Universität Trier

HARNESSING COMPLEX STRUCTURES AND COLLECTIVE DYNAMICS IN LARGE NETWORKED COMPUTING SYSTEMS

Dem Fachbereich IV / Informatik
an der Universität Trier
zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
eingereichte

Dissertation

vorgelegt von

Diplom-Informatiker
Ingo Scholtes

geboren am 1. Februar 1980 in Trier

Dekan:	**Prof. Dr. Ralf Münnich**
Berichterstatter:	**Prof. Dr. Peter Sturm** University of Trier Department of Computer Science D-54286 Trier Germany
Mitberichterstatter:	**Prof. Dr. Tamás Vicsek** Eötvös Loránd University (ELTE) Department of Biological Physics H-1117 Budapest Hungary
Tag der Einreichung:	10. November 2010
Tag der mündlichen Prüfung:	27. Mai 2011

Acknowledgements

During the studies that led to this work, I was lucky enough to be surrounded by people who generously offered their support. My sincere thankfulness goes to my parents, my family and to Beate for their endless support, patience and their affectionate encouragement. I am also deeply grateful for the inspiring and encouraging attitude of my supervisor Peter Sturm and my referee Tamás Vicsek. Without them this work would not have been possible. Finally, I would like to thank my colleagues and friends Markus Esch, Jean Botev, Bernd Klasen, Daniel Görgen, Hermann Schloss and Katharina Anna Zweig for inspiring feedback and discussions on numerous occasions.
Despite the support provided by all these colleagues and friends, I feel obliged to say that I take full responsibility for all remaining errors and misconceptions.

Thank you very much!

Zürich, July 2012 *Ingo Scholtes*

To expect the unexpected shows a thoroughly
modern intellect.

An Ideal Husband
OSCAR WILDE

Kurzfassung

Wir leben in einer zunehmend vernetzten Welt, umgeben von einer Vielzahl immer größerer, miteinander verflochtener, technischer Systeme. Da diese immer mehr Bereiche unseres täglichen Lebens durchdringen, gewinnt ein genaues Verständnis ihrer Struktur und Dynamik einen immer größeren Stellenwert. Gleichzeitig entstehen viele der heute größten technischen Infrastrukturen wie zum Beispiel das globale Router-Netzwerk des Internets, das World Wide Web, große Peer-to-Peer Systeme oder globale Stromnetze nicht "am Reißbrett" sondern entwickeln sich - von vielerlei Bedingungen und Wechselwirkungen beeinflusst - verteilt und oftmals außerhalb des Einflussbereichs einer zentralen Kontrollinstanz. Für heutige und zukünftige, immer größer werdende, vernetzte Rechnersysteme wird es aufgrund dieser zunehmenden Komplexität immer schwieriger belastbare Aussagen über deren Struktur und Verhalten zu treffen.

Eine Reihe von Ausfällen und Störungen hat hierbei insbesondere gezeigt, dass in großen, vernetzten Systemen ungewollt komplexe Strukturen entstehen können, welche jenen die zum Beispiel in physikalischen, biologischen und sozialen Systemen auftreten nicht unähnlich sind. In dieser Dissertation gehen wir der Frage nach, wie solche, in großen vernetzten Rechnersystemen auftretende *komplexe Phänomene* beschrieben, beherrscht und aktiv genutzt werden können. Hierzu betrachten wir zunächst methodische Ansätze aus der Erforschung zufälliger und komplexer Netzwerke, welche seit ungefähr einem Jahrzehnt verstärkt zur empirischen Untersuchung natürlicher, sozialer und technischer Systeme verwendet werden und wertvolle Einblicke in deren Struktur und Dynamik geliefert haben. Einen besonderen Stellenwert nimmt hier die Beobachtung ein, dass die oftmals bemerkenswerte Effizienz, Zuverlässigkeit und Anpassungsfähigkeit vieler natürlicher Systeme auf verhältnismäßig einfache, lokale und oftmals randomisierte Interaktionen zwischen einer großen Zahl von Elementen zurückgeführt werden kann. Wir fassen zunächst einige interessante Ergebnisse über die Entstehung komplexer Netzstrukturen und kollektiver Dynamik zusammen und untersuchen, inwiefern diese in konstruktiver Weise beim Entwurf und Betrieb großer vernetzter Rechnersysteme nutzbar sind.

Ein besonderes Augenmerk dieser Dissertation richtet sich auf die Anwendung von Prinzipien, Methoden und Resultaten aus dem Umfeld komplexer Netzwerke im Kontext verteilter Systeme, welche auf sogenannten Overlay-Netzwerken beruhen. Die Tatsache, dass die (virtuelle) Konnektivität in solchen Systemen weitgehend unabhängig von physikalischen Beschränkungen dynamisch angepasst werden kann, ermöglicht

beispielsweise eine Konstruktion auf der Grundlage von Analogien zwischen komplexen Netzstrukturen und Phänomenen der statistischen Physik.

Aufbauend auf theoretischen Erkenntnissen über die Eigenschaften sogenannter skalenfreier Netze, stellen wir ein einfaches und effizientes Verbindungsprotokoll vor, mit dessen Hilfe randomisierte, skalenfreie Overlaynetze mit anpassbarem Knotengradexponenten auf verteilte Art und Weise erzeugt werden können. In diesem Zusammenhang demonstrieren wir, dass auf Grundlage des vorgestellten Protokolls Phasenübergangsphänomene, wie sie häufig im Umfeld der statistischen Physik auftreten, aktiv zur Anpassung makroskopischer, statistischer Netzparameter genutzt werden können, welche die Stabilität und Leistung vernetzter Systeme massiv beeinflussen. Im konkreten Fall erlaubt die Anpassung des Exponenten der Knotengradverteilung zufälliger skalenfreier Overlays in bestimmten kritischen Bereichen eine schnelle Veränderung relevanter struktureller sowie dynamischer Eigenschaften. Insofern ermöglicht es das untersuchte Protokoll, belastbare Aussagen über die Beziehung zwíschen dem mikroskopischen, lokalen Verhalten einzelner Netzknoten sowie den makroskopischen Eigenschaften resultierender, komplexer Netzstrukturen zu treffen. Für Systeme in denen der Knotengradexponent nicht auf einfache Art und Weise auf Grundlage lokaler Protokollparameter ableitbar ist wird zudem ein verteiltes, probabilistisches Verfahren vorgestellt, welches eine Überwachung des Exponenten und damit die Ableitung wichtiger Netzeigenschaften erlaubt.

Schließlich wendet sich die Dissertation der Untersuchung komplexer, nichtlinearer Dynamik in vernetzten Rechnersystemen zu. Wir betrachten ein einfaches, nachrichtenbasiertes Protokoll, welches auf Basis des Kuramoto-Modells für gekoppelte Oszillatoren eine stabile, globale Synchronisation periodischer Ereignisse in verteilter, selbstorganisierter Art und Weise erreicht. Die Effizienz und Stabilität des vorgestellten Verfahrens wird in einer Reihe von Netztopologien untersucht. Darüberhinaus deuten wir an, dass das vorgestellte Verfahren - basierend auf theoretischen und empirischen Erkenntnissen über die Wechselwirkungen zwischen spektralen Eigenschafter komplexer Netze und der Synchronisationsdynamik gekoppelter Oszillatoren - eine verteilte Überwachung topologischer Eigenschaften vernetzter Rechnersysteme erlaubt.

Ein wichtiger Aspekt der vorliegenden Dissertation ist die Betrachtung eines interdisziplinären Ansatzes zur Schaffung eines bewussten und konstruktiven Umgangs mit komplexen Strukturen und kollektiver Dynamik in großen, vernetzten Rechnersystemen. Die damit verknüpfte Untersuchung verteilter Systeme aus der Perspektive der nicht-linearen Dynamik, der statistischen Mechanik sowie der theoretischen Biologie zeigt interessante Parallelen zu physikalischen, biologischen und sozialen Systemen auf. Dies verspricht die Schaffung selbstorganisierender, vernetzter Rechnersysteme, die *wachsen* und *reagieren* statt konstruiert und gesteuert zu werden und deren Struktur und Verhalten mittels Methoden der komplexen Systeme und der statistischen Physik modelliert, analysiert und verstanden werden kann.

