity, were used. With a limited number of households that could afford their services, expansion of networks slowed, and many companies were barely profitable. With a lack of water and sanitation claiming thousands of lives in periodic epidemics, municipalisation proceeded rapidly after 1860, and it was municipalities which were able to raise the finance for investment which private companies in many cases could not.

A few well-run private companies which worked together with their local towns and cities (gaining legal monopolies and thereby the financial security to invest as required) did survive, providing around 20% of the population with water even today.

The rest of the water industry in England and Wales was reprivatized in the form of 10 regional monopolies in 1989.

Origins of the term

The original concept of natural monopoly is often attributed to John Stuart Mill, who (writing before the marginalist revolution) believed that prices would reflect the costs of production in absence of an artificial or natural monopoly.[2] In *Principles of Political Economy* Mill criticized Smith's neglect of an area that could explain wage disparity. Taking up the examples of professionals such as jewelers, physicians and lawyers, he said,

"The superiority of reward is not here the consequence of competition, but of its absence: not a compensation for disadvantages inherent in the employment, but an extra advantage; a kind of monopoly price, the effect not of a legal, but of what has been termed a natural monopoly... independently of... artificial monopolies [i.e. grants by government], there is a natural monopoly in favor of skilled laborers against the unskilled, which makes the difference of reward exceed, sometimes in a manifold proportion, what is sufficient merely to equalize their advantages. If unskilled laborers had it in their power to compete with skilled, by merely taking the trouble of learning the trade, the difference of wages might not exceed what would compensate them for that trouble, at the ordinary rate at which labor is remunerated. But the fact that a course of instruction is required, of even a low degree of costliness, or that the laborer must be maintained for a considerable time from other sources, suffices everywhere to exclude the great body of the laboring people from the possibility of any such competition.

So Mill's initial use of the term concerned natural abilities, in contrast to the common contemporary usage, which refers solely to market failure in a particular type of industry, such as rail, post or electricity. Mill's development of the idea is that what is true of labor is true of capital.

"All the natural monopolies (meaning thereby those which are created by circumstances, and not by law) which produce or aggravate the disparities in the remuneration of different kinds of labor, operate similarly between different employments of capital. If a business can only be advantageously carried on by a large capital, this in most countries limits so narrowly the class of persons who can enter into the employment, that they are enabled to keep their rate of profit above the general level. A trade may also, from the nature of the case, be confined to so few hands, that profits may admit of being kept up by a combination among the dealers. It is well known that even among so numerous a body as the London booksellers, this sort of

combination long continued to exist. I have already mentioned the case of the gas and water companies.

Mill also used the term in relation to land, for which the natural monopoly could be extracted by virtue of it being the only land like it. Furthermore, Mill referred to network industries, such as electricity and water supply, roads, rail and canals, as "practical monopolies", where "it is the part of the government, either to subject the business to reasonable conditions for the general advantage, or to retain such power over it, that the profits of the monopoly may at least be obtained for the public." So, a legal prohibition against competition is often advocated and rates are not left to the market but are regulated by the government.

Regulation

As with all monopolies, a monopolist who has gained his position through natural monopoly effects may engage in behavior that abuses his market position, which often leads to calls from consumers for government regulation. Government regulation may also come about at the request of a business hoping to enter a market otherwise dominated by a natural monopoly.

Common arguments in favor of regulation include the desire to control market power, facilitate competition, promote investment or system expansion, or stabilize markets. In general, though, regulation occurs when the government believes that the operator, left to his own devices, would behave in a way that is contrary to the government's objectives. In some countries an early solution to this perceived problem was government provision of, for example, a utility service. However, this approach raised its own problems. Some governments used the state-provided utility services to pursue political agendas, as a source of cash flow for funding other government activities, or as a means of obtaining hard currency. These and other consequences of state provision of services often resulted in inefficiency and poor service quality. As a result, governments began to seek other solutions, namely regulation and providing services on a commercial basis, often through private participation.

As a quid pro quo for accepting government oversight, private suppliers may be permitted some monopolistic returns, through stable prices or guaranteed through limited rates of return, and a reduced risk of long-term competition. (See also rate of return pricing). For example, an electric utility may be allowed to sell electricity at price that will give it a 12% return on its

capital investment. If not constrained by the public utility commission, the company would likely charge a far higher price and earn an abnormal profit on its capital.

Regulatory responses:

- doing nothing
- setting legal limits on the firm's behavior, either directly or through a regulatory agency
- setting up competition *for* the market (franchising)
- setting up common carrier type competition
- setting up surrogate competition ("yardstick" competition or benchmarking)
- requiring companies to be (or remain) quoted on the stock market
- public ownership

Since the 1980s there is a global trend towards utility deregulation, in which systems of competition are intended to replace regulation by specifying or limiting firms' behavior; the telecommunications industry is a leading example globally.

Doing nothing

Because the existence of a natural monopoly depends on an industry's cost structure, which can change dramatically through new technology (both physical and organizational/institutional), the nature or even existence of natural monopoly may change over time. A classic example is the undermining of the natural monopoly of the canals in eighteenth century Britain by the emergence in the nineteenth century of the new technology of railways.

Arguments from public choice suggest that regulatory capture is likely in the case of a regulated private monopoly. Moreover, in some cases the costs to society of overzealous regulation may be higher than the costs of permitting an unregulated private monopoly. (Although the monopolist charges monopoly prices, much of the price increase is a transfer rather than a loss to society.)

