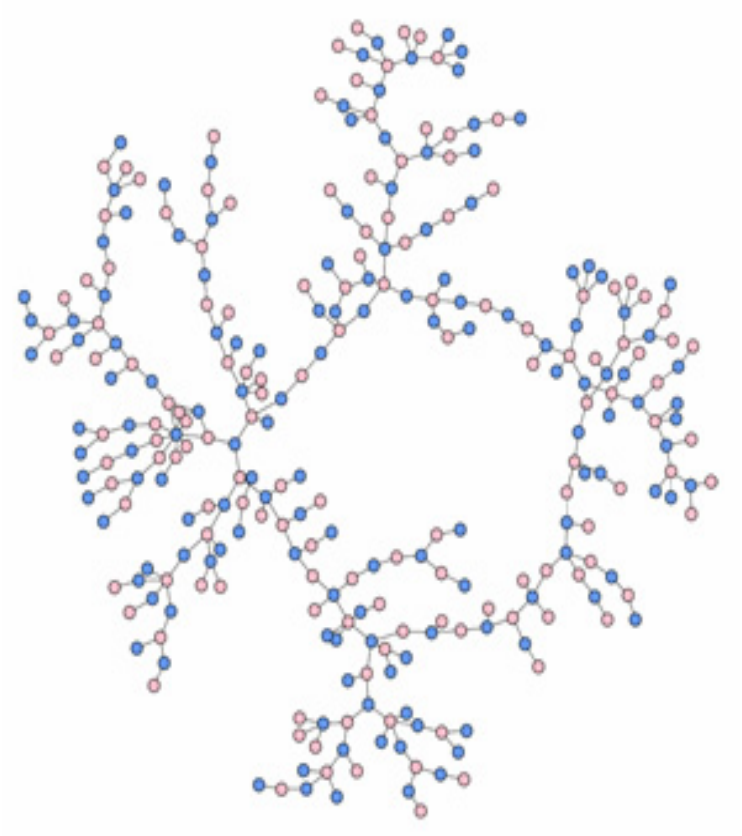
Complex networks: Models and dynamics

Pontus Svenson, Christian Mårtenson, Christian Carling



FOI-R-- 1766 --SE
ISSN 1650-1942

Command and Control Systems
Systems Technology
Methodology report

November 2005

FOI is an assignment-based authority under the Ministry of Defence. The core activities are research, method and technology development, as well as studies for the use of defence and security. The organization employs around 1350 people of whom around 950 are researchers. This makes FOI the largest research institute in Sweden. FOI provides its customers with leading expertise in a large number of fields such as security-policy studies and analyses in defence and security, assessment of different types of threats, systems for control and management of crises, protection against and management of hazardous substances, IT-security and the potential of new sensors.

FOI
Defence Research Agency
Command and Control Systems
P.O. Box 1165
SE-581 11 Linköping

Phone: +46 13 37 80 00
Fax: +46 13 37 81 00

www.foi.se

Complex networks: Models and dynamics

FOI-R--1766--SE
ISSN 1650-1942

Command and Control Systems
Systems Technology
Methodology report

November 2005

Issuing organization	Report number, ISRN	Report type
FOI – Swedish Defence Research Agency Command and Control Systems P.O. Box 1165 SE-581 11 Linköping Systems Technology SE-164 90 Stockholm	FOI-R--1766--SE	Methodology report
	Research area code 4. C4ISTAR	
	Month year November 2005	**Project no.** E 60707
	Sub area code 49 Interdisciplinary Projects regarding C4ISTAR	
	Sub area code 2	

Author/s (editor/s)	
Pontus Svenson Christian Mårtenson Christian Carling	**Project manager** Peter Krylstedt
	Approved by Martin Rantzer
	Sponsoring agency FM
	Scientifically and technically responsible Pontus Svenson

Report title

Complex networks: Models and dynamics

Abstract

This report is a brief introduction to complex networks. It describes the most important models for random, small world, and scale-free networks, and discusses their properties. Measures such as the clustering coefficient and the local efficiency are defined, and dynamical properties such as network searchability and synchronization are briefly discussed.

The report provides necessary background material for further studies on how the networks in a network-based defence should be designed in order to facilitate information spreading and synchronization.

Keywords

Complex networks, complex systems, random graphs, scale-free graphs, information spreading

Further bibliographic information	Language English
ISSN 1650-1942	**Pages** 26 p.
	Price acc. to pricelist

Utgivare
FOI - Totalförsvarets forskningsinstitut
Ledningssystem
Box 1165
581 11 Linköping

Systemteknik
164 90 Stockholm

Rapportnummer, ISRN
FOI-R--1766--SE

Klassificering
Metodrapport

Forskningsområde
4. Ledning, informationsteknik och sensorer

Månad, år
November 2005

Projektnummer
E 60707

Delområde
49 Breda projekt inom ledning, informationsteknik och sensorer

Delområde 2

Författare/redaktör
Pontus Svenson
Christian Mårtenson
Christian Carling

Projektledare
Peter Krylstedt

Godkänd av
Martin Rantzer

Uppdragsgivare/kundbeteckning
FM

Tekniskt och/eller vetenskapligt ansvarig
Pontus Svenson

Rapportens titel
Komplexa nätverk: modeller och dynamik

Sammanfattning
Rapporten ger en kort introduktion till området komplexa nätverk. De viktigaste modellerna för slumpgrafer, ”small world”-grafer och s k ”scale-free graphs” beskrivs. Olika mått, som klustringskoefficienten och den lokala effektiviteten, definieras och dynamiska egenskaper (sökbarhet, synkronisering) diskuteras kort.
Rapporten ger en nödvändig bakgrund för vidare studier om hur nätverk i NBF ska designas för att underlätta informationsspridning och synkronisering.

Nyckelord
Komplexa nätverk, komplexa system, slumpgrafer, scale-free graphs, informationsspridning, sökbarhet, synkronisering

Övriga bibliografiska uppgifter

Språk Engelska

ISSN 1650-1942

Antal sidor: 26 s.

Distribution enligt missiv

Pris: Enligt prislista

Chapter 1

Introduction

This report is intended to serve as a brief introduction to and survey of information spreading in complex networks. It is not intended to be a complete survey of the area, but is intended to provide enough information about the subject so that a researcher who is not familiar with the subject can be able to quickly learn some of the terminology and the most important models used. Since this is a very active research area, it is necessary for FOI to continue to keep track of developments in it.