Abstract

We are living in an increasingly connected world, surrounded by a multitude of interwoven technical systems constantly growing in size. Since they pervade more and more aspects of our everyday lives, a thorough understanding of the structure and dynamics of these systems is becoming increasingly important. However - rather than being blueprinted and constructed at the drawing board - many of today's largest technical infrastructures like for example the Internet's global router network, the World Wide Web, large scale Peer-to-Peer systems or the power grid - evolve in a distributed fashion, often beyond the control of a central controlling instance and influenced by various surrounding conditions and interdependencies. Hence, due to this increase in complexity, making substantiated statements about the structure and behavior of today's and tomorrow's networked systems is becoming increasingly complicated.

A number of failures and disruptions has furthermore shown that in a number of large networked systems complex structures can emerge unintentionally that resemble those which can be observed for example in in biological, physical and social systems. In this dissertation, we investigate the question how such *complex phenomena* arising in large networked computing systems can be modeled, controlled and actively used. For this, we first review methodologies stemming from the field of random and complex networks, which are since roughly a decade increasingly being used for the empirical study of natural, social and technical systems, thus delivering valuable insights into their structure and dynamics. A particularly interesting finding is the fact that the sometimes remarkable efficiency, dependability and adaptivity of many natural systems can be related to rather simple local interactions between a large number of elements. We review a number of interesting findings about the formation of complex structures and collective dynamics and investigate how these are applicable in a constructive way in the design and operation of large scale networked computing systems.

A particular focus of this dissertation will be laid upon potential applications of principles, methods and results stemming from the study of complex networks in distributed computing systems that are based on so-called overlay networks. Here we argue how the fact that the (virtual) connectivity in such systems is alterable and widely independent from physical limitations facilitates a design that is based for example on

analogies between complex network structures and phenomena studied in statistical physics.

Based on theoretical results about the properties of so-called scale-free networks, we present a simple and efficient membership protocol by which randomized, scale-free overlay networks with adjustable degree distribution exponent can be created in a distributed fashion. With this protocol we further exemplify how phase transition phenomena - as occurring frequently in the domain of statistical physics - can actively be used to quickly adapt macroscopic statistical network parameters which are known to massively influence the stability and performance of networked systems. In the case considered in this dissertation, the adaptation of the degree distribution exponent of a random, scale-free overlay allows - within certain critical regions - a change of relevant structural and dynamical properties. As such, the proposed scheme allows to make sound statements about the relation between the local behavior of individual nodes and large scale properties of the resulting complex network structures. For systems in which the degree distribution exponent cannot easily be derived for example from local protocol parameters, we further present a distributed, probabilistic mechanism which allows the monitoring of a network's degree distribution exponent and thus a reasoning about important structural qualities.

Finally, the dissertation shifts its focus from the investigation of network structures towards the study of complex, non-linear dynamics in networked computing systems. We consider a simple, message-based protocol which - based on the Kuramoto model for coupled oscillators - achieves a stable, global synchronization of periodic heartbeat events in a distributed fashion. The protocol's performance and stability is then evaluated in different network topologies. We further argue that - based on existing theoretical and empirical findings about the interrelation between network structures and the synchronization dynamics of coupled oscillators - the proposed protocol allows a distributed monitoring of the structural and modular properties of networked computing systems.

An important aspect of this dissertation is the consideration of an interdisciplinary approach towards a sensible and constructive handling of complex structures and collective dynamics in large scale networked computing systems. The associated investigation of distributed systems from the perspective of non-linear dynamics, statistical physics and theoretical biology highlights interesting parallels both to biological and physical systems. This promises the construction of self-organizing networked systems which *grow* and *react* rather than being constructed and controlled and whose structures and dynamics can be modeled, analyzed and understood in the conceptual frameworks of statistical physics and complex systems.

Contents

Chapter 1

Motivation

> ... and as far as it is possible to do so, these two wonderful vessels are designed to be unsinkable.
>
> *Publicity Brochure of White Star Line on the Olympic and the Titanic, 1910*

History of Distributed Systems

Since the development of the packet-switched *Arpanet* in the late 1960s, a triumphal course of distributed computing technologies has taken place. The Arpanet starting in 1969 with no more than four nodes, 40 years later its successor, the Internet, currently interconnects an estimated number of 1.5 billion computing devices and creates the basis for distributed computing systems of unprecedented scale. Today most people naturally possess and use a multiplicity of interconnected devices like desktop, laptop or tablet computers, smart phones, mobile music players, video game consoles, electronic readers, digital picture frames or network-connected television equipment. Given the popularity and ubiquity of applications like the World Wide Web (WWW), online communities, distributed virtual environments, electronic commerce, Internet video telephony and electronic mail, one can hardly imagine a modern society without these technological achievements.

Scale of Current Systems

Many distributed systems that we are using routinely today rely on thousands or even millions of communicating machines. The routing infrastructure of the Internet is comprised of several hundred thousand routers, the Domain Name System of the Internet is based on tens of millions of servers, each of the data centers behind popular web-based services like Google, Amazon, YouTube, eBay or Facebook contains tens of thousands of machines and Peer-to-Peer-based services like SKYPE interconnect the machines of tens of millions of active users.

Challenges in Massive-Scale Distributed Systems

The design, operation and management of networked computing systems at such massive scales is a challenging task. Due to the large number of devices, errors and failures of individual components are rather ubiquitous than exceptional and must be dealt with in an efficient manner that does not influence the functioning of the system as a whole. Explicit algorithmic interactions as well as implicit interdependencies between individual machines can result in correlated failures that threaten a system's dependability. As these systems run continuously over long periods of time, the occurrence of exceptional situations - which are unlikely to occur on short time scales - are becoming more probable. The operation conditions often depend on human, social or environmental factors that are hardly predictable and cannot be tested exhaustively in advance. At the same time, short-term modifications to adapt to unforeseen situations are rendered impossible by laborious and expensive deployment routines. Finally - due to complex and possibly subtle interactions - even carefully designed systems can be prone to complex emergent phenomena that are hardly predictable. Experiences with today's largest distributed systems underpin these concerns as some of them have exhibited systemic failures that can be attributed to one or more of the aforementioned aspects. Prominent examples for such failures can increasingly be found in networked computing but also in other systems consisting of a sufficiently large number of interacting components like e.g. power grids, ecological and - being a recent issue - economic systems.

Future of Distributed Systems

All these problems are likely to worsen in future, as distributed computing systems are becoming larger, more dynamic and complex. Among the main driving forces fueling this development are the ongoing miniaturization and price decline of networked devices as well as the increasing incorporation of Internet connectivity into "smart" everyday objects. In addition to this trend towards the so-called "Internet of things", the increasing spread of Internet-connected devices in emerging and developing countries is likely to further the scale of prevalent and future networked computing systems. Besides their mere size, the dynamics and degrees of freedom in connectivity are constantly increasing. While early distributed systems were typically comprised of few, rather statically connected components, the development of flexible, light-weight and universally applicable protocols and associated middleware support foreshadows a world in which billions of heterogeneous devices can potentially interact with each other across the Internet [Scholtes and Sturm, 2006].

Challenges in Future Systems

The increase in scale and dynamics of these systems gives rise to unprecedented complexities. At the same time, the relevance of massively distributed systems for everyday's life is continuously increasing. Even today, the Internet and the WWW have become critical infrastructures whose reliability is crucial for the functioning of many aspects of our society. A precondition for the use of massively distributed systems in practice is the ability to make credible statements about their performance, behavior and reliability. Considering the challenges and problems motivated above and their importance for current and future networked systems, the question arises how these goals can be met at a very large, possibly global scale.