More fundamentally, the theory of contestable markets developed by Baumol and others argues that monopolists (including natural monopolists) may be forced over time by the mere *possibility* of competition at some point in the future to limit their monopolistic behavior, in order to deter entry. In the limit, a monopolist is forced to make the same production

decisions as a competitive market would produce. A common example is that of airline flight schedules, where a particular airline may have a monopoly between destinations A and B, but the relative ease with which in many cases competitors could also serve that route limits its monopolistic behavior. The argument even applies somewhat to government-granted monopolies, as although they are protected from competitors entering the industry, in a democracy excessively monopolistic behavior may lead to the monopoly being revoked, or given to another party.

Nobel economist Milton Friedman, said that in the case of natural monopoly that "there is only a choice among three evils: private unregulated monopoly, private monopoly regulated by the state, and government operation." He said "the least of these evils is private unregulated monopoly where this is tolerable." He reasons that the other alternatives are "exceedingly difficult to reverse" and that the dynamics of the market should be allowed the opportunity to have an effect and are likely to do so (*Capitalism and Freedom*). In a Wincott Lecture, he said that if the commodity in question is "essential" (for example: water or electricity) and the "monopoly power is sizeable," then "either public regulation or ownership may be a lesser evil." However, he goes on to say that such action by government should not consist of forbidding competition by law. Friedman has taken a stronger laissez-faire stance since, saying that "over time I have gradually come to the conclusion that antitrust laws do far more harm than good and that we would be better off if we didn't have them at all, if we could get rid of them" (*The Business Community's Suicidal Impulse*).

Advocates of laissez-faire capitalism, such as libertarians, typically say that permanent natural monopolies are merely theoretical. Economists from the Austrian school claim that governments take ownership of the means of production in certain industries and ban competition under the false pretense that they are natural monopolies.

Franchising and outsourcing

Although competition within a natural monopoly market is costly, it is possible to set up competition *for* the market. This has been, for example, the dominant organizational method for water services in France, although in this case the resulting degree of competition is limited by contracts often being set for long periods (30 years), and there only being three major competitors in the market.

Equally, competition may be used for *part* of the market (e.g. IT services), through outsourcing contracts; some water companies outsource a considerable proportion of their operations. The extreme case is Welsh Water, which outsources virtually its entire business operations, running just a skeleton staff to manage these contracts. Franchising different parts of the business on a regional basis (e.g. parts of a city) can bring in some features of "yardstick" competition (see below), as the performance of different contractors can be compared. See also water privatization.

Common carriage competition

This involves different firms competing to distribute goods and services via the same infrastructure - for example different electricity companies competing to provide services to customers over the same electricity network. For this to work requires government intervention to break up vertically integrated monopolies, so that for instance in electricity, generation is separated from distribution and possibly from other parts of the industry such as sales. The key element is that access to the network is available to any firm that needs it to supply its service, with the price the infrastructure owner is permitted to charge being regulated. (There are several competing models of network access pricing.) In the British model of electricity liberalization, there is a market for generation capacity, where electricity can be bought on a minute-to-minute basis or through longer-term contracts, by companies with insufficient generation capacity (or sometimes no capacity at all).

Such a system may be considered a form of deregulation, but in fact it requires active government creation of a new system of competition rather than simply the removal of existing legal restrictions. The system may also need continuing government fine tuning, for example to prevent the development of long-term contracts from reducing the liquidity of the generation market too much, or to ensure the correct incentives for long-term security of supply are present. See also California electricity crisis. Whether such a system is more efficient than possible alternatives is unclear; the cost of the market mechanisms themselves are substantial, and the vertical de-integration required introduces additional risks. This raises the cost of finance - which for a capital intensive industry (as natural monopolies are) is a key issue. Moreover, such competition also raises equity and efficiency issues, as large industrial consumers tend to benefit much more than domestic consumers.

Stock market

One regulatory response is to require that private companies running natural monopolies be quoted on the stock market. This ensures they are subject to certain financial transparency requirements, and maintains the possibility of a takeover if the company is mismanaged. The latter in theory should help ensure that company is efficiently run. By way of example, the UK's water economic regulator, Ofwat, sees the stock market as an important regulatory instrument for ensuring efficient management of the water companies.

In practice, the notorious short-termism of the stock market may be antithetical to appropriate spending on maintenance and investment in industries with long time horizons, where the failure to do so may only have effects a decade or more hence (which is typically long after current chief executives have left the company).

Public ownership

A traditional solution to the regulation problem, especially in Europe, is public ownership. This 'cuts out the middle man': instead of government regulating a firm's behavior, it simply takes it over (usually by buy-out), and sets itself limits within which to act.

Network effects

Network effects are considered separately from natural monopoly status. Natural monopoly effects are a property of the producer's cost curves, whilst network effects arise from the benefit to the consumers of a good from standardization of the good. Many goods have both properties, like operating system software and telephone networks.