Networks are gaining more and more attention both in the scientific and military communities. The number of papers published on the subject has risen dramatically, as is evidenced by the data shown in figure 1.1. Several good books and collections on complex networks have also recently been published [1, 2, 3, 4]

From a practical point of view, knowledge of complex networks is needed for the military for several reasons [5]. When designing the communication and command networks of future network-centric systems, we must ensure that they are robust against enemy attacks and component failures. The networks must also be designed so that information, orders, and service-requests are spread as fast as possible. Lastly, complex network knowledge is necessary in order to analyse the social networks [6, 7] of the opponents facing us in military operations other than war situations [8]. Figure 1.2 shows the social network of dating among students in an american high-school [9].[1] By analyzing the social network in various ways, many important properties of the subjects can be inferred [6, 7].

Future command and control systems will also need to be designed to facilitate the use of distributed fusion. There will be several local fusion nodes that gather and fuse data collected at different places in the network of platforms. This means that in addition to the robustness and communication requirements men-

[1] See also http://researchnews.osu.edu/archive/chainspix.htm

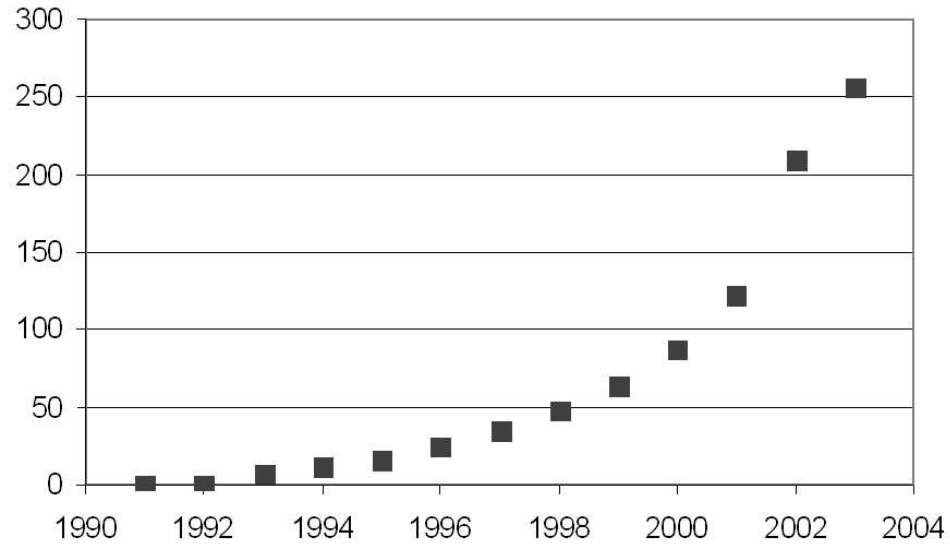


Figure 1.1: In recent years, the number of papers published on networks has grown tremendously. This figure shows the number of papers with "network" in the title appearing in the **cond-mat** section of **arxiv** (http://www.arxiv.org).

tioned above, the specific fusion algorithms chosen might also impose constraints on the network used.

In the report, we will distinguish between the physical and the logical communication networks. The physical network consists of the actual hardware that transmits information and will not be much discussed here. The logical communication network uses the physical to transmit information between sender and recipient.

The report is outlined in the following way. Chapter 2 gives a brief overview of various types of network models, while chapter 3 discusses some dynamic properties of networks, including search and information spreading on networks. Finally, section 4 contains some conclusions and a list of possible future work in this area.

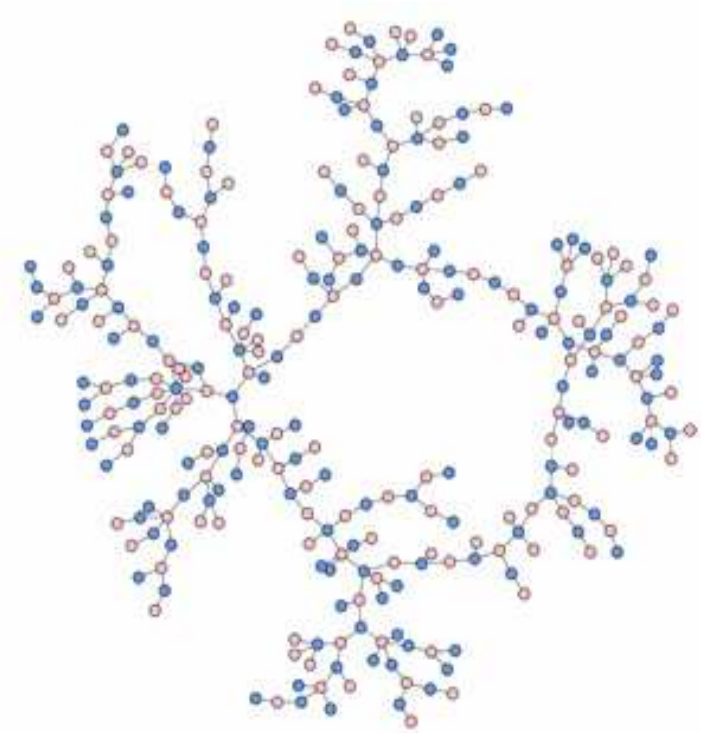


Figure 1.2: An example of a social network. The nodes correspond to students in a high-school. Two students are linked if they have been romantically involved with each other. Boys and girls are coloured differently. Can you determine which colour corresponds to which gender by analyzing the network?

Chapter 2

Network models

Parts of this chapter are based on chapter 5 of [10].

2.1 Simple networks

In order to properly describe a general network or graph, two things are needed. First, we need a list of the *nodes* of the graph. The nodes can be named and have various properties associated to them, but for describing the graph it is enough that they can be enumerated from 0 to $N-1$. Second, we must know which nodes are connected to which. This is most easily thought of as a list of *edges* (i,j) that are connected. Each edge can have various properties associated to it (*e.g.*, a weight w_{ij}). It is sometimes convenient to consider the graph as a function $\phi(i)$ that gives a list of the neighbours of node i.

Edges can be either directed (meaning, for example, that an edge (i,j) can only transmit information from i to j) or undirected. A graph is connected if there is a chain of edges connecting any pair of nodes in it. A natural generalisation of graphs is to replace the edges by triples i,j,k or even n-tuples. Such structures are called hypergraphs.

The simplest kinds of networks are regular, like the ones shown in figures 2.1 and 2.2.

All regular lattices have some features in common. By looking at the graphs in figure 2.1 it is for instance apparent that these graphs are clustered, in the sense that if we remove one node, its neighbours will still have a short path between them. Another interesting characteristic of regular lattices is that the average distance between nodes is quite large. For a lattice with N sites in D dimensions[1], it grows

[1]The simplest example to think of is $\mathcal{Z}_l^D$, where nodes are placed at integer coordinates and edges link nodes that whose coordinates differ by ± 1 in exactly one dimension. Choose $l = N^{1/D}$ to get

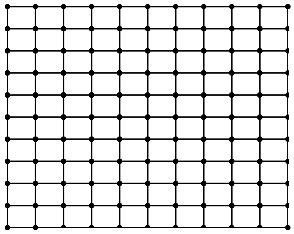


Figure 2.1: A square lattice.