Decentralized Systems in Nature

In the history of engineering, nature has been a frequent source of inspiration for the design and improvement of a variety of systems. Large distributed computing systems with their elaborate protocols and intricate technical details do not seem like a domain predestined for applying concepts from natural systems. For individual computing devices this may be true but things seemingly change when addressing large numbers of interacting machines. Looking at nature, we can identify a variety of systems whose functioning as a whole is based on large numbers of interacting, possibly error-prone elements. A single human body is composed of an estimated number of 10 trillion cells of more than 200 different types. Flocks and colonies of birds or insects can grow to millions of individual animals and yet they show signs of globally coherent, organized behavior despite lacking central coordination. Constituting an example from the inanimate world, a magnet consists of an intractable number of atoms and yet the spins of their electrons can order spontaneously and thus lead to effects that are observable on a macroscopic scale.

Complex Structures and Dynamics in Natural Systems

During the last decades, the mechanisms underlying such large scale systems in nature have received a lot of attention. Research in disciplines like mathematics, physics, biology and sociology have unveiled principles underlying their evolution as well as their sometimes surprising robustness and adaptiveness. Related strands of research of these different scientific disciplines are commonly summarized by the terms *complex systems science*, *complexity theory*[1] and - as such complex systems are increasingly being viewed in the conceptual framework of networks - *complex network science*. The interdisciplinary character of this emerging field has contributed significantly to a cross-pollination of research. In this study it has been uncovered that some of the phenomena originally studied in models for natural systems can also occur inadvertently in sufficiently large technical systems, in some cases leading to

[1] This is not to be confused with the field of *computational complexity theory* in computer science.

detrimental effects or even catastrophic failures. Prior to formulating the theses underlying this dissertation, in the following we review some prominent examples for the occurrence of such problems in practice.

1.1 Complex Phenomena in Distributed Systems

SKYPE's "Perfect Storm"

The morning of August 16th, 2007 gave an unpleasant surprise to millions of people across the globe. The Internet's most popular voice and video communication service SKYPE fell silent as any login attempt of the majority of its then 220 million users remained unsuccessful. Due to its massive popularity and importance for the business world, the incident arouse immediate interest in media across the world, casting a damning light on a service that by so many had been viewed as essential infrastructure. In its issue of August 17th, the New York Times wrote:

> *"The online telephone service* SKYPE *was not working for much of the day on Thursday, leaving its 220 million users, some of them small businesses that had given up their landlines, without a way to call colleagues, customers and friends. [...] "There is a chance this could go on beyond tomorrow, but it's our hope that it's going to be resolved," Kurt Sauer,* SKYPE*'s chief security officer, said. "What happened today was caused by a unique set of events, the genesis of which is not entirely understood."*[2]

Reaction to SKYPE *Outage*

Although the SKYPE engineers did take immediate action, the until then so reliable service could not be fully restored until August 18th. To many users and computer science professionals, the magnitude of this outage was all the more surprising as the SKYPE service is based on a Peer-to-Peer technology, whose decentralized architecture had so far been seen as a warrant for reliability and scalability. On August 20th, two days after the service had been restored, the problem was identified and communicated to the public in the following statement that was published by Villu Arak on the SKYPE Blog:

> *"On Thursday, 16th August 2007, the* SKYPE *peer-to-peer network became unstable and suffered a critical disruption. The disruption was triggered by a massive restart of our users' computers across the globe within a very short time frame as they re-booted after receiving a routine set of patches through Windows Update. The high number of restarts*

[2] see `http://www.nytimes.com/2007/08/17/business/17ebay.html`, accessed on October 5th, 2010

> *affected* SKYPE*'s network resources. This caused a flood of log-in requests, which, combined with the lack of peer-to-peer network resources, prompted a chain reaction that had a critical impact."*[3]

Aftermath of SKYPE *Disruption*

Since no further technical details on the incident were disclosed, one can only speculate about the actual reasons behind the outage. Updates to Windows operating systems are routinely applied on a fixed day and no similar problems had been observed on prior occasions. Disclaiming any responsibility, Microsoft engineers immediately affirmed that there was nothing special about either amount, type or timing of the updates applied on the day in question. Due to this and the fact that the outage occurred almost two days after updates had been distributed, speculations were running wild with alternative explanations. At least those conjecturing that the massive reboots were not the exclusive reason for the capital crash were later affirmed by a statement published by SKYPE. Comparing the situation to a "perfect storm", it was stated that the outage occurred because a

> *"[...] combination of factors created a situation where the self-healing needed outside intervention and assistance by our engineers."*[4]

Recent Evidence of the SKYPE *Outage*

Based on recently published external measurements of SKYPE traffic that were performed during the downtime [Rossi *et al.*, 2009], there is some evidence that dynamic interactions of SKYPE's adaptation scheme, network traffic as well as the removal and overload of machines caused a cascading effect which eventually brought down the whole system. In the aftermath of the SKYPE outage, commentaries were dealing with the consequences for the trust in a service that had been viewed by many as essential and reliable infrastructure. On August 20th, 2007 "Business Week" writer Stephen Baker commented:

> *"Time was, our phones in the United States were nearly as reliable as AAA bonds. They were engineered to be mission critical — and long distance calls cost an arm and a leg. Now we have nearly free long distance calling on* SKYPE*. But it's nowhere near mission critical."*[5]

Although the reference to the reliability of AAA-rated financial products seems rather grotesque in the light of the then imminent meltdown of the financial markets, the comment shows that a crucial question is entering public

[3] see `http://heartbeat.skype.com/2007/08/what_happened_on_august_16.html`, accessed on October 5th, 2010

[4] see `http://heartbeat.skype.com/2007/08/the_microsoft_connection_explained.html`, accessed on October 5th, 2010

[5] see `http://www.businessweek.com/the_thread/blogspotting/archives/2007/08/skype_aftermath.html`, accessed on October 5th, 2010

Dependability of Distributed Infrastructures

awareness: How dependable are massively distributed Internet-based services and how much do we depend on them? The consequences of an outage of a few popular services have proven to be grave today and are likely to worsen in future. Since then, SKYPE has further strengthened its position, becoming one of the most popular Internet services. As of April 2010, SKYPE reportedly had more than 560 million registered users across the globe[6], with at times more than 23 million of them being online concurrently[7]. A small but increasing fraction of users is even dropping landlines and relying on SKYPE as primary telephony service. The events of August 2007 have demonstrated that a thorough understanding of network dynamics is crucial, shall massively-distributed Internet-based services one day replace vital infrastructures like for example the telephone system. The current trend towards using Internet services like SKYPE on smartphones and other mobile devices with inherently unreliable connectivity is likely to pose additional challenges to the provision of dependable Peer-to-Peer based services.

GMail Outage of 2009

Roughly 18 months after SKYPE's "perfect storm", the provider of some of the today most popular Internet services experienced a similar glitch. On February 24th, 2009, Google's E-Mail service GMail was globally unavailable for roughly two and a half hours, leaving a majority of its then 100 million users without access to their mailboxes. The potential impact of such a failure becomes apparent when considering that GMail's corporate mail service provides the basis for the electronic communication of numerous companies. According to an entry that was published on the day of the incident on the GMail Blog, the reason for the outage was the introduction of a new load balancing scheme in one of the company's data centers.