References

1. DiLorenzo, Thomas J. (1996), *The Myth of the Natural Monopoly*, The Review of Austrian Economics 9(2)
2. *Principles of Political Economy*, Book IV 'Influence of the progress of society on production and distribution', Chapter 2 'Influence of the Progress of Industry and Population on Values and Prices', para. 2
3. *Wealth of Nations* (1776) Book I, Chapter 10
4. *Principles of Political Economy* Book II, Chapter XIV 'Of the Differences of Wages in different Employments',
5. *Principles of Political Economy* Book II, Chapter XV, 'Of Profits', para. 9

6. *Principles of Political Economy*, Book II, Chapter XVI, 'Of Rent', para. 2
7. *Principles of Political Economy*, Book V, Chapter XI 'Of the Grounds and Limits of the Laisser-faire or Non-Interference Principle'
8. On subways, see also, McEachern, Willam A. (2005). *Economics: A Contemporary Introduction*. Thomson South-Western. pp. 319.
9. Body of Knowledge on Infrastructure Regulation "General Concepts"
10. DiLorenzo, Thomas J. (1996). "The Myth of Natural Monopoly". *The Review of Austrian Economics* **9** (2): 43–58. doi:10.1007/BF01103329. http://www.mises.org/journals/rae/pdf/rae9_2_3.pdf.
11. Berg, Sanford (1988). *Natural Monopoly Regulation: Principles and Practices*. Cambridge University Press. ISBN 9780521338936.
12. Baumol, William J.; Panzar, J. C., and Willig, R. D. (1982). *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovich. ISBN 978-0155139107.
13. Filippini, Massimo (June 1998). "Are Municipal Electricity Distribution Utilities Natural Monopolies?". *Annals of Public and Cooperative Economics* **69** (2): 157. doi:10.1111/1467-8292.00077.
14. Foldvary, Fred E.; Daniel B. Klein (2003). *The Half-Life of Policy Rationales: How New Technology Affects Old Policy Issues*. Cato Institute. http://lsb.scu.edu/~dklein/Half_Life/Half_Life.htm.
15. Sharkey, W. (1982). *The Theory of Natural Monopoly*. Cambridge University Press. ISBN 9780521271943.
16. Thierer, Adam D (1994). "Unnatural Monopoly: Critical Moments in the Development of the Bell System Monopoly". *The Cato Journal* **14** (2). http://www.cato.org/pubs/journal/cjv14n2-6.html.
17. Train, Kenneth E. (1991). *Optimal regulation: the economic theory of natural monopoly*. Cambridge, MA, USA: MIT Press. ISBN 978-0262200844.
18. Waterson, Michael (1988). *Regulation of the Firm and Natural Monopoly*. New York, NY, USA: Blackwell. ISBN 0-631-14007-7.

23. The rich get richer and the poor get poorer

"The rich get richer and the poor get poorer" is a catchphrase and proverb, frequently used (with variations in wording) in discussing economic inequality.

Predecessors

Andrew Jackson, in his 1832 bank veto, said that when the laws undertake... to make the rich richer and the potent more powerful, the humble members of society... have a right to complain of the injustice to their Government.

William Henry Harrison said, in an October 1, 1840 speech, I believe and I say it is true Democratic feeling, that all the measures of the government are directed to the purpose of making the rich richer and the poor poorer.

In 1821, Percy Bysshe Shelley argued, in A Defense of Poetry (not published until 1840), that in his England, "the promoters of utility" had managed to exasperate at once the extremes of luxury and want. They have exemplified the saying, "To him that hath, more shall be given; and from him that hath not, the little that he hath shall be taken away." The rich have become richer, and the poor have become poorer; and the vessel of the State is driven between the Scylla and Charybdis of anarchy and despotism. Such are the effects which must ever flow from an unmitigated exercise of the calculating faculty.

The phrase resembles the Bible verse: For whosoever hath, to him shall be given, and he shall have more abundance: but whosoever hath not, from him shall be taken away even that he hath.

However, in this verse Jesus is not referring to economic inequality at all. Rather it is part of his answer to the question "Why speakest thou unto them in parables?" Jesus says his parables give fresh understanding only to those who already have accepted his message.

"Ain't We Got Fun"

A version of the phrase was popularized in 1921 in the wildly successful song Ain't We Got Fun, and the phrase sometimes attributed to the song's lyricists, Gus Kahn and Raymond B. Egan. Oddly, the lyrics never actually say

that the poor get "poorer;" instead it takes off from or alludes to the line, showing that it was already proverbial. They cue the listener to expect the word "poorer," but instead say

There's nothing surer: The rich get rich and the poor get—children;

and, later:

There's nothing surer: The rich get rich and the poor get laid off;

Note too that the Kahn and Egan lyrics say "the rich get *rich,*" not *richer.*

The line is sometimes mistakenly attributed to F. Scott Fitzgerald. It appears in The Great Gatsby, as

the rich get richer[sic] and the poor get—children!

The character Gatsby orders the character Klipspringer, sitting at the piano, "Don't talk so much, old sport.... "Play!" and Klipspringer breaks into the Kahn and Egan song.

In political and economic rhetoric

The line is often cited by opponents of capitalism as a statement of fact and by supporters of capitalism as an example of an erroneous belief. Thus,

The modern-day statistical work of Stanley Lebergott and Michael Cox confirms this Smithian view and disputes the commonly held criticism that under a free markets the rich get richer and the poor get poorer.