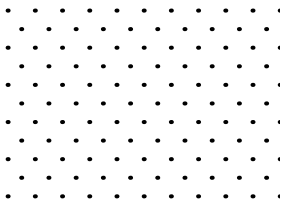


Figure 2.2: A triangular lattice

as $N^{1/D}$.

A natural extension of the regular lattice is to consider other graphs where all nodes are equivalent (i.e., have the same neighbourhood). The simplest example of such a graph is the complete graph with N nodes, K_N. This consists of N nodes where each node is connected to each of the other (so the graph has $\left(\begin{array}{c} N \\ 2 \end{array} \right)$ edges).

2.2 Classical random graphs

Traditionally, two different models of random graph processes have been used [11]. In the first model, $\mathcal{G}(N, p)$, each possible edge (i, j) is considered and included in the graph with a probability p. The other model, $\mathcal{G}(N, M)$, instead selects without replacement M of the $\left(\begin{array}{c} N \\ 2 \end{array} \right)$ possible edges. Note that these models are *not* completely equivalent. For the latter model, the graph is guaranteed to have exactly M edges, while the number of edges is a stochastic variable for the former. In the

N nodes.

thermodynamic limit, choosing

$$M = p \left(\begin{array}{c} N \\ 2 \end{array} \right)$$

gives graphs that should share all relevant properties. An important quantity characterizing different random graphs is their *connectivity* or *average degree* γ, which measures the average number of neighbours that the nodes have. For random graphs with N nodes and M edges, this is given by $\gamma = 2\frac{M}{N} = p(N-1)$ for the two ensembles.

Graph theory is a fascinating mathematical subject with many deep results; see for instance [11, 12]. One of the most interesting results is that there is a phase transition as the connectivity γ of a random graph grows. For small γ, the random graph consists of many isolated trees[2] of nodes. At $\gamma = 1$ this suddenly changes and a giant component emerges. The size of the giant component scales linearly with the number of nodes, N. This percolating transition is somewhat surprising — note that the graph can not be connected until it has a connectivity of at least $2(N-1)/N$. Another important result is that the average path-length between two nodes scales as $\log N$ for large N.

The random graph model, however, is not sufficient to describe many naturally occurring networks. Examples of networks that require more detailed models include the Internet, telephone networks, airline timetables, electric power grids, neural networks, interactions between monomers in a folded protein, metabolic pathways of biological organisms, electrical circuits, networks of sexual contacts and collaboration graphs of film actors, business people, and scientists (see [13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23] and references therein).

2.3 Small world graphs

There are many different kinds of networks in Nature. Perhaps the first that comes to mind is the social network of a society. Here each node represents a person, while there is an edge between two persons if they know each other. What does this graph look like? It is very unlikely that it would be a regular lattice — our acquaintances are not ordered in such a simple way. The social network however shares an important feature with regular lattices: they are clustered. Clustering means that there is a high probability that two neighbours of a given node also are direct neighbours themselves. An alternative way to think about it is to consider the average path length between two neighbours of a node i. Since both nodes

[2] A tree is a connected graph without loops.

are neighbours of i, this is obviously smaller than 2. If node i is now removed from the graph, we have to find a new shortest path between the nodes. If this new path length is still small, the graph is clustered. All regular lattices are obviously clustered, and social networks are clustered too: if person A knows persons B and C, there is a high probability that B and C will also know each other.

(Real social networks are clustered in several ways: everybody's acquaintances can be divided into several distinct clusters, *i.e.*, the people one knows from work all know each other, while the overlap between this group and one's neighbours often is zero. This can be modelled by allowing the network to have edges of different kinds, or by superimposing several different networks on top of each other.)

Mathematically, we can measure the degree of clustering in a graph by the *clustering coefficient*, C, defined as the average over all nodes of the local clustering coefficient C_i. For a given node i, consider its immediate neighbourhood, *i.e.*, the set of nodes to which it is linked. The *local clustering coefficient* is now given by the fraction of all possible edges between nodes in the neighbourhood that actually appear in the graph. Figure 2.3 shows an example that should make the definition clear.

Another important feature of social networks is the so called small world effect: When two strangers meet, it sometimes happens that the two people turn out to have mutual acquaintances.

The idea behind small world networks was first introduced by Milgram [24] in 1967. Milgram's experiment consisted of studying the path of letters addressed to a stockbroker in Pittsburgh. The letters were given to people in rural Nebraska with the rule that the current holder of the letter must hand it over to somebody with whom they were on a first-name basis. The average number of links in the chain of people between Nebraska and Pittsburgh was six, hence the term "Six degrees of separation". The number is of course not exact (a severe shortcoming of the experiment was that only one third of the letters were actually delivered!), but the phenomenon that people are linked via a small number of nodes has been verified by later, more careful experiments (e.g, [25]).

The small world effect has later been popularised by occurring in media, such as the movie "Six Degrees of Separation". There are also various amusing games using the same concept, such as the web site
`http://www.cs.virginia.edu/oracle/` where a user can find the distance between an arbitrary actor and Kevin Bacon. Actors here represent the nodes of the graph, and two actors are linked if they have participated in the same movie. It should be noted that the actors represented in the database are American and European ones. The network of actors in Indian movies, for instance, probably has few connections to this.

Another example are the Erdös numbers. Named after the famous mathemati-

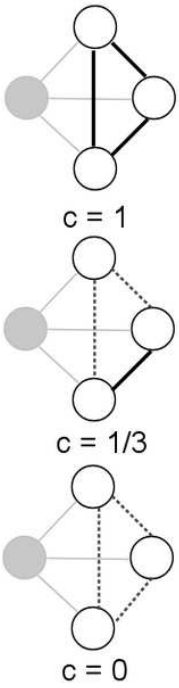


Figure 2.3: Examples of clustering coefficients. We wish to calculate the clustering coefficient of the grey node on the left, which has three neigbours (indicated by grey lines). The dark lines show the edges in the neighbourhood of the grey node that actually appear in the graph, while the dotted lines show all three possible such edges.

cian Paul Erdös [26], these are defined recursively: Erdös has Erdös number 0; a person has Erdös number $n + 1$ if they have co-authored a paper with somebody who has Erdös number n (there are at least 507 persons with Erdös number 1; see the web site
`http://www.oakland.edu/~grossman/erdoshp.html`).