> *"This morning, there was a routine maintenance event in one of our European data centers. This typically causes no disruption because accounts are simply served out of another data center. Unexpected side effects of some new code that tries to keep data geographically close to its owner caused another data center in Europe to become overloaded, and that caused cascading problems from one data center to another. It took us about an hour to get it all back under control."*[8]

[6] see `http://gigaom.com/2010/04/20/skype-q4-2009-number/`, accessed on October 5th, 2010

[7] According to own observation on October 5th, 2010

[8] see `http://gmailblog.blogspot.com/2009/02/update-on-todays-gmail-outage.html`, accessed on October 5th, 2010

Internet Services vs. Public Utility Infrastructures

The "cascading problems" in this statement resemble - though informally - the "chain reaction" that had previously caused the disruption of the SKYPE service. In contrast to SKYPE's decentralized Peer-to-Peer-based infrastructure, GMail is provided by a cluster of data centers that are centrally managed by Google. The incident thus demonstrates that this kind of issue is not a specialty of Peer-to-Peer-based services like SKYPE that inherently suffer from unreliable resources and constant topology changes. Based on these incidents, one might argue that - as their importance for society is steadily increasing - future Internet-based services need to be engineered and controlled in a way that makes them as reliable as today's public utility infrastructures. Unfortunately, as their complexity is steadily increasing, a closer inspection of these critical infrastructures reveals examples for very similar systemic failures.

Power Failure in North-Eastern America

On August 14th, 2003, public life in most of northeastern America came to a sudden rest as one of the largest blackouts in history hit the region. The shutdown of a single nuclear power station in Eastlake, Ohio had triggered a series of events that eventually left more than 55 million people across the northeastern US and Canada without electricity for an average of 18 hours. The loss of this power station initially resulted in a slow but progressive failure of numerous high voltage transmission lines in north-eastern Ohio. When two and a half hours later the line trips had reached southeastern Michigan, a high-speed cascade of failures and power plant shutdowns was set off that propagated throughout the neighboring states and the Canadian province of Ontario within less than 60 seconds. The resulting largest power outage in the history of the United States eventually led to the shutdown of generating units in a total of 265 power plants, brought both land and air traffic to a temporary halt and caused the death of at least 11 people. The scale of this incident was so massive that it could even be observed from space. Figure 1.1 shows the illumination of north-eastern America on a normal night as opposed to the night of the power failure, as photographed by a satellite of the National Oceanic and Atmospheric Administration (NOAA).

Power Failure in Central Europe

Immediately after the events in the United States and Canada, a number of German electricity suppliers reassured the public that the occurrence of a similar power failure in Europe or Germany would be extremely unlikely[9]. They were proven wrong roughly three years later when on November 4th, 2006 an incident of similar magnitude hit Europe. A rapid cascade of power line

[9] see `http://www.strom-magazin.de/strommarkt/stromausfall-chaos-in-amerika-zusammenbruch-der-hochspannungsnetze-in-deutschland-unwahrscheinlich_9792.html`, accessed on October 5th, 2010

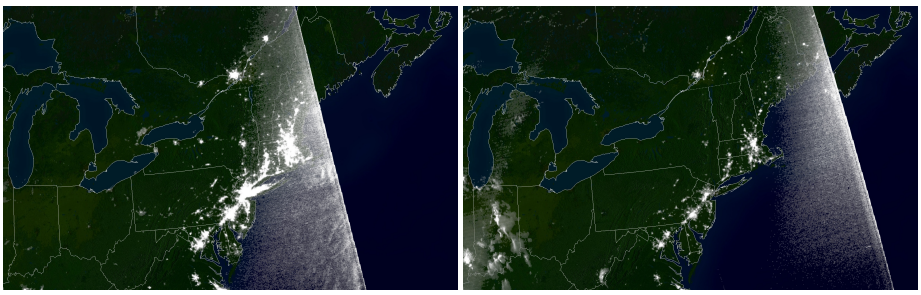


Figure 1.1: Northeastern America before (left) and during (right) the Blackout of 2003, Image Courtesy of National Oceanic and Atmospheric Administration

failures and power plant shutdowns spread across Europe, eventually leaving an estimated number of 10 million households in Germany, France, Belgium, Austria, Spain, Portugal, Italy and Morocco without electricity. On November 5th, 2006, the online news portal *Bloomberg* reported:

> *"Power failure in a German electricity grid operated by E.ON AG caused blackouts across western Europe last night, depriving millions of homes of electricity, disrupting trains and risking outages to hospitals and airports. About 5 million households in France went without power for as much as an hour in the nation's biggest outage since 1978, Andre Merlin, the director of Reseau de Transport d'Electricité, France's power-grid operator, told the press today. Overall, some 10 million households across Belgium, Germany, Italy, Portugal, Spain, and Eastern Europe may have been affected, Merlin said. The grid failure in Germany led to the biggest pan-European power collapse in at least 30 years through a domino effect that swept through Western and Eastern Europe, he said."*[10]

Aftermath of European Blackout

Roughly two hours after the incident, power could be restored and the investigations for the cause of the failure were initiated. The result of these investigations discovered that - in contrast to the events in North America - the blackout was caused by a planned disconnection rather than an unforeseen failure. Upon request of a dockyard, a power transmission line across the river Ems in northern Germany had been disconnected in order to let pass a ship. Although this had been done routinely on several prior occasions, the final report on the incident discovered that - due to the nonobservance of regulations - the disconnection of this single line had driven the system into a critical state in

[10] see `http://www.bloomberg.com/apps/news?pid=20601085&sid=a4J9_1zeDuEo`, accessed on October 5th, 2010

Smart Grid Initiatives

which even small power deviations could set off cascading failures that propagated far through the system [UCTE, 2007]. Since then it has been argued that the increasing importance of renewable energy sources like wind, sun or water, whose output is inherently harder to predict than that of traditional power plants, necessitates new adaptive infrastructures in order to forestall similar failures in future. Consequently, the development of such adaptive power utility infrastructures - so-called *smart grids* - is currently a highly active area of research.

The Swaying Millennium Bridge

In all the examples considered so far, an explicit network topology was involved that interconnected power stations, data centers or computing devices. However even problems occurring in traditional construction engineering, where an explicit network topology is absent, can be related to the networked computing domain. When on June 10th, 2000 the Millennium bridge in London was officially opened, thousands of pedestrians were eager to use the new construction in order to cross the river Thames. As the number of people simultaneously crossing the bridge increased, the newly-opened bridge suddenly started to swing horizontally so intensely that people cramped to the handrails. Two days later, the bridge had become the city's swaying attraction and it had to be closed for further investigation. In an interview with the German radio station D-Radio that was broadcasted on July 1st, 2008, bridge engineer Pat Dallard remembered the opening day:

> *"We had to witness that the bridge was behaving totally different than planned. We had invested much time in the construction and we were sure that we have everything under control. But then something happened which we never ever had expected. [...] Apart from this we observed something very bizarre. The swaying depended in a strange way on the number of people currently on the bridge. Up to a certain number it was stable. Only ten people more and the swaying suddenly appeared."*[11]

Millennium Footbridge Investigations

Needless to say, that a gloating media coverage befell the engineers of the prestigious project even though they had seemingly followed all existing guidelines and regulations. An analysis of video footage of the opening day revealed that the swaying was apparently caused by 2000 pedestrians crossing the bridge in lockstep. The fact that 2000 people marching in lockstep caused the bridge to swing is not at all surprising. That all these pedestrians walked in perfect synchrony was however all the more remarkable. As done routinely,

[11] Translation from German text available at `http://www.dradio.de/dlf/sendungen/forschak/809985/`, accessed on October 5th, 2010

the engineers had assumed that the small random horizontal displacements caused by the steps of a large number of pedestrians would average out to zero. Apparently, this was not the case.

