Network Structure
What is a Network?
Technological Networks

The Internet

Vertices are computers or computer routers
Edges are cables or optical fiber lines linking the computers
Technological Networks

A flight path network

Vertices indicate airports
Edges indicate connecting flights
Biological Networks

A (directed) food chain network

Vertices indicate species
Directed edge indicates that the target species depends on the other species
Biochemical Networks

A (directed) gene transcription network

Vertices indicate proteins
Directed edge indicates that one protein regulates the transcription of the target protein
C. Elegans Connectome

From Mitya Chklovskii.

**Vertices** = neuron

**Edges** = synaptic connection between two neurons
Social Networks

Pattern of friendships among members of a karate club

Vertices indicate individuals
Edges indicate friendships

Zachery (1977)
At each node, with probability $p$ add an edge from a node to some other non-neighboring node. This is called the Watts-Strogatz algorithm.
Erdos-Renyi Random Network

130 nodes
215 edges

Red = five nodes with highest degree
Green = their neighbors
Degree Distribution Follows a Poisson Distribution

Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}

\lambda > 0 \quad \text{is a parameter}
Scale-Free Network

130 nodes
215 edges

Red = five nodes with highest degree
Green = their neighbors

Scale-free networks are characterized by a few hubs with very high degree
Scale-Free Degree Distribution Follows a Power Law

\[ \Pr[X = k] = Ck^{-\gamma} \]

\( \gamma > 0 \) is a parameter

For \( k \) sufficiently large
Degree Distribution Follows a Poisson Distribution

“Homogeneous”

“Heterogeneous”
Algorithm For Building a Scale-Free Network

1. Start with $m_0$ nodes randomly connected
Algorithm For Building a Scale-Free Network

1. Start with $m_0$ nodes randomly connected

2. Add additional nodes where probability of connecting to each existing node is larger if that node has high degree
Algorithm For Building a Scale-Free Network

1. Start with $m_0$ nodes randomly connected

2. Add additional nodes where probability of connecting to each existing node is larger if that node has high degree

"The Rich Get Richer"
Scale-Free Metabolic Networks

Scale-Free Signaling and Transcription Networks

Protein interaction network
*H. sapiens*

Gene co-expression network
*S. cerevisiae*

*Stelzl et al. Cell, 122:957, 2005*

*Noort et al, EMBO Reports, 5:280, 2004*
Short Path Lengths in Scale-Free Metabolic Networks

The histogram of the path lengths in the *E. coli* metabolic network

The average path lengths for metabolic networks of 43 organisms with different complexity

letters to nature

Important consequence of nonlinear gravitational processes if the initial conditions are gaussian, and is a potentially powerful signa
ture to exploit in statistical tests of this class of models see Fig. 1.

The information needed to fully specify a non-gaussian field (or, in a wider context, the information needed to define an
image) resides in the complete set of Fourier phases. Unfortunately, relatively little is known about the behaviour of Fourier
phases in the nonlinear regime of gravitational clustering[6]-[7] -- it is essential to understand phase correlations in order to
design efficient statistical tools for the analysis of clustering data. A first step on the road to useful quantitative description of phase
information is to represent it visually. We do this using colour, as shown in Fig. 2. To view the phase coupling in an X body
simulation, we Fourier-transform the density field; this produces a complex array containing the real (R) and imaginary (I) parts of the
transformed image, with the pixels in this array shd by wavenumber k rather than position x. The phase for each wavenumber, given by
φ = atan2(I, R), is then represented in a box for that pixel.

The rich pattern of phase information revealed by this method (see Fig. 3) can be quantified, and related to the gravitational
dynamics of its origin. For example, in our analysis of phase coupling we introduced a quantity D(k):

\[ D(k) = \phi_{r} - \phi_{l} \]  

This quantity measures the difference in phase of modes with neighbouring wavenumbers in one dimension. We refer to D(k)
as the phase gradient. To apply this idea to a two-dimensional simulation, we simply calculate gradients in the u and v directions independently. Because the difference between two circular random variables is itself a circular random variable, the distribution of D(k) should initially be uniform. As the fluctuations evolve waves begin to collapse, spawing higher-frequence modes in phase with the original fluctuations. These interacting waves produce the non-
uniform distribution of D(k) seen in Fig. 3.