Regular lattices do get shorter and shorter distances between nodes as the dimensionality increase (the diameter scales as $N^{1/d}$ for a d-dimensional lattice with N nodes, so it decreases if we increase d and keep N constant), but this is still too large to explain the small world effect. Instead, new graph models are needed.

A small world graph is intermediate between a regular lattice and a random graph — it has both clustering (like a regular lattice) and short maximum distances (like the random graph). It is constructed by considering in turn all the bonds (i, j) of a start graph (most often a regular lattice) and with some probability p replacing them with (i, k), where k is a new, randomly chosen, node. So by changing the rewiring probability p we can interpolate between the regular lattice and a random graph. An example of a small world obtained by rewiring the square lattice is shown in figure 2.4. Note that the small world for $p = 1$ differs slightly from a random graph, since all nodes are guaranteed to have a local connectivity of at least $\gamma/2$ where γ is the connectivity of the regular lattice. The distribution of connectivities is more broad for the small world with $p = 1$ than for the corresponding random graph.

The advance of the Internet and other communications networks has highlighted the need to be able to not only describe but also design networks that communicate efficiently. Efficiently here has two distinct meanings — the obvious one that a message from A to B should be transmitted along the shortest possible path, and also an equally important one that the network should be fail-safe. If a node suddenly disappears, it should be possible to quickly find alternate paths between the rest of the nodes that don't involve the dead node. A very clear definition of small world behaviour in terms of *efficiency* has been given by Latora and Marchiori [27]. They measure the local efficiency as the time needed to communicate in the network, assuming unit velocity of signal propagation. The efficiency between two nodes is thus

$$\epsilon_{ij} = \frac{1}{d_{ij}} \tag{2.1}$$

where d_{ij} is the shortest distance between nodes i and j and $d_{ij} = \infty$ if there is no path between the nodes. The global efficiency is the average of this over all pairs of nodes in the graph. A high global efficiency corresponds to a small diameter of the graph. The local efficiency for a node i is calculated as an average of ϵ_{ik} over all neighbours k of i, and the total local efficiency of the graph is then the average of this over all nodes. The local efficiency is a measure of the fault tolerance of the

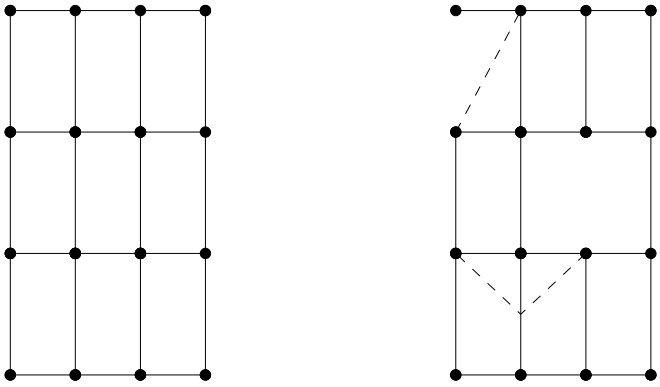


Figure 2.4: This figure shows the construction of a small world starting from a $2D$ square lattice (left). In the right figure, two edges have been rewired and are shown as dashed lines.

network.

In addition to efficiency and clustering, there are a large number of measures that can be used to characterize a graph's properties. Many of these measures come from sociology, and have been used to determine, *e.g.*, the influence and power of individuals in different social networks. Others come from computer science, or have been suggested by physicists. A very good survey of such measures is [28].

A small world graph still has the same poissonian distribution of node-connectivities as random graphs, whereas many graphs in Nature instead have a power-law distribution. Such graphs are called scale free and have attracted a lot of interest, see, *e.g.*, [29, 14, 30, 31], and below.

Much of the analytical work on small world networks has used a variation of the standard rewiring model [32]. In this model edges are never deleted from the graph, only added. This means that the mean connectivity of the graph will not be preserved.

The eigenvalues of the adjacency matrix J_{ij} (this is the symmetric graph whose entries at position i, j is non-zero and 1 iff i and j are neighbours) of both scale-free and small world graphs have been calculated by Farkas et al [33]. For random matrices, the spectral density of eigenvalues is known to converge to a semi-circle function, but they find completely different behaviour for realistic graphs. In particular, small-world networks turn out to have spectra that can not easily be de-

scribed by any function. This is yet another indication that it is not enough to put models of interacting agents on random graphs. A similar study has been made by Monasson [34]. It would be interesting future work to do the same for small worlds starting from d-dimensional regular lattices.

Exact calculations of many properties of these models are still missing; the analogue between the occurrence of the small world phenomenon for any rewiring probability $p > 0$ and a first order phase transition [35, 32] can perhaps lead to better result in the future.

Another alternative model has been introduced by Holme and Kim [36]. Here, instead of adding single nodes, triangles are added. This leads to a model that is both scale-free (*i.e.*, has power-law distribution of connectivities) and shows clustering. The most important advantage of this model is that it provides for an easy way to control the amount of clustering in the graph.

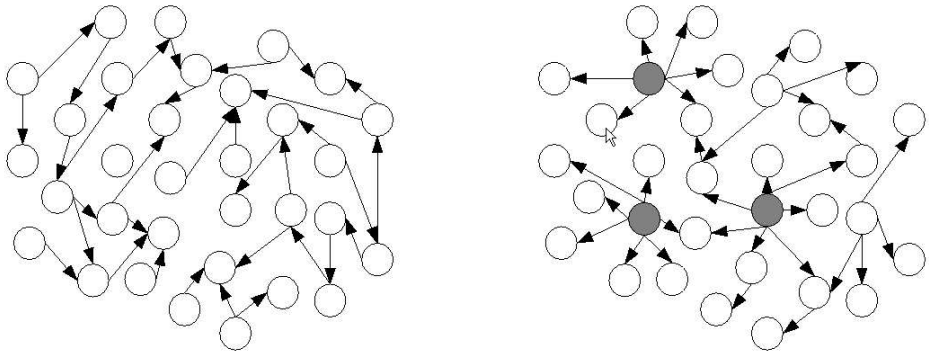
2.4 Scale free and growing graphs

A network is called scale free if there is no characteristic length scale in it. In contrast to lattices, whose characteristic length scale is the lattice spacing, a scale free graph divides its edges unequally among its nodes: the degree distribution follows a power-law. This means that there are a few nodes (called hubs) that have very many edges, whereas most of the nodes have very few. An important characteristic of scale free networks is that while they are robust against accidental failures, they are very vulnerable to deliberate attacks against hubs.

A deterministic model for generating scale free graphs has been introduced by Barabási and Ravasz [30]; this model generates the graph by iteratively replacing nodes with small graphs, in a manner similar to the construction of self-similar fractals.