Figure 1.2: Millennium Bridge in London, Image Courtesy of Wikimedia Commons

Analysis of the Millennium Bridge Problem

A full explanation for the phenomenon was found no earlier than five years later in the complex and subtle mutual interactions between the pedestrians and the bridge [Strogatz *et al.*, 2005]. Here one is tempted to ask why this has not been a frequent problem in the engineering of footbridges. The first important reason was the fact that the Millennium footbridge has a resonance frequency that is in the range of the human step rate and thus a sufficient number of people walking in lockstep could cause an increasing swaying of the bridge by means of mechanical resonance. However the force needed to overcome the damping of the bridge was so large that a significant swaying could only be generated by a massive number of people walking in lockstep at or near the resonance frequency of the bridge. If however a certain critical number of people hit the resonance frequency by chance, this could cause a small horizontal displacement that was large enough to force more people to fall into lockstep as they - possibly unconsciously - tried to keep balance. The result of this self-energizing dynamical process was a perfect synchronization of pedes-

trians' steps, possibly more precise than one could have achieved when trying to synchronize the steps of 2000 people on purpose. Although possible effects of pedestrians on the bridge were routinely considered during the design of the bridge, the complex mutual interaction between the bridge's hardly noticeable swinging and people's reaction had been omitted. In the end, the problem could be resolved by installing additional dampers that increased the critical number of people required to hit the resonance frequency by chance to a value that made it sufficiently improbable. Following this event, the guidelines for structural engineering were updated to incorporate the observed phenomenon and a number of bridges around the world were identified that have comparable resonance frequencies and might thus suffer from the same problem [Dallard *et al.*, 2001].

Inadvertent Synchronization in Computing Systems

Referring to the problem with the Millennium Bridge, the computer scientist Jeffrey C. Mogul argued that the fact that such phenomena occur in a "well-regulated engineering profession with decades or centuries of experience [...] and with regular use of computer modeling" should "keep us humble" [Mogul, 2006]. At this point one might be tempted to ask what - if anything - the Millennium Footbridge example has to do with the domain of computer science. In fact, very similar phenomena have occurred in networked computing systems. In order to build and maintain information on shortest routes in the Internet, routers need to periodically exchange protocol messages. As the majority of routers are operated and set up independently, one would expect the traffic induced by these periodic messages to be scattered uniformly across time. The intuition that the independence of individual routers would result in an uncorrelated traffic pattern was proven wrong in the early 1990s, when traffic studies exhibited periodic traffic peaks that resulted from a highly synchronized exchange of routing table updates. It was subsequently argued that these periodic peaks were the reason for earlier observations of periodic packet losses and periodic changes in round trip times.

Investigation of Synchronized Traffic

Shortly after these observations, the phenomenon had been studied and in [Floyd and Jacobson, 1994], implicit interactions between routers were identified as cause of the problem: It was conjectured that the processing of incoming routing updates can subtly affect the timing of future routing table updates sent by a router. Based on this it was shown that - even when considering a fairly large amount of random fluctuations - these subtle interactions eventually lead to a synchronized exchange of protocol messages. Similar to the Millennium bridge example, this can cause more and more routers to fall into lockstep by

means of a self-energizing process. Based on these findings, in [Floyd and Jacobson, 1994] it has been argued that

> *"[...] the architect's intuition that independent sources give rise to uncorrelated aggregate traffic is simply wrong."*

Since then, the prevention of inadvertent synchronization phenomena has been studied intensively and in today's Internet, Active Queue Management techniques like Random Early Detection [Braden *et al.*, 1998] are used to alleviate related synchronization problems.

1.2 Implications for Networked Computing Systems

Characteristics of Complex Phenomena

Above we have considered some prominent examples for detrimental phenomena that occur in sufficiently large and complex distributed systems both within and outside the domain of computing. In fact, one could instance many more which showcase similar effects (see for example the reviews of [Gribble, 2001; Mogul, 2006]). Clearly, once the causes for detrimental phenomena have been investigated, it is usually possible to adjust or redesign systems in a way that alleviates the resulting problems. Nevertheless the fact that the possibility of such problems has not been foreseen during their design must worry any engineer. What are the implications of this for the engineering of networked computing systems? To answer this question, we first take a closer look at the mechanisms and characteristics underlying the emergence of the phenomena above.

Self-Organized Formation of Structures and Patterns

The examples of the Millennium Footbridge and the Internet's router infrastructure illustrate a remarkable feature that can arise from complex interactions, namely globally coherent patterns emerging without central coordination. In the above examples, simple temporal oscillatory patterns (in terms of a synchronization of seemingly independent, periodic processes) were formed due to the subtle interactions between a large number of pedestrians or routers. In other types of systems much more complex, spatial, topological or temporal structures can emerge. *These structures are neither blueprinted nor coordinated centrally which is why they are often said to emerge in a self-organized fashion. The phenomenon of self-organization typically occurs in non-linear dynamical systems and an analytical prediction of the structures and patterns resulting from such non-linear interactions is rarely possible.*

Order from Noise

A nice example for the self-organized and unexpected formation of a temporal pattern is the synchronization of routing messages in the early Internet. An interesting finding of the subsequent study of the phenomenon that has been performed in [Floyd and Jacobson, 1994] is that a surprisingly large amount of randomness in the router's intrinsic update periods is required to break the synchrony. Contrariwise the application of moderate amounts of random fluctuations has even been shown to result in synchronization to occur more quickly and to be more robust. *The seemingly paradoxal principle that random fluctuations can facilitate the self-organized formation of ordered states has been dubbed "order-from-noise" [von Foerster, 1985] and it has since been argued that this principle is abundant in natural self-organization processes like for example the organization of coherent patterns in swarms of particles, animals or people [Vicsek* et al.*, 1995; Helbing and Vicsek, 1999; Yates* et al.*, 2009].*

Hidden Threshold Phenomena

Another interesting aspect of many complex phenomena is the fact that they do not develop gradually but occur rather spontaneously. For the synchronization of routing messages it was found that a certain critical number of routers is required for synchrony to emerge. Systems whose size is below this critical number will show no signs of synchronization while synchrony emerges spontaneously as soon as a certain critical number is exceeded [Floyd and Jacobson, 1994]. The same has been found for the Millennium Bridge: The swaying occurred abruptly when the number of people crossing it exceeded a critical number [Strogatz *et al.*, 2005]. Such "hidden thresholds" are typical for complex phenomena. The fact that it is often hard if not impossible to derive them analytically poses a threat to the engineering of systems with predictable performance. For distributed computing systems being deployed at massive - possibly global - scale, full scale tests are not a feasible option. At the same time small scale test bed deployments may show no signs of detrimental behavior as long as a certain critical size is not exceeded. Furthermore, simulations may not capture implicit, non-linear interactions that lead to detrimental patterns and a formal analysis is often out of reach due to the non-linear, dynamical nature of the processes involved.

Self-Organized Criticality

A number of the examples for systemic failures that have been considered above can be related to the occurrence of cascading failures. Here one observes that - although for entirely different reasons - systems were driven to a state with insufficient over-provisioning of resources. This took away the "damping" necessary to prevent the spreading of cascading effects. If taken alone this is not a surprising phenomenon: Once the domino stones are set it clearly is not astonishing that the toppling of the first stone causes the other stones to top-

ple as well. In this picture, the question what is setting up the stones is much more interesting. Why do many technological and natural systems seemingly self-organize into critical states in which even infinitesimal perturbations can trigger cascades that affect the whole system? In physics, systems with this property are said to exhibit the property of self-organized criticality [Bak *et al.*, 1987] and the mechanisms underlying it have been investigated extensively. For power transmission grids, a combination of technological requirements, engineering guidelines and economic incentives has been identified to be responsible [Motter, 2002; Carreras *et al.*, 2002]. More recently, in [Buldyrev *et al.*, 2010] it has been argued that the interdependence of communication networks and power grids can further cascading failures and thus set off large scale power outages.