According to Marx, capitalism will inevitably lead to ruin in accordance with certain laws of economic movement. These laws are "the Law of the Tendency of the Rate of Profit to Fall," "the Law of Increasing Poverty," and "the Law of Centralization of Capital." Small capitalists go bankrupt, and their production means are absorbed by large capitalists. During the process of bankruptcy and absorption, capital is gradually centralized by a few large capitalists, and the entire middle class declines. Thus, two major classes, a small minority of large capitalists, and a large proletarian majority are formed.[13]

A use of the phrase by a free market advocates disputing the claim is:

Relative cohort inequity decreased markedly, with the poor improving their position much faster than the rich. Relative percentile inequity increased slightly. In terms of buying power, both the poor cohort and the poor percentile became significantly wealthier. These data indicate that the view that the rich are getting richer and the poor are getting poorer is clearly over-generalized.

In the United States the phrase has been used frequently (in the past tense) to describe alleged socioeconomic trends under the Ronald Reagan and George H. W. Bush presidencies. Defenders of the Reagan policies characterize this claim as inciting class warfare.

Commentators refer to the idea as a cliché in discussions of economic inequality, but one that they argue to be accurate nonetheless:

It's a cliché, perhaps, to say that "the rich get rich and the poor get poorer". But in the 1980s and 1990s, cliché or not, that is what took place in some regions of the world, particularly in South Asia and sub-Saharan Africa.

References

1. Watson, Harry L. (1998). *Andrew Jackson V. Henry Clay: Democracy and Development in Antebellum America*. Palg. Macmillan. ISBN 0-312-17772-0.
2. "William Henry Harrison quotes". James Richard Howington (personal web page). http://home.att.net/~howington/whh.html. Retrieved 2006-08-12.
3. Degregorio, William (1997). *Complete Book of U.S. Presidents: From George Washington to George W. Bush*. Gramercy. ISBN 0-517-18353-6., p. 146; quotes "all the measures of the government are directed to the purpose of making the rich richer and the poor poorer" and sources it to Schlesinger, Arthur (1946). *The Age of Jackson*. Boston: Little, Brown., p. 292
4. Shelley, Percy Bysshe (1909–14). [A Defence of Poetry (from he Harvard Classics: English Essays: Sidney to Macaulay]. Retrieved 2006-08-12.
5. Matthew 13:12, King James translation
6. "Ain't We Got Fun". Everything2. http://everything2.net/index.pl?node_id=1101156. Retrieved 2006-08-11.
7. "Ain't We Got Fun". Don Ferguson. http://www.oocities.com/dferg5493/ aintwegotfun.htm. 2006-08-11.
8. 1921 recording by Billy Jones: mp3 file, 3.4 mb from the UCSB cylinder preservation project
9. Fitzgerald, F. Scott (1998) [1921]. *The Great Gatsby*. Oxford University Press. ISBN 0-19-283269-7. p. 76; also at Project Gutenberg of Australia
10. Skousen, Mark. *The Making of Modern Economics: The Lives and Ideas of the Great Thinkers*. p. 26
11. http://www.unification.org/ucbooks/cncc/cncc-05.htm.

12. http://taylorandfrancis.metapress.com/(lrmptp55doytq0bsyarkhxnb)/app/home/contribution.asp?referrer=parent&backto=issue,5,8;journal,37,124;linkingpublicationresults,1:107889,1.
13. James K. Galbraith. "The rich got richer." *Salon Magazine*, June 8, 2004
14. Edward N. Wolf. "The rich get increasingly richer: on household wealth during the 1980s." Economic Policy Institute Briefing Paper #36. 1993.
15. Alan Reynolds. "Upstarts and Downstarts (The Real Reagan Record)." *National Review*, August 31, 1992.
16. Jude Wanniski citing Bruce Bartlett, "Class Struggle in America?," Commentary Magazine, July 2005
17. James L. Dietz, James M. Cypher (2004-04-01). *The Process of Economic Development*. United Kingdom: Routledge. p. 15. ISBN 0-415-25416-7.
18. Brian Hayes (September 2002). "Follow the Money". *American Scientist* 90 (5): 400. doi:10.1511/2002.5.400. http://americanscientist.org/Issues/Comsci02/02-09Hayes.html.Hayes analyzes several computer models of market economies, applying statistical mechanics to questions in economic theory in the same way that it is applied in computational fluid dynamics, concluding that "If some mechanism like that of the yard-sale model is truly at work, then markets might very well be free and fair, and the playing field perfectly level, and yet the outcome would almost surely be that the rich get richer and the poor get poorer."
19. http://www.americanscientist.org/issues/pub/follow-the-money
20. Rieman, J. (1979). *The Rich Get Richer and The Poor Get Poorer*. New York: Wiley.
21. David Hapgood (1974). *The Screwing of the Average Man. How The Rich Get Richer and You Get Poorer*. Bantom Books. ISBN 0-553-12913-9.
22. Rolf R Mantel (1995). *Why the rich get richer and the poor get poorer*. Universidad de San Andrés: Victoria, prov. de Buenos Aires. OCLC 44260846 44260846.
23. S. Ispolatov, P.L. Krapivsky, and S. Redner (March 1998). "Wealth distributions in asset exchange models". *The European Physical Journal B — Condensed Matter and Complex Systems* (Berlin / Heidelberg: Springer) 2: 267–276. doi:10.1007/s100510050249. ISSN 1434-6028/1434-6036. — Ispolatov, Krapivsky, and Redner analyze the wealth distributions that occur under a variety of exchange rules in a system of economically interacting people.
24. Kee H. Chung and Raymond A. K. Cox (March 1990). "Patterns of Productivity in the Finance Literature: A Study of the Bibliometric

Distributions". *Journal of Finance* **45** (1): 301–309. doi:10.2307/2328824. — Chung and Cox analyze bibliometric regularity in finance literature, relating Lotka's law of scientific prductivity to the maxim that "the rich get richer and the poor get poorer", and equating it to the maxim that "success breeds success".