There are many models of growing networks. In these models, one starts with a single node at time $t = 0$. In each new time-step a new node is added to the graph and a new edge is created that connects this node to one of the older ones with a probability that depends on the connectivity of that node. If this probability is simply proportional to the node's connectivity (k), the model is reduced to the scale-free graph model of Barabasi and Albert [37]; see also [29]. It has been shown that the case where the probability is proportional to the connectivity is the *only* case which also leads to a power-law for the distribution of connecitivities in the entire graph [38]. If the probability is proportional to k^γ for any $\gamma \neq 1$, we get stretched exponential (if $\gamma < 1$) distributions or graphs where the majority of the edges share a common central node (for $\gamma > 1$).

The best example of a growing network is the Internet — each node that is



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

Figure 2.5: This figure shows the difference between a random network and a scale free network. Note the presence of hubs (nodes with many neighbours) in the right network.

added could be interpreted either as a new computer that is connected to it or a new web-site that is created. The edges created between this node and the older ones are then the hyperlinks that the addition of a new site entails. [21] is a good easy introduction. It turns out, however, that a more complicated model is needed to model the Internet, see below.

So-called acquaintance networks, such as the Erdös graph, have been studied, among others, by Newman [31]. Such network have the characteristic that a small number of nodes have many edges. These nodes cannot be ignored when studying communication on such networks. Figure 2.6 shows an example of an acquaintance network: a collaboration network, where nodes are scientists and edges link scientists that have co-authored a paper.

The crucial point of Newman's new model is that the probability distribution of the number of neighbours, that the neighbours of a specified node has, is not independent of that node. In social networks, a node with very few neighbours is likely to be linked to other nodes that also have few neighbours; while a node with many neighbours have similar nodes among its neighbours.Clustering is important for this calculation. If we know both the degree distribution[3] and the clustering coefficient of a network, it is possible to calculate the number of neighbours at distance two from a given node. This is important to know when conducting research on social networks.

A very interesting approach to the problem of analysing growing networks

[3] *i.e.*, the probability that a node has a certain number of neighbours

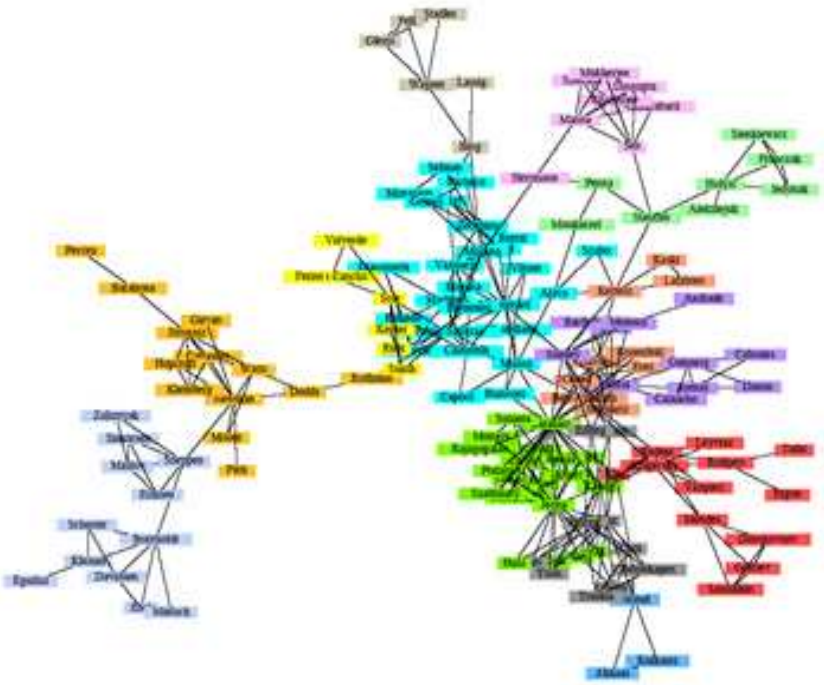


Figure 2.6: This figure shows the largest component of the network of collaboration among physicists who work in areas related to complex networks. (Figure from [39].)

has recently been introduced by Kleinberg and co-authors [40]. Perhaps the most prominent example of a growing network today is the Internet. The paper examines data representing routers on the Internet for a tvo year period, and finds that the link-density of the network increases with time, *i.e.*, that the number of edges $e(t)$ is related to the number of nodes $n(t)$ by a power-law

$$e(t) \propto n(t)^a, \tag{2.2}$$

with $a = 1.18$.

Having such a relation between $e(t)$ and $n(t)$ means that the connectivity $\gamma = \frac{e}{n}$ is time-dependent, in sharp contrast to most models of graphs and networks. The authors find a similar relation (but with different a's) also for three different kinds of citation networks. In addition to the super-linear scaling of edges with nodes, they also find that the average distance (also, somewhat non-standard, referred to as the effective diameter) between nodes *decreases* as a function of time. Recall that the average distance for a random graph grows as the logarithm of the number of nodes. While the two observations at first glance seem to be related, the authors note that it is possible to construct graphs that satisfy one of them but not the other. For large graphs, it is not practical to compute the diameter exactly. Instead, various approximate algorithms can be used. Since the shrinking diameter propery is very surprising, the authors check that their result is robust by using several different such approximations to calculate the diameter. They also check for errors due to the presence of a giant component; the shrinking diameter is present also if the calculation is restricted to just the giant component.

The paper presents several simple models that exhibit the densification property, and also give one model, the *Forest Fire model* that possessess both it and also displays shrinking effective diameters.

The Forest Fire starts from a graph with just one node and then adds one additional node at each time-step. At time t, let G_t be the current graph and v the added node. A node $w \in G_t$ is now selected randomly, and an edge $v \to w$ is formed. Most graph models would now continue by selecting another w and possibly adding an edge to it. In contrast, the Forest Fire model select a random number of the nodes in G_t that were linked to w, and adds edges from v to these nodes. This process is then repeated recursively for each of those nodes. (The process terminates if it reaches a node that has already been encountered. It is also possible to distinguish between out- and in-links when selecting the neighbours of w; see the paper [40] for details.) Intuitively, the graph is generated in a similar way as friendships are formed: a newcomer finds one friend and with a certain probability becomes friends also with the friend's friends, and so on. The name Forest Fire comes from a certain similarity to lattice cellular automata models used for study-

ing Forest Fires. A natural extension of the model is to select several starting points w at each time-step.