Limits to Predictability

From the perspective of a computer system engineer, the main problem with all these phenomena is that they pose a threat to the predictability of the behavior of systems: Hidden threshold effects complicate their design as they prohibit to extrapolate the behavior observed in small scale tests to larger settings. Furthermore self-organized critical states and non-linear interactions can give rise to situations in which infinitesimal fluctuations can have arbitrarily large effects that lie far away both in space and time and that are hardly predictable. By means of subtle interactions and self-organizing processes, apparently independent events can exhibit surprising correlations and they may benefit from noise and random fluctuations that are usually thought to prevent them. Since at sufficiently large scale, even systems comprised of simple elements with rather simple interactions can exhibit surprisingly complex behavior, all these aspects are likely to become increasingly important as distributed systems grow in size. *Increasing interconnectedness, heterogeneity and dynamics in future distributed systems are thus likely to facilitate the inadvertent occurrence of complex detrimental phenomena.*

Complexity is taking over

Since roughly ten years, the exuberant complexities involved in the design and operation of large scale computing systems are increasingly acknowledged in both industry and science. Fueled by the rise of the World Wide Web, in the early 2000s complexities involved with the operation of high performance data centers had become one of the prime challenges in the computing industry. In order to address these issues, in 2001 IBM launched the Autonomic Computing initiative, soon being joined by similar projects of major companies like Sun, Microsoft and Hewlett Packard. In IBM's manifest from 2001, the problem being tackled by Autonomic Computing is summarized as follows:

> *"In fact, the growing complexity of the I/T infrastructure threatens to undermine the very benefits information technology aims to provide. Up until now, we've relied mainly on human intervention and administration to manage this complexity. Unfortunately, we are starting to gunk up the works." [IBM, 2001]*

The Autonomic Computing Vision

As suggested by this quote as well as the name of the initiative, the main objective of the Autonomic Computing vision was to repel the need for human intervention and thus simplify the practical management of large scale data centers. Technically, this can sometimes be achieved by an integration of monitoring and control capabilities as well as closed control loops into computing infrastructures. By this, operational parameters (like resource utilization or quality of service) can be kept in a range specified by corporate policies ideally without requiring outside intervention. Since then, most major companies have incorporated similar technologies into their products. In recent years the wide-spread adoption of virtualization techniques has further facilitated the use of related mechanisms.

Global Scale Networked Computing Systems

While the complexities arising in the operation of data centers may be severe, the situation is even worse when considering networked computing systems being deployed at a global scale. In many of these systems, structures emerging at a large scale are not subject to centralized control and detailed planning. Instead, in systems like for instance the Internet's routing infrastructure (a picture of which is shown in Figure 1.3) or large-scale Peer-to-Peer systems, network topologies rather result from distributed processes being influenced by economic incentives, dynamic user behavior and technological constraints. What are the large-scale properties of such networked systems? Is there a bound on the Internet's diameter? To what extent can a Peer-to-Peer system compensate node and link failures without being fragmented in different subsystems? And how are such properties related to the local mechanisms building the global network topology? In [Gkantsidis *et al.*, 2003], the complexities emerging in such networked computing systems are summarized as follows:

> *"The network paradigm is shifting. Today's open, distributed and dynamic networks are no longer artifacts that we construct, but phenomena that we study." [Gkantsidis* et al.*, 2003]*

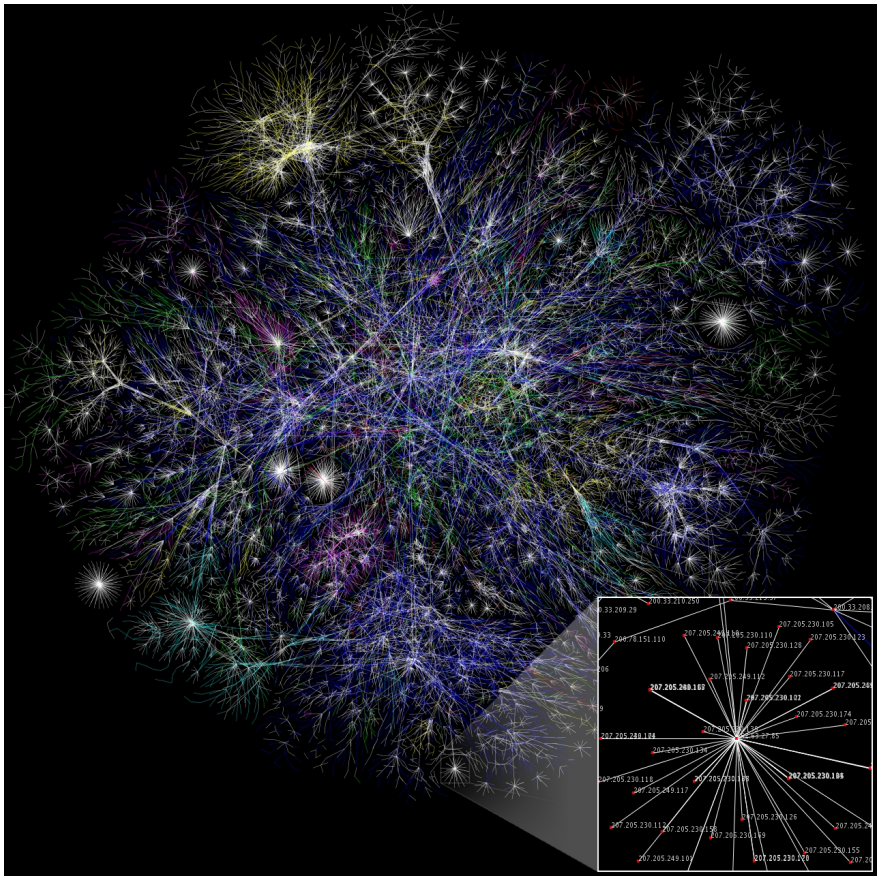


Figure 1.3: Partial Map of IP Routes based on data from `http://www.opte.org` obtained in January 2005. Image courtesy of Matt Britt.

Large Dynamic Networked Systems - Reclaiming Control

Based on these observations, the question arises how large scale networked systems can be managed in the face of these complexities. In the examples considered above, we have so far only considered detrimental complex phenomena. This is however only one side of the coin since the last decade's research has shown that the principles underlying these problems are in fact the very same that underlie the efficiency, robustness and adaptiveness of a variety of natural systems. Apart from giving rise to detrimental patterns, self-organization processes can actively be used to produce advantageous spatial, temporal or topological structures. By means of the "order-from-noise" principle, these structures can be robust despite random fluctuations and their robustness can in fact benefit from a certain amount of randomness. Similarly, a controlled use of critical states and "hidden thresholds" could be used to intentionally change a system's properties and thus make it more adaptive.

The main question being studied in this dissertation is how such aspects can be actively used in the design and operation of current and future networked computing systems. In particular, we investigate how a targeted application of the last decade's findings about the formation of complex network structures and the collective dynamics emerging in such networked systems can contribute to enhance their scalability, reliability and predictability. Related issues are increasingly being acknowledged by the research community and have recently been addressed for example in the "Organic Computing" priority program of the German Research Foundation DFG [Müller-Schloer *et al.*, 2004; Schmeck, 2005]. In the following chapters, we will study both theoretical and practical aspects that are related to this question, by this proposing what one may call *emergent organization principles* for future networked computing systems.

Emergent Organization Principles

1.3 Overlay Networks and Peer-to-Peer Systems

Overlay Networks

A class of distributed computing systems for which this approach is particularly promising are those being built upon so-called *overlay networks* or short *overlays*. In these systems, a virtual network topology - the overlay - is used in which links represent the graph of communication across a set of nodes representing network devices, applications or users. Actual communication between two nodes connected by a link in the overlay is realized over connections that exist in one or more underlying network layers. This so-called underlay provides the basic transport, addressing and routing mechanisms that are used by higher level services. At the level of the overlay network, custom addressing or routing schemes can possibly be realized. An example for a simple overlay topology can be seen in Figure 1.4. Here, nodes in the overlay are virtually connected to a ring topology, each node representing for example the machine of a user of a distributed application or service. Using the routing and addressing scheme provided by the underlying network infrastructure, in this example a message exchange between two overlay nodes across a virtual link e involves sending the message across four links $e_1, \ldots, e_4$ in the corresponding underlay. The actual overlay topology being used to provide a particular service is typically chosen such that it facilitates the efficient operation of certain distributed algorithms for example for distributed search, data distribution or decentralized routing.

