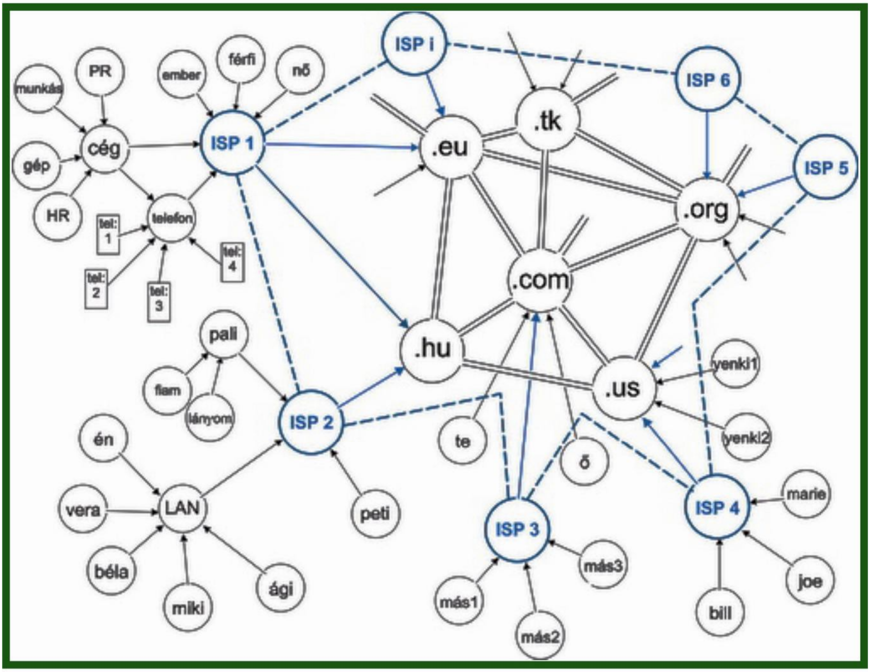
PETER G. GYARMATI

SOME WORDS ABOUT NETWORKS



TCC COMPUTER STUDIO

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

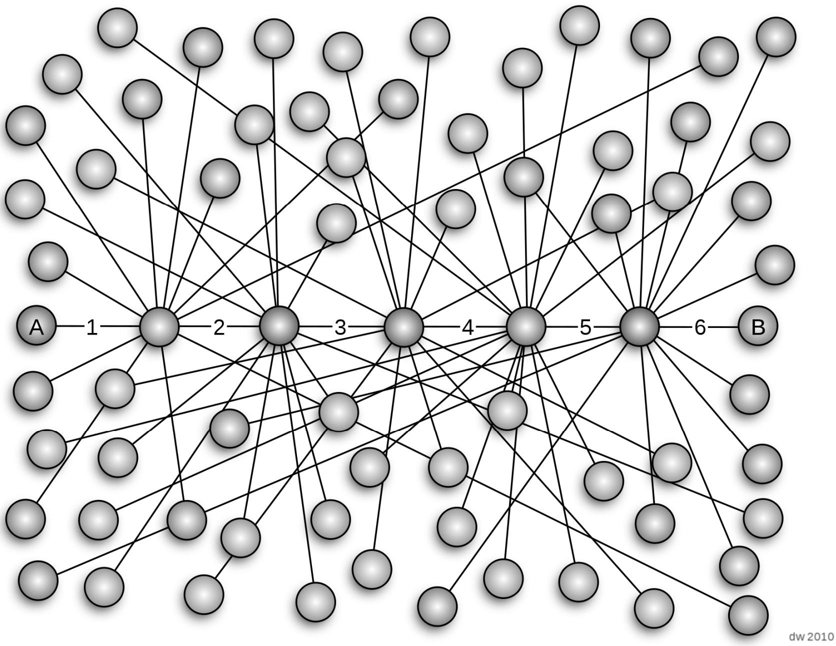
ISBN 978-963-08-1468-3

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.

~~~~~

.

.

.

.

.
~~~~~

.

.

.

.

.

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 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 $\lim_{x\to\infty} L(t\,x)/L(x) = 1$ 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 $\lambda = 0$.) 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 $x \rightarrow \infty$ 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.

6. 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.

~~~~~
~~~~~