24.Ain't We Got Fun?

For the 1937 Merrie Melodies cartoon, see Ain't We Got Fun (cartoon).

Cover page to the sheet music.
By Billy Jones, 1921. for Edison Records.

"Ain't We Got Fun?" is a popular foxtrot published in 1921 with music by Richard A. Whiting, lyrics by Raymond B. Egan and Gus Kahn.

It was first performed in 1920 in the revue *Satires of 1920*, then moved into vaudeville and recordings. "Ain't We Got Fun?" and both its jaunty response to poverty and its promise of fun "Every morning / Every evening", and "In the meantime, / In between time" have become symbolic of the Roaring

Twenties, and it appears in some of the major literature of the decade, including *The Great Gatsby* by F. Scott Fitzgerald and in Dorothy Parker's award-winning short story of 1929, "Big Blonde".

Composition

"Ain't We Got Fun" follows the structure of a foxtrot. The melody uses mainly quarter notes, and has an unsyncopated refrain made up largely of variations on a repeated four-note phrase.

The *Tin Pan Alley Song Encyclopedia* describes it as a "Roaring Twenties favorite" and praises its vibrancy, "zesty music", and comic lyrics.

Philip Furia, connecting Kahn's lyrics to the song's music, writes that:

Not only does Kahn use abrupt, colloquial--even ungrammatical--phrases, he abandons syntax for the telegraphic connections of conversation. Truncated phrases like *not much money* are the verbal equivalent of syncopated musical fragments.

—Philip Furia, *The Poets of Tin Pan Alley:*

A History of America's Great Lyricists Critical appraisals vary regarding what view of poverty the song's lyrics take. Nicholas E. Tawa summarizes the refrain *Ain't we got fun* as a *satirical and jaunty rejoinder* toward hard times. Diane Holloway and Bob Cheney, authors of *American History in Song: Lyrics from 1900 to 1945*, concur, and describe the black humor in the couple's relief that their poverty shields them from worrying about damage to their nonexistent Pierce Arrow luxury automobile.

Yet George Orwell highlights the lyrics of "Ain't We Got Fun" as an example of working class unrest:

All through the war and for a little time afterwards there had been high wages and abundant employment; things were now returning to something worse than normal, and naturally the working class resisted. The men who had fought had been lured into the army by gaudy promises, and they were coming home to a world where there were no jobs and not even any houses. Moreover, they had been at war and were coming home with a soldier's attitude to life, which is fundamentally, in spite of discipline, a lawless attitude. There was a turbulent feeling in the air.

—George Orwell, *The Road to Wigan Pier*

After quoting a few of the song's lines Orwell refers to the era as a time when *people had not yet settled down to a lifetime of unemployment*

mitigated by endless cups of tea, a turn of phrase which the later writer Larry Portis contests.

He [Orwell] could just as easily have concluded that the song revealed certain fatalism, a resignation and even capitulation to forces beyond the control of working people. Indeed, it might be only a small step from saying, "Ain't we got fun" in the midst of hardship to the idea that the poor are happier than the rich--because, as the Beatles intoned, "Money can't buy me love." It is possible that "Aint We Got Fun", a product of the music industry (as opposed to 'working-class culture') was part of a complex resolution of crisis in capitalist society. Far from revealing the indomitable spirit of working people, it figured into the means with which they were controlled. It is a problem of interpretation laying at the heart of popular music, one which emerged with particular clarity at the time of the English Industrial Revolution.

—Larry Portis, Soul Trains

However, others concentrate on the fun that they got. Stephen J. Whitfield, citing lyrics such as "Every morning / Every evening / Ain't we got fun", writes that the song "set the mood which is indelibly associated with the Roaring Twenties", a decade when pleasure was sought and found constantly, morning, evening, and "In the meantime / In between time".

Philip Furia and Michael Lasser see implicit references to sexual intercourse in lyrics such as *the happy chappy, and his bride of only a year.*[11] Looked at in the context of the 1920s, an era of increasing sexual freedom, they point out that, while here presented within the context of marriage (in other songs it is not), the sexuality is notably closer to the surface than in previous eras and is presented as a delightful, youthful pleasure. There are several variations on the lyrics. For example, *American History in Song* quotes the lyrics:

They won't smash up our Pierce Arrow, / We ain't got none
They've cut my wages / But my income tax will be so much smaller
When I'm paid off, / I'll be laid off
Ain't we got fun?

The sheet music published in 1921 by Jerome K. Remick and Co. leaves this chorus out completely, whereas a recording for Edison Records by Billy Jones keeps the reference to the Pierce Arrow, but then continues as in the

sheet music: "There's nothing surer / The rich get rich and the poor get laid off / In the meantime,/ In between time/ Ain't we got fun?"