Chapter 3

Dynamical properties of networks

Dynamical properties are often more difficult to analyze than static. One of the first dynamical models that was studied for small world graphs is the density classification problem. Here a cellular automaton (CA) is given the problem of determining whether the majority of its inputs is 1 or 0. This problem was shown to be easier for CA placed on small worlds than on regular lattices [15, 41]. Neural networks [42] have been found to both have fast responses and possess coherent oscillation on small world graphs. No other network structure has been found that combines these two properties.

Rumour propagation is another example where the network plays a vital role, as is voter models of various kinds.

Disease spreading has been investigated in among others [43] and the position of the percolation [44] point has been found analytically for small world networks by Moore and Newman [45, 46]. It has been shown [47] that for a scale-free graph an epidemic will spread for an arbitrary small transmission probability. Dezsö and Albert have however shown that immunizing a sufficiently large number of nodes with high connectivities can lead to a finite effective epidemiological threshold [48].

Some work has been done on evolutionary models on small world graphs. Kulkarni and co-workers [49] have found that the nodes with the highest connectivities have the most evolutionary activity in the Bak-Sneppen model [50]. In most models of evolution, only point-mutations that change one gene are allowed (this corresponds to flipping one spin in a Monte Carlo simulation). Bagnoli and Bezzi [51] have studied the effect of allowing a small fraction of larger mutations, which has the effect of turning the hypercube of genotypes into a small world network. They find that by allowing a small number of such short-cuts the same results are obtained as when arbitrary mutations are allowed. This means that the fitness

landscape will be explored very quickly, but unfortunately also that the model lacks stability.

3.1 Network searchability

When analysing a network, an important feature to consider is its searchability, *i.e.*, how easy it is to search it for information. Depending on what information is available in the nodes and the search task at hand, different topologies entail using different search strategies.

For instance, if the network is stationary and known to you (through, *e.g.*, a map), finding the shortest path between two nodes is a standard procedure using an algorithm like Dijkstra's. On the other hand, for dynamically changing networks where nodes enter and leave at will, such complete information is difficult to obtain and maintain. In this case other methods have to be used in order to find the target node quickly.

Section 3.1.1 surveys some of the recent approaches dealing with network search based on local connectivity information. The success of the different methods relies heavily on network structure. The section also describes a method for determining network topology based on random walks.

A special category of networks are those that emerge between people in every day life. In these social networks nodes represent individuals and a link represents some kind of relation between two individuals, *e.g.*, business, family or sexual. People are in general very good at performing local search in social networks, which was illustrated in 1967 by Stanley Milgram in his famous small world experiment (see section 2.3). Methods for searching social networks are discussed in section 3.1.2.

3.1.1 Local search in model networks

Adamic et al. [52] show with theoretical calculations and simulations that efficient local search can be performed on a random graph with power-law connectivity. Local search in this case means that the only information available is the connectivities of a node's neighbours and its neighbours' neighbours. By directing the search to nodes with high degree neighbours it is possible to achieve a search scaling better than $O(N)$, where N is the number of network nodes. Although efficient on a power-law network, this high degree strategy is not always appropriate. On a graph where the connectivity follows a Poisson distribution (as is the case for the standard Erds-Renyi random graph model) it performed comparatively worse.

Yang [53] investigates how efficient different random-walk strategies are for

search on a number of different types of networks. The random-walk methods used include normal random walk (RW), no-back walk (NB), no-triangle-loop (NTL) walk, no-quadrangle-loop (NQL) walk, and self-avoiding (SA) walk. Networks studied are a classical random-graph (the Gilbert model), a small-world graph (the Watts-Strogartz model) and two scale-free networks (the Barabsi-Albert model and the deactivation model by Klemm and Eguiluz). It concludes that if the topology of the network is unknown the self-avoiding walk is the most efficient search strategy.

The search performance of the different walking strategies can be used to classify the topological features of a network. As is to be expected, on a clustered network (*i.e.*, one which has a lot of triangles and quadrangles) NTL and NQL will be more efficient than RW and NB. To reveal small-world characteristics SA can be used for probing as it performs sligthly better than NTL and NQL.

A different approach to network searchability is presented by Rosvall et al. [54]. The focus is shifted from finding the shortest path in the sense of the minimum number of links to the target node, to finding the path that requires the least information. The information cost for a node is defined as the number of yes/no questions necessary to determine which link is on the shortest path to the target node. The total search information between two nodes s and t is given by

$$S(s \to t) = -\log_2 \left(\sum_{\{p(s,t)\}} \frac{1}{k_s} \prod_{j \in p(s,t)} \frac{1}{k_j - 1} \right), \tag{3.1}$$

where k_s is the degree distribution of the startnode, k_j the degree distribution of node j on the current path and the sum is taken over the set $\{p(s,t)\}$ of degenerate shortest paths between s and t. Typically a shortest path will avoid highly connected nodes like hubs in a scale-free topology. To be able to characterize a complete network in terms of searchability the search information S is averaged over all shortest paths

$$S = \frac{1}{N(N-1)} \sum_{s \neq t} S(s \to t), \tag{3.2}$$

where N is the number of nodes in the network. The authors apply this measure to a number of models and real world network topologies and conclude that nature seems to "disfavour distant specific communication".

Another interesting concept defined in [54] is the *hide*, defined as the average search information needed to reach a target node from a random start node

$$\mathcal{H}_t = \sum_s S(s \to t)/N. \tag{3.3}$$

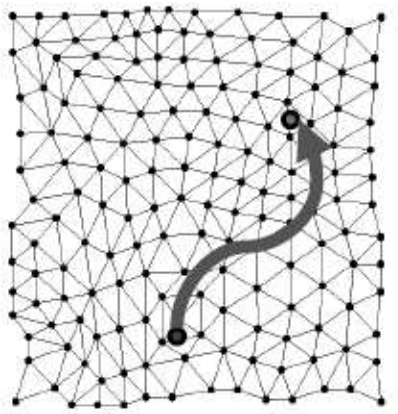


Figure 3.1: Illustration of a greedy search algorithm on a planar network.

The hide describes how visible the node t is to the rest of the network. An "important" node will have a small hide and the concept can be used to define a generalized hierarchy measure. It would be interesting to compare the hide to measures from social network analysis [6, 7] that attempt to measure the same thing.

3.1.2 Search in social networks

As mentioned above, social network often have small world characteristics.

Kleinberg [55] shows that finding short paths between nodes is easier in some small world networks than others. The model used is a simple square lattice, where each node is connected to its closest neighbours and also has one additional short-cut. The probability for the short-cut to lead to a randomly chosen node is proportional to $r^{-\alpha}$, where r is the lattice distance ("Manhattan") to the node. The possibility for a local greedy search algorithm to find short paths depends heavily on the value of α. Kleinberg reports that $\alpha = 2$ is optimal and that values above and below rapidly deteriorate the search.