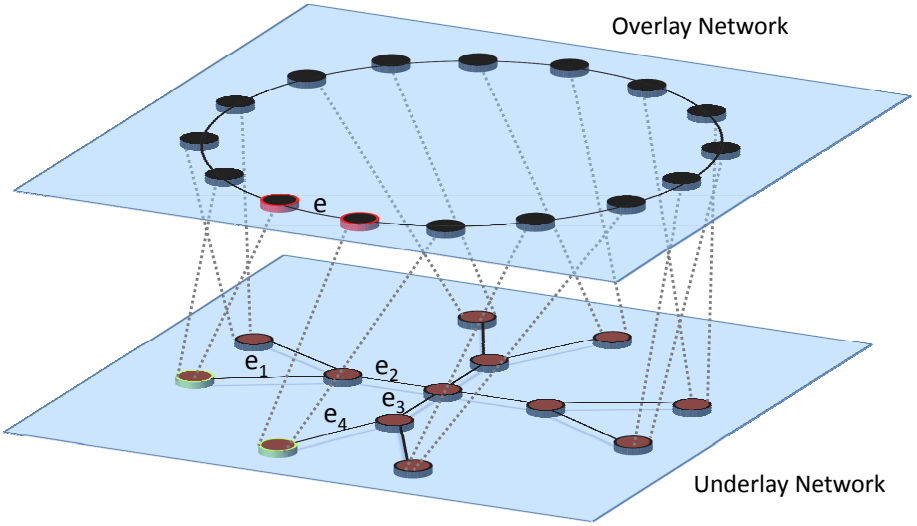


Figure 1.4: Example Overlay Network built on top of an Internet-style Underlay

Internet-based Overlays

Overlay networks, whether being named so explicitly or not, are used regularly in the design of networked computing systems since several decades. In fact, the Internet can be viewed as a virtual overlay spanning vastly different organizational domains, network technologies and transport mechanisms, thus providing a homogeneous communication space in which all devices can potentially communicate with each other by means of addressing and routing schemes provided by the *IP* protocol suite. Higher level overlays can then be used to introduce novel services and applications without requiring to deploy new devices or protocols in the underlay. Examples for widely used Internet services that are provided by means of overlay topologies can be found in the Domain Name System, content distribution networks like AKAMAI as well as multicast protocol extensions. It has been argued for example in [Waldhorst *et al.*, 2010], that overlay networks constitute a crucial abstraction for the large scale deployment of novel services and applications in future iterations of the Internet.

Overlay Networks and Peer-to-Peer Systems

A special class of distributed systems typically employing overlay topologies at the application-layer are Peer-to-Peer based services and applications. Blurring the traditional distinction between resource provider and consumer, Peer-to-Peer systems typically use application-specific overlays to leverage the processing power, bandwidth and storage capacity of user machines. By harnessing the resources of users, Peer-to-Peer systems can achieve a remarkable

scalability at low cost. Since novel services and applications can be deployed without the need to set up, maintain and control costly centralized infrastructures, Peer-to-Peer systems have raised significant interest both in research and business. For services and applications being deployed at very large, possibly global scale, the decentralized approach underlying Peer-to-Peer systems is particularly interesting because it can avoid scalability bottlenecks, mitigate single points of failures and circumvent censorship. Having grown popular mostly in illegal file-sharing systems in the 1990s, today large scale decentralized Peer-to-Peer systems like the video-telephony service SKYPE or the content distribution protocol BITTORRENT [Cohen, 2003] are among the most widely used services in the Internet. It has since been argued that the decentralized paradigm underlying these services is emerging as a leading pattern in the design of scalable and reliable distributed systems [Wehrle *et al.*, 2005].

Decentralized Peer-to-Peer Systems

The personal motivation for the research underlying much of this thesis is - amongst others - based on experiences obtained during the practical development of such a Peer-to-Peer system, namely the *Event Monitoring* service EMON. It is currently being used for the distribution of particle collision data from the read-out systems of CERN's ATLAS detector to the machines of physicists participating in the Large Hadron Collider (LHC) experiment [Scholtes, 2005; Kolos and Scholtes, 2005; The community of testers and developers in ATLAS DAQ/HLT, 2005]. In a nutshell, EMON is a distributed service allowing users to sample collision event data from different points of the data flow chain of the ATLAS detector. A scalable dissemination of these data to possibly a few thousand subscribers is achieved by employing a Peer-to-Peer based content distribution scheme. For this, the machines of those users wanting to sample event data are interconnected in a multicast overlay topology, thus harnessing the bandwidth of user machines for redistributing events to other users. Based on event selection criteria and different sampling points in the ATLAS detector's data acquisition infrastructure, several k-nary tree overlays are used for this purpose. A schematic view of this overlay for two sampling points in the data flow chain is depicted in Figure 1.5. Here, each tree node represents a machine of a user - a so-called Monitoring Task - subscribing to events from a particular sampling point, edges in the overlay are virtual connections used to redistribute incoming events to machines of other users. In each of these distribution trees, the most capable machines are moved towards the root of the tree while the less capable constitute the leaves. Further details on algorithmic and imple-

The Event Monitoring Service

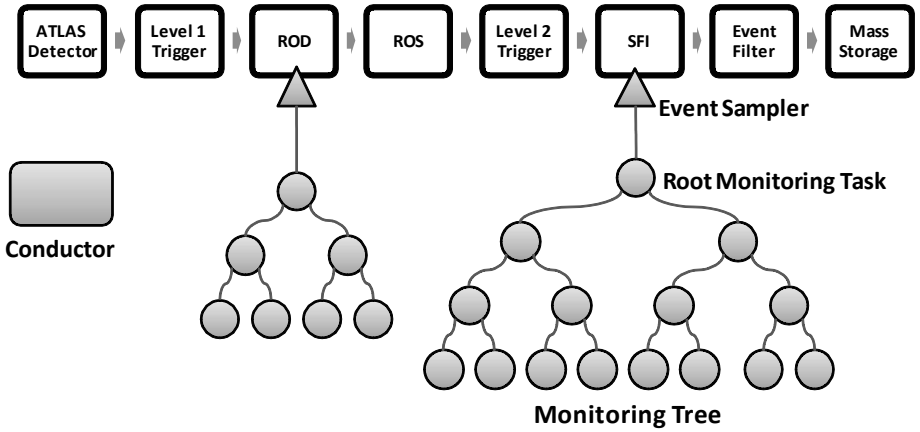


Figure 1.5: Event Monitoring Data Distribution Overlay

mentation aspects of the EMON service can be found in [Scholtes *et al.*, 2006; Scholtes *et al.*, 2008c].

Managing the EMON *Overlay*

Clearly, the overlay topology being used in EMON is extremely simple and with at most a few thousand users the number of involved machines is comparably moderate. Furthermore, a rather tightly controlled scientific setting is addressed in which malicious user behavior can be disregarded for the most part. Nevertheless, practical experiences obtained during the design, testing and maintenance of this system have demonstrated that the efficient and reliable management of overlay networks in the face of unreliable and heterogeneous machines is non-trivial. The algorithmic aspects of highly structured overlays like EMON's tree topology may appear simple. However the actual implementation of such systems is complicated significantly by error-prone links and machines, heterogeneous resources and the unpredictable dynamics of users. In such highly dynamic distributed systems, maintaining a particular overlay structure required by deterministic distributed algorithms is non-trivial. Furthermore, in the practical implementation of sophisticated distributed maintenance algorithms, preventing deadlocks, race conditions and overlay inconsistencies has proven to be a considerable challenge due to the massive concurrency involved.