Reception and performance history

It premièred in the show *Satires of 1920*, where it was sung by Arthur West, then entered the vaudeville repertoire of Ruth Roye.[15] A hit recording by Van and Schenck increased its popularity, and grew into a popular standard.

The song appears in the F. Scott Fitzgerald novel *The Great Gatsby*, when Daisy Buchanan and Gatsby meet again after many years, and appears in Dorothy Parker's 1929 short story, "Big Blonde". Warner Brothers used the song in two musicals during the early 1950s: The Gus Kahn biopic *I'll See You in My Dreams* and *The Eddie Cantor Story*.[15] Woody Allen used the song in his 1983 film *Zelig*.

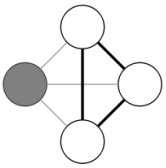
Notable Recordings

Doris Day for her album "By the Light of the Silvery Moon
The song was featured in the film "By the Light of the Silvery Moon" (1953), and performed by Doris Day and Gordon MacRae.

25. Clustering coefficient

In graph theory, a **clustering coefficient** is a measure of degree to which nodes in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterized by a relatively high density of ties (Holland and Leinhardt, 1971; Watts and Strogatz, 1998). In real-world networks, this likelihood tends to be greater than the average probability of a tie randomly established between two nodes (Holland and Leinhardt, 1971; Watts and Strogatz, 1998).

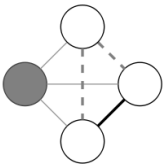
Two versions of this measure exist: the global and the local. The global version was designed to give an overall indication of the clustering in the network, whereas the local gives an indication of the embeddedness of single nodes.



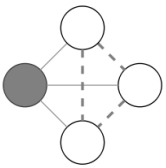
c = 1

Global clustering coefficient

The global clustering coefficient is based on triplets of nodes. A triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) undirected ties. A triangle consists of three closed triplets, one centered on each of the nodes. The global clustering coefficient is the number of closed triplets (or 3 x triangles) over the total number of triplets (both open and closed). The first attempt to measure it was made by Luce and Perry (1949). This measure gives an indication of the clustering in the whole network (global), and can be applied to both undirected and directed networks (often called transitivity, see Wasserman and Faust, 1994, page 243).



c = 1/3



c = 0

Formally, it has been defined as:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}} =$$

$$= \frac{\text{number of closed triplets}}{\text{number of connected triples of vertices}}$$

A generalization to weighted networks was proposed by Opsahl and Panzarasa (2009), and a redefinition to two-mode networks (both binary and weighted) by Opsahl (2009).

Local clustering coefficient

The local clustering coefficient of the light blue node is computed as the proportion of connections among its neighbors which are actually realized compared with the number of all possible connections. In the figure, the light blue node has three neighbors, which can have a maximum of 3 connections among them. In the top part of the figure all three possible connections are realized (thick black segments), giving a local clustering coefficient of 1. In the middle part of the figure only one connection is realized (thick black line) and 2 connections are missing (dotted red lines), giving a local cluster coefficient of 1/3. Finally, none of the possible connections among the neighbors of the light blue node are realized, producing a local clustering coefficient value of 0.

The **local clustering coefficient** of a vertex in a graph quantifies how close its neighbors are to being a clique (complete graph). Duncan J. Watts and Steven Strogatz introduced the measure in 1998 to determine whether a graph is a small-world network.

A graph $G = (V,E)$ formally consists of a set of vertices V and a set of edges E between them. An edge e_{ij} connects vertex i with vertex j.

The neighborhood N for a vertex v_i is defined as its immediately connected neighbors as follows:

$$N_i = \{v_j : e_{ij} \in E \wedge e_{ji} \in E\}$$

We define k_i as the number of vertices, $|N_i|$, in the neighborhood, N_i, of a vertex.

The local clustering coefficient C_i for a vertex v_i is then given by the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them. For a directed graph, e_{ij} is distinct from e_{ji}, and therefore for each neighborhood N_i there are $k_i(k_i - 1)$ links that could exist among the vertices within the neighborhood (k_i is the total (in + out) degree of the vertex). Thus, the **local clustering coefficient for directed graphs** is given as

$$C_i = \frac{|\{e_{jk}\}|}{k_i(k_i - 1)} : v_j, v_k \in N_i, e_{jk} \in E$$

An undirected graph has the property that e_{ij} and e_{ji} are considered identical. Therefore, if a vertex v_i has k_i neighbors, $\frac{k_i(k_i-1)}{2}$ edges could exist among the vertices within the neighborhood. Thus, the **local clustering coefficient for undirected graphs** can be defined as

$$C_i = \frac{2\,|\{e_{jk}\}|}{k_i(k_i-1)} : v_j, v_k \in N_i, e_{jk} \in E$$

Let $\lambda_G(v)$ be the number of triangles on $v \in V(G)$ for undirected graph G. That is, $\lambda_G(v)$ is the number of subgraphs of G with 3 edges and 3 vertices, one of which is v. Let $\tau_G(v)$ be the number of triples on $v \in G$. That is, $\tau_G(v)$ is the number of subgraphs (not necessarily induced) with 2 edges and 3 vertices, one of which is v and such that v is incident to both edges. Then we can also define the clustering coefficient as

$$C_i = \frac{\lambda_G(v)}{\tau_G(v)}$$

It is simple to show that the two preceding definitions are the same, since

$$\tau_G(v) = C(k_i, 2) = \frac{1}{2}k_i(k_i-1)$$

These measures are 1 if every neighbor connected to v_i is also connected to every other vertex within the neighborhood, and 0 if no vertex that is connected to v_i connects to any other vertex that is connected to v_i.