A more realistic model for social interactions is presented by Watts et al. [56]. The network is a small world construct but based on the concept of cognitive hierarchies instead of a lattice. People tend to break down the world into layers, where the top layer represents the entire population and the bottom layer is a small closely related group or even a single individual. In between these extremes are groups of groups, hierarchically ordered. The distance between two nodes in such a network is equal to the lowest hierarchical level of a group comprising both nodes[1]. When

[1] In physics, this is referred to as an ultrametric distance.

a person searches the network it is possible to zoom out and target a larger group and then narrow it to pin point the desired node.

In their model, Watts et al. also utilize the fact that people in real life are not limited to a single type of social distance. For instance, guiding a search based on distance in physical location can be combined with distance in profession and/or hobby. Each type of social distance gives rise to a separate hierarchy and the combined distance between two nodes is set to equal the smallest of these. The authors investigate how the number of parallel hierarchies influence searchability and conclude that more than one can be useful but too many will decrease performance, similar to the confusion described by Rosvall et al. [54].

Interesting work has been done to compare the models of Kleinberg and Watts with real world networks. Adamic and Adar [57] derived a social network by analyzing email communication within their lab at HP. A social link was created for each pair of individuals that had exchanged at least 6 emails in both directions during a period of 3 months. Mass emails sent to more than 10 people were removed. These restrictions were introduced to try to catch an as realistic as possible network of aquaintances. Interestingly, these restrictions also transformed the degree distribution of the network from a power-law to something looking more like a Poisson form. As mentioned earlier, according to Adamic et al. [52] this would imply that a high degree search, *i.e.*, directing search to nodes that are the best connected, would not work efficiently, which also was confirmed. Two other search strategies were evaluated on the HP lab email network. The first is based on the observation that people closer in the organizational hierarchy email each other more often. The relationship between separation in distance and probability of correspondence matched Watts prediction of a searchable network (figure 3.1.2). The last search strategy used the nodes' physical location in the lab buildings to determine the social distance. The relationship between linking probability and distance r was shown to be $1/r$, different from the optimal $1/r^2$ calculated by Kleinberg. The results of the simulations corresponded well to this finding as well. A similar approach to model search in real world networks was conducted on a network constructed from an online networking site [57]. Unfortunately, the network was incomplete in many respects and good results were therefore hard to achieve.

3.2 Synchronization

How we achieve synchronization of intent among own units and coalition partners is an important and difficult question whose answer depends more on the organization and methodology used than on technical innovations and systems. Nevertheless, it is interesting to study how various technological system could help facilitate

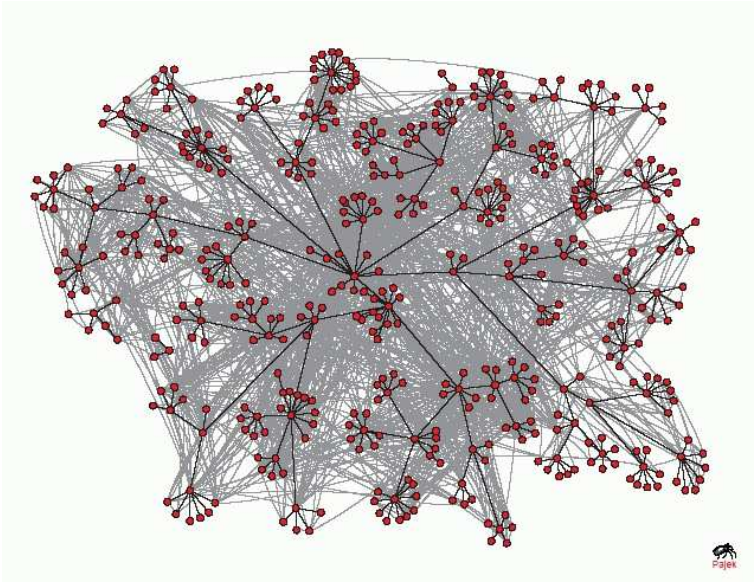


Figure 3.2: HP Labs email communication (light grey lines) mapped onto the organizational hierarchy (black lines). Note that communication tends to 'cling' to the formal organizational chart [57].

such synchronization. For example, how should the communication networks be constructed to improve the speed by which commander's intent and situation pictures are spread?

One way of answering this question is by studying simple models of fusion and command nodes on various networks. By using a simple model that can be simulated (or possibly even solved exactly) on different kinds of networks and study the differences in behaviour that arise, we can determine at least qualitatively what the differences between the networks are.

In order to use the results from such experiments for designing the military networks, it is of course necessary that the model is sufficiently similar to real fusion system.

One possible model that could be used for such experiments is the so-called *voter model* version of the Ising model.

The voter model consists of a number of agents that vote either "yes" or "no". How they chose to vote depends on how their neighbours vote. In the most simple version of the model, each agent selects one of its neighbours randomly and adjusts its vote to be the same. There are also versions with more interesting interactions.

Chapter 4

Conclusions and future work

Most suggestions for network-based defence system rely on a service-based architecture. In such systems, it is necessary to match the commander or analyst that is requesting a service with the platform or fusion node that can provide it. In order to do this as efficiently as possible, it is necessary to design the logical communication networks so that this type of communication is facilitated.

As shown in the chapters on information spreading and network search, different network topologies can have significantly different impact on the ease with which information is found in a network.

As stated in the introduction, it is important to know how to model networks for several different reasons [5]. We must be able to analyze the enemy's social and organizational structures as well as their communication networks. We must also be able to model the interaction network that will emerge when our commanders and operators communicate with each other. Not all of this communication will emanate from the hierarchical structure of the task-force: if two people know each other, they will most likely communicate (by phone, email, or instant messaging) even if they are not supposed to. Instead of banning such communication, the network-based defence system needs to exploit it and use it in order to achieve synchronization and fast information spreading. In order to do this as well as possible, it is necessary to model and simulate the emerging networks, using, *e.g.*, the models for search in social networks described in chapter 3

This report presented a brief survey of the field of complex networks and information spreading. As stated in the introduction, it is far from complete. We hope that it can serve as a starting point for researchers interested in further explorations of the area.

Bibliography

[1] S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, 2003.

[2] Romualdo Pastor-Satorras and Alessandro Vespignani. *Evolution and Structure of the Internet : A Statistical Physics Approach*. Cambridge University Press, 2004.

[3] Eli Ben-Naim, Hans Frauenfelder, and Zoltan Toroczkai, editors. *Complex Networks*, volume 650 of *Lecture Notes in Physics*. Springer, 2004.