It is necessary to develop quantitative measures of phase information that can describe the structure displayed in the colour
representation. To begin, the phases are wavenumber and are the D(k) obtained from them. This corresponds to a state of
minimal information, or, in other words, minimal entropy. As information flows into the phases, the information content must
increase and the entropy decrease. This can be quantified by defining an information entropy for the set of phase gradients. We
construct a frequency distribution, D(k), of the values of D(k) obtained from the whole map. The entropy is then defined as

\[ S = -\int D(k) \log(D(k)) D(k) \]  

where the integral is taken over all values of D(k) that is, from 0 to 2π. The use of D(k) rather than ϕ, to define entropy is one way of
accounting for the lack of translation invariance of a problem that was missed in previous attempts to quantify phase
entropy[7]. A uniform distribution of D(k) is a state of maximum entropy (mini
imum information), corresponding to gaussian initial conditions (random phases). The minimal value of Smin = \log(2π) is a
characteristic of gaussian fields. As the system evolves, it moves into states of greater information content (that is, lower entropy).

The scaling of S with clustering growth displays interesting properties, establishing an important link between the spatial
pattern and the physical processes driving clustering growth. This phase information is a unique `imprint' of gravitational instabi
lity, and it therefore also furnishes statistical tests of the presence of any initial non-gaussianity[8]-[11].

References:


Publication date: 2000
Citation number: 8149

Error and attack tolerance of complex networks

Riko Albev, Haoang Jiang, and Albert-Laszlo Barabasi

Department of Physics, 322 Norris-Benedi Science Hall, University of Notre Dame, Notre Dame, Indiana 46556, USA

Many complex systems display a surprising degree of tolerance against errors. For example, relatively simple organisms possess
perseverence despite drastic pharmacological or environmental interventions, an error tolerance attributed to the robustness of
the underlying metabolic network. Complex communication networks display a surprising degree of robustness although
key components regularly malfunction, local failures rarely lead to the loss of the global information-carrying ability of the
network. The stability of these and other complex systems is often attributed to the redundant wiring of the functional
network defined by the systems' components. Here we demonstrate that error tolerance is not shared by all redundant systems:
if it is displayed only by a class of inherently sequenced networks,
Their Definition of Network Diameter

Shortest path from node 1 to node 6 is (1,5,4,6), path length=3

Do this calculation for all node pairs and sum

Divide by the number of node pairs to get diameter $d$
The Scale-Free Network is **More Resilient** to Random Failure Than the Exponential Network

![Graph showing the comparison between Scale-Free (SF) and Exponential (E) networks in terms of Diameter (d) vs. Failure fraction (f). The graph illustrates a linear increase in d for the Exponential network and almost no change for the Scale-Free network.]

- Linear increase in d
- Almost no change
But the Scale-Free Network is **More Vulnerable** to Attack Than the Exponential Network

Attack means sequentially taking out the nodes with highest degree
Giant Components

\[ S = \text{Fraction of nodes in the giant component} \]
\[ <s> = \text{mean size of a component} \]
In an Exponential Network There is **No Difference** Between Failure and Attack on Giant Component Size

$f_c$ is the threshold for network disintegration
In an Exponential Network the Network Disintegration is Roughly Homogenous
In an Exponential Network the Network Disintegration is
**Roughly Homogenous**
In a Scale-Free Network There is a **Big Difference** Between Failure and Attack on Giant Component Size

\[ f_c \text{ is the threshold for network disintegration} \]
In a Scale-Free Network the Network Disintegration in Response to Failure is **Very Heterogeneous**
End users are the computers and other devices we all use.

Interior vertices are routers, which are special-purpose computers at the junctions between data lines.
The World Wide Web is a Large Directed Network

**Internet**
Interconnected computers
Massive network of networks

**World Wide Web**
One service that runs on the internet
System of interlinked hypertext documents

- **Vertices** are the web pages
- **Directed Edges** are the hyperlinks
Degree Distribution of the Internet Obeys a Power Law for $k>1$

$\Pr[X = k] = Ck^{-2.48}$ in the year 2000
In-Degree and Out-Degree Distributions of the WWW Obey Power Laws

\[ \gamma_{in} = 2.1 \]
\[ \gamma_{out} = 2.45 \]

Exponent data from the year 2000
The Internet and WWW are Resilient to Random Failure, but Sensitive to Attack
In the Internet and WWW There is a **Big Difference**
Between Failure and Attack on Giant Component Size

$f_c$ is the threshold for network disintegration
Subway Systems are Undirected Networks

Nodes=stations
Edges=rail lines

Washington, DC metro
How Many Important Stations are There?

Hub?
Degree Centrality for Washington, DC System

hubs

Not a hub
Eigenvector Centrality for Washington, DC System

hubs

Now a hub
How About Another Subway System?
Degree Centrality for the Cleveland System

Not a hub

Vertices 1-15

Vertices 16-30

Vertices 31-45

Vertices 46-60

Vertices 61-70

Vertices 76-90

Vertices 91-100

Vertices 101-110
Eigenvector Centrality for the Cleveland System

Now a hub?

Vertices 1–15

Vertices 16–30

Vertices 31–45

Vertices 46–60

Vertices 61–70

Vertices 76–90

Vertices 91–100

Vertices 101–110