Network average clustering coefficient

The clustering coefficient for the whole network is given by Watts and Strogatz as the average of the local clustering coefficients of all the vertices n:

$$\overline{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$$

A graph is considered small-world, if its average clustering coefficient $\overline{C}$ is significantly higher than a random graph constructed on the same vertex set, and if the graph has approximately the same mean-shortest path length as its corresponding random graph.

A generalization to weighted networks was proposed by Barrat et al. (2004), and a redefinition to bipartite graphs (also called two-mode networks) by Latapy et al. (2008) and Opsahl (2009).

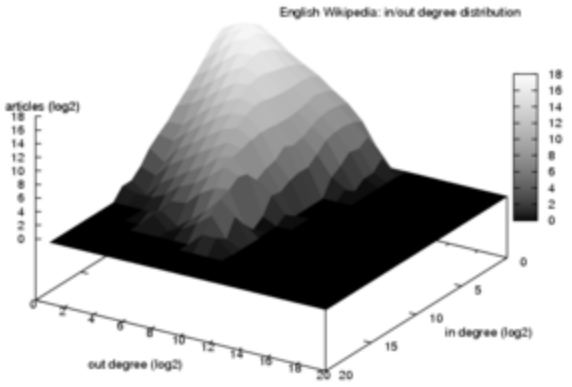
This formula is not, by default, defined for graphs with isolated vertices; see Kaiser, (2008) and Barmpoutis et al, The networks with the largest possible average clustering coefficient are found to have a modular structure, and at the same time, they have the smallest possible average distance among the different nodes.

References

1. P. W. Holland and S. Leinhardt (1971). "Transitivity in structural models of small groups". *Comparative Group Studies* 2: 107–124.
2. D. J. Watts and Steven Strogatz (June 1998). "Collective dynamics of 'small-world' networks". *Nature* 393 (6684): 440–442. doi:10.1038/30918. PMID 9623998. http://tam.cornell.edu/tam/cms/manage/upload/SS_nature_smallworld.pdf
3. R. D. Luce and A. D. Perry (1949). "A method of matrix analysis of group structure". *Psychometrika* 14 (1): 95–116. doi:10.1007/BF02289146. PMID 18152948 .
4. Stanley Wasserman, Kathrine Faust, 1994. *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.
5. Tore Opsahl & Pietro Panzarasa (2009). "Clustering in Weighted Networks". *Social Networks* 31 (2): 155–163. doi:10.1016/j.socnet. 2009.02.002. http://toreopsahl.com/2009/04/03/article-clustering-in-weighted-networks.
6. Tore Opsahl (2009). "Clustering in Two-mode Networks". *Conference and Workshop on Two-Mode Social Analysis (Sept 30-Oct 2, 2009).* http://toreopsahl.com/2009/09/11/clustering-in-two-mode-networks.
7. A. Barrat and M. Barthelemy and R. Pastor-Satorras and A. Vespignani (2004). "The architecture of complex weighted networks". *Proceedings of the National Academy of Sciences* 101 (11): 3747–37523747–3752 . doi:10.1073/pnas.0400087101. PMC 374315. PMID 1500716515007165 . http://www.pubmedcentral.nih.gov/articlerender.fcgi?tool=pmcentrez &artid=374315.
8. M. Latapy and C. Magnien and N. Del Vecchio (2008). "Basic Notions for the Analysis of Large Two-mode Networks". *Social Networks* 30 (1): 31–48. doi:10.1016/j.socnet.2007.04.006.
9. D. Barmpoutis and R. M. Murray (2010). "Networks with the Smallest Average Distance and the Largest Average Clustering". arXiv:1007.4031.

26.Degree distribution

In the study of graphs and networks, the degree of a node in a network is the number of connections it has to other nodes and the **degree distribution** is the probability distribution of these degrees over the whole network.
This figure here is an in/out degree-distribution for a hyperlink graph (logarithmic scales).



Definition
The degree of a node in a network (sometimes referred to incorrectly as the connectivity) is the number of connections or edges the node has to other nodes. If a network is directed, meaning that edges point in one direction from one node to another node, then nodes have two different degrees, the in-degree, which is the number of incoming edges, and the out-degree, which is the number of outgoing edges.
The degree distribution $P(k)$ of a network is then defined to be the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and n_k of them have degree k, we have $P(k) = n_k/n$.
The same information is also sometimes presented in the form of a *cumulative degree distribution*, the fraction of nodes with degree greater than or equal to k.

Observed degree distributions
The degree distribution is very important in studying both real networks, such as the Internet and social networks, and theoretical networks. The

simplest network model, for example, the (Bernoulli) random graph, in which each of *n* nodes is connected (or not) with independent probability *p* (or 1 – *p*), has a binomial distribution of degrees:

$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$, (or Poisson in the limit of large *n*).

Most networks in the real world, however, have degree distributions very different from this. Most are highly right-skewed, meaning that a large majority of nodes have low degree but a small number, known as "hubs", have high degree. Some networks, notably the Internet, the world wide web, and some social networks are found to have degree distributions that approximately follow a power law: $P(k) \sim k^{-\gamma}$, where γ is a constant. Such networks are called scale-free networks and have attracted particular attention for their structural and dynamical properties.