[4] Romualdo Pastor-Satorras, Miguel Rubi, and Albert Diaz-Guilera, editors. *Statistical Mechanics of Complex Networks*, volume 625 of *Lecture Notes in Physics*. Springer, 2003.

[5] FMV. Teknisk prognos 2005. FMV dokument 57900/2005.

[6] John P Scott. *Social Network Analysis: A Handbook*. SAGE Publications, 2nd edition, 2000.

[7] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[8] John Arquilla and David F. Ronfeldt. *The Advent of Netwar*. RAND, 1996.

[9] Peter S Bearman, James Moody, and Katherine Stovel. Chains of affection. *American Journal of Sociology*, 110:44–91, 2004.

[10] Pontus Svenson. *A random walk in statistical physics*. PhD thesis, Chalmers University of Technology, 2001.

[11] Béla Bollobás. *Random Graphs*. Academic Press, New York, 1985.

[12] Béla Bollobás. *Graph Theory: An Introductory Course*. Springer-Verlag, New York, 1979.

[13] M. E. J. Newman. Models of the Small World: A Review. *J. Stat. Phys*, 101:819, 2000. e-print cond-mat/0001118.

[14] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distribution and their applications. *Phys. Rev. E*, 64:026118, 2001.

[15] D. J. Watts. *Small Worlds*. Princeton University Press, 1999.

[16] S.Jespersen, I. M. Sokolov, and A. Blumen. Small-world Rouse networks as models of cross-linked polymers. *J, Chem. Phys*, 113(17):7652, 2000.

[17] Mark Steyvers and Joshua B. Tenenbaum. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. Submitted to Cognitive Science.

[18] Brian Hayes. The Web of Words. *American Scientist*, March-April 1999.

[19] R. Ferrer i Cancho, C. Janssen, and R. V. Solé. The Topology of Technology Graphs: Small World Patterns in Electronic Circuits. Manuscript 2001.

[20] Sidney Redner. How Popular is Your Paper? An Empirical Study of the Citation Distribution. *Europ. J. Phys. B*, 4:131, 1998.

[21] Brian Hayes. Graph Theory in Practice, I and II. *American Scientist*, January-February and March-April 2000.

[22] R. Albert H. Jeong and A.-L. Barabasi. Internet - Diameter of the World-Wide Web. *Nature*, 401:130, 1999.

[23] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg. The web of human sexual contacts. *Nature*, 411:907, 2001.

[24] S. Milgram. The small world problem. *Psychology Today*, 2:60, 1967.

[25] C. Korte and S. Milgram. Acquaintance linking between white and negro populations: Application of the small world problem. *J. Personality and Social Psychology*, 15:101, 1970.

[26] P. Hoffman. *The Man Who Loved Only Numbers*. Hyperion, New York, 1998.

[27] Vito Latora and Massimo Marchiori. Effi cient Behaviour of Small-World Networks. *Phys. Rev. Lett.*, 87:198701, 2001.

[28] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. eprint cond-mat/0505185 (http://www.arxiv.org/abs/cond-mat/0505185), 2005.

[29] R. Albert and A-L. Barabási. Statistical Mechanics of Complex Networks. eprint cond-mat/0106096.

[30] A-L. Barabási and E. Ravasz. Deterministic Scale-Free Networks. eprint cond-mat/0107419.

[31] Mark E. J. Newman. Ego-centered networks and the ripple effect. eprint cond-mat/0111070.

[32] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Phys. Lett. A*, 263:341, 1999.

[33] E. J. Farkas, I. Derényi, A-L. Barabási, and T. Vicsek. Spectra of "Real-World" Graphs: Beyond the Semi-Circle Law. *Phys. Rev. E*, 64:026704, 2001.

[34] R. Monasson. Diffusion, localization and dispersion relations on "small-world" lattices. eprint cond-mat/9903347.

[35] M. A. de Menezes, C. F. Moukarzel, and T. . P Penna. First-order transition in small-world networks. *Europhys. Lett.*, 50:5, 2000.

[36] Petter Holme and Beom Jun Kim. Growing Scale-Free Networks with Tunable Clustering. eprint cond-mat/0110452.

[37] A. L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.

[38] P. L. Krapivsky, S. Redner, and F. Leyvraz. Connectivity of Growing Random Networks. *Phys. Rev. Lett.*, 85(21):4629, 2000.

[39] M E J Newman and M Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69:026113, 2004.

[40] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densifi cation laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 177, 2005.

[41] D. Watts and S. Strogatz. Collective dynamics of 'small world' networks. *Nature*, 393:440–2, 1998.

[42] L. F. Lago-Fernández, R. Huerta, F. Corbacho, and J. A. Sig¨uenza. Fast response and temporal coherent oscillations in small-world networks. *Phys. Rev. Lett.*, 84:2758, 2000.

[43] M. E. J. Newman and D. J. Watts. Scaling and percolation in the small-world network model. *Phys. Rev. E*, 60:7332, 1999.

[44] D Stauffer and A Aharony. *Introduction to Percolation Theory*. CRC Press, 1994.

[45] M. E. J. Newman and C. Moore. Epidemics and percolation in small-world networks. *Phys. Rev. E*, 61:5678, 2000.

[46] M. E. J. Newman and C. Moore. Exact solution of site and bond percolation on small-world networks. *Phys. Rev. E*, 62:7059, 2000.

[47] R. Pastor-Satorras and A. Vespignani. Epidemic Spreading in Scale-Free Networks. *Phys. Rev. Lett.*, 86:3200, 2001.

[48] Z. Denzs¨o and A-L. Barabási. Can we stop the AIDS epidemic? eprint cond-mat/0107420.

[49] R. V. Kulkarni, E. Almaas, and D. Stroud. Evolutionary dynamics in the Bak-Sneppen model on small-world networks. eprint cond-mat/9905066.

[50] K Bak and K Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters*, 24:4083, 1993.

[51] F. Bagnoli and M. Bezzi. Small world effect in evolution. eprint cond-mat/0007458.

[52] R. M. Lukose L. A. Adamic and B. A. Huberman. Local search in unstructured networks. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks: From the Genome to the Internet*. Wiley-VCH, Berlin, 2002.

[53] S-J. Yang. Exploring complex networks by walking on them. *Physical Review E*, 71, 2005.

[54] M. Rosvall, A. Gr¨onlund, P. Minnhagen, and K. Sneppen. Searchability of networks. arXiv:cond-mat/0505400 v1, May 2005.

[55] J. M. Kleinberg. Navigation in a small world. *Nature*, 406:845, 2000.

[56] P. S. Dodds D. J. Watts and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

[57] L. Adamic and E. Adar. How to search a social network. arXiv:cond-mat/0310120 v2, Oct 2004.