References

1. Albert, R.; Barabasi, A.-L. (2002). "Statistical mechanics of complex networks". *Reviews of Modern Physics* 74: 47–97. arXiv:cond-mat/0106096. Bibcode 2002RvMP...74...47A. doi:10.1103/RevModPhys.74.47.
2. Dorogovtsev, S.; Mendes, J. F. F. (2002). "Evolution of networks". *Advances in Physics* 51 (4): 1079–1187. arXiv:cond-mat/0106144. doi:10.1080/00018730110112519.
3. Newman, M. E. J. (2003). "The structure and function of complex networks". *SIAM Review* 45 (2): 167–256. doi:10.1137/S003614450342480. http://siamdl.aip.org/getabs/servlet/GetabsServlet?prog=normal&id=SIREAD000045000002000167000001.

27. Epilogue

I have to tell you some word about my success using the contents of this book. As a computer scientist I have a great deal of network knowledge, by doing many years of research works and giving lectures. Recently, some friends of mine who form a community named Szentendre Szalon – reachable on web at www.szalon.tk - asked me to tell about networks understandable by most of them. They are from a wide scale of human knowledge fields spread from the science, engineering through medical as far as artists.

This collection made me possible to systematize the network knowledge getting me possible giving commonly understandable performances. We are already over half a dozen lectures and looking forward some more in the next season. I know that this book requires more thorough groundwork from the readers but for a lecturer it is a must.

I also have other occasions to use this book for. Students in scientific circles require more deep knowledge on this field. I already gave lecture-series about networks based on this collection and also I looking forward to continue, to repeat such series. These students always asked me getting electronic copies of the relating chapters.

This is why I decided to publish the whole collection in one.

The people must come to learn that the small world relation in written format can be first found in a publication from 1929 by a Hungarian humorist *Karinthy Frigyes: Láncszemek (F. Karinthy: Chain of links).*

It was published in one of his collected work, in which he propagates that *"Everything is different as you would think of"*. And I tell you this is no wonder at all! Wise people could prove it by their statements.

Let me represent this by citing two famous man wise, and full opposite sayings:

P. Erdős: "God likes take risks."

A. Einstein: "God not plays roulette with the Universe."

In conclusion I wish everybody success using this networking breviary.

~~~~
~~~~

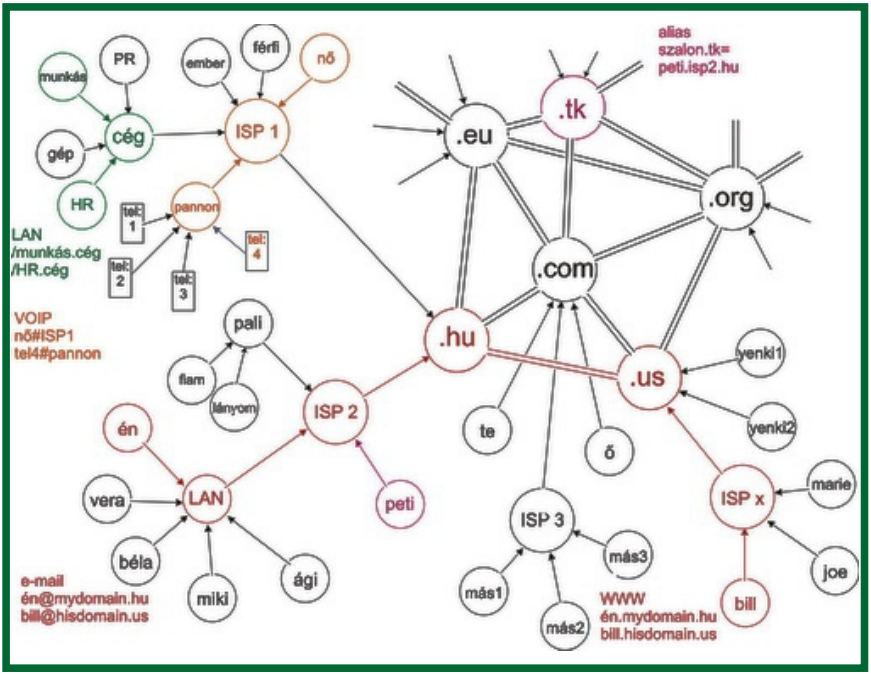
The TCC COMPUTER STUDIO is the publisher, a member of the 1988 founded TCC-GROUP. Address: Hungary, 2000 Szentendre, Pannonia str. 11.

Phone: +36-26-314-590, e-mail: tcc66@hotmail.com.
Editor in chief : A. Terjék.

Printing & binding by CORBIS DIGITAL Hungary, 1033 Budapest, Szentendrei str. 89-93., PP Center, phone: +36-1-999-6593, e-mail: corbiskft@gmail.com, www.corbis-digital.hu.
Printed in size 11(A/5) sheet.

ISBN 978-963-08-1468-3

PETER G. GYARMATI (1941) a mathematician and electronic engineer worked trough years as a research professor with networks. The object of his many lectures gave us an overview about network science and also some of us went into deep research with him on details using hard math disciplines. This book gives a collection about his lectures and offered as a breviary for scientists. 2011.



ISBN 978-963-08-1468-3