Harnessing Complex Phenomena in Peer-to-Peer Systems

These problems are in no way unique to the Event Monitoring service but rather showcase the complexities involved with the management of overlay networks in general. In fact, these difficulties are aggravated considerably when designing systems that are deployed at a very large, possible global scale. In Peer-to-Peer systems, overlays typically interconnect user-contributed ma-

chines that are not managed professionally. They may crash frequently and can join or leave the system at any time, thus resulting in a highly dynamic topology with continuously and concurrently fluctuating nodes and links. In such systems, efficiently maintaining highly structured overlay topologies is a considerable challenge. The continuous fluctuation of participants, an effect commonly called *churn*, requires the use of distributed algorithms which *repair and maintain* network structures being required for distributed algorithms to operate correctly and/or efficiently. *However, at the same time one can argue that the virtual and alterable character of connections in systems built upon overlays facilitates the use of phenomena in which favorable complex structures emerge in a self-organized fashion based on simple local, possibly randomized connection schemes. Furthermore, we will see that the emergence of globally coherent patterns from local stochastic interactions suggests and active use of complex collective behavior in large dynamic Peer-to-Peer systems.*

1.4 Contributions and Outline of this Dissertation

In the remainder of this dissertation, different facets will be considered that are related to the question how complex phenomena can be used in a targeted fashion in large scale networked computing systems. Based on their relevance for the deployment of services in the future Internet, a particular focus has been laid upon systems being built upon overlay networks. Coarsely resembling the structure of this thesis, in the following paragraphs those aspects that constitute the main contributions of this work are briefly summarized. As will be argued in more detail later, an important aspect of this dissertation is to take an interdisciplinary perspective on the management of very large, dynamic and networked computing networks. In particular, we seek to incorporate the views of random graph theory, complex network science and statistical mechanics on the management of overlay networks whose characteristics are favorable for an application in distributed computing systems. In chapter 2 we thus introduce some necessary concepts and review a number of findings from these disciplines that seem relevant for the engineering of networked computing systems.

Overlay Networks - A Thermodynamic Perspective

Having introduced the necessary conceptual framework of random graphs and complex networks, in chapter 3 we briefly review the strengths and weaknesses of *structured* and *unstructured* approaches to the management of large dynamic overlay topologies. We then consider a *thermodynamic perspective on the management of overlay networks* which is based on analogies between com-

plex networks and *statistical mechanics* that has been introduced in chapter 2. We further discuss how a study of models from complex network science and an application of analogies to particle systems gives rise to what one may call *thermodynamically structured* overlays, actively making use of strong stochastic guarantees stemming from simple, randomized protocols and employing phase transition phenomena to adapt macroscopic network qualities. Being aware that an application of theoretical results from the study of random graphs and complex networks in engineered networked systems requires caution, we argue that the virtual character of connections in overlay networks facilitates a well-grounded, deliberate and constructive use of these findings. Parts of the concepts that will be discussed in chapter 3 have originally been published in [Scholtes *et al.*, 2008a; Scholtes *et al.*, 2010].

Creating and Adapting Scale-Free Overlays

Based on the ideas put forth in chapter 3, in chapter 4 we first critically appraise the use of findings about random scale-free graphs in computer networks. In particular, we argue that the use of randomization techniques facilitates the application of analytical results summarized in chapter 2. We then turn to some practical aspects that arise when wanting to create random scale-free overlays with freely adjustable degree distribution exponent in practice. In section 4.2, we introduce a distributed, probabilistic rewiring scheme that can be used for this purpose. The scheme has recently been presented in [Scholtes, 2010] and experimental as well as analytical results suggest that it constitutes a practicable approach to efficiently create randomized overlay topologies with scale-free characteristics and variably pronounced hubs. We further suggest that the scheme can be used to actively trigger phase transitions in order to adapt structural properties of scale-free networks that are relevant for the performance of networked computing systems.

Monitoring Scale-Free Overlays

Based on the notion of *thermodynamic guarantees* that has been introduced in chapter 3, in section 4.3 we introduce a distributed monitoring scheme by which the degree distribution exponent of power law networks can be derived in a distributed and probabilistic fashion. This scheme has first been proposed in [Scholtes *et al.*, 2008a], where we have argued that it can be used for a *macro-level reasoning* about the topological properties of power law networks as well as about the performance of dynamical processes and distributed algorithms operating upon them.

Self-Organized Synchronization in Peer-to-Peer Systems

In chapter 5 we somewhat shift the focus away from the formation, adaptation and monitoring of overlay networks with complex, probabilistic structures. Here, we rather consider the use of complex collective phenomena emerging by means of non-linear dynamical processes operating - for example

- on such networks. We present a probabilistic, message-based variation of the Kuramoto model for coupled oscillators that has been introduced in [Scholtes *et al.*, 2009; Scholtes *et al.*, 2010] We argue that a limit cycle attractor of this dynamical system can be used to efficiently synchronize a set of oscillating signals in overlay networks with good mixing properties. The proposed scheme can thus be used for example to provide synchronous heartbeats in decentralized Peer-to-Peer systems. Compared to existing schemes that rely on pulse coupled models, the proposed algorithm provides the benefit that associated message exchanges are randomly distributed randomly across time, rather than occurring synchronously.

Monitoring Networks by Synchronization

In section 5.2.3 we further consider the question whether this self-organized synchronization scheme can be used to monitor spectral properties of overlay networks. We particularly present experimental results suggesting that a time series analysis of local oscillations and incoming coupling offsets can facilitate a distributed reasoning about a network's algebraic connectivity. While this idea has originally been put forth in [Scholtes *et al.*, 2010], we further present initial experimental evidence that the synchronization scheme allows nodes to locally assess community structures in their neighborhood.

An Interdisciplinary Perspective on Network Engineering

Having considered the use of both complex structures and complex collective phenomena in networked computing systems, we conclude this dissertation in chapter 6. Here we reflect on the potential of applying recent findings from the field of complex networks and complex systems science in the engineering of networked computing systems that actively use emergent organizational principles. Furthermore we critically appraise the contributions of this dissertations towards achieving this goal. Finally, we present some perspectives of an interdisciplinary approach to the engineering of networked computing systems and give an outlook to future work.

Chapter 2

Introduction to Random Graphs and Complex Networks

> Smooth shapes are very rare in the wild but extremely important in the ivory tower and the factory.
>
> BENOÎT MANDELBROT

Random Graphs

When wanting to reason about the structures and dynamics unfolding in very large networked computing systems, details of the processes shaping these networks are often either unknown or uncontrollable for example because they are influenced by environmental conditions or user behavior. Apart from the well-established study of networks with highly regular structures like trees, rings or regular lattices, it is thus often helpful to consider networks that are generated by stochastic processes. Related questions are being studied in the field of random graph theory, which was pioneered in 1959 by Pál Erdös and Alfréd Rényi [Erdös and Rényi, 1959]. Even though the models that underly these so-called Erdös/Rényi random graphs are extremely simple, they have been an active area of research for more than five decades. An exhaustive and authoritative review of the findings of this research can be found for example in [Bollobás, 2001]. In this section we briefly introduce some basic notions and findings of classical random graph theory only in so far as they relate to the topic of this thesis and demonstrate the reasoning about networks with more complex, probabilistic structures. As we shall see in subsequent sections and chapters, some of the results stemming from this study can have

important implications for the design and operation of networked computing systems and large scale overlay networks. At this point however, we first need to precisely state our notion of the terms *graph* and *network*, both of which will be used interchangeably throughout this thesis.

Definition (Network, Graph)**.** *Let V be a set and $E \subseteq V \times V$. We call the ordered pair $G = (V, E)$ a graph or network with distinguishable vertices (or nodes) V. An element $(v, w) \in E$ is called an edge (or link) from v to w. In this case, w is said to be adjacent to v and the edge (v, w) is said to be incident to v. An edge (v, v) is called a self-loop. A network or graph is said to be undirected if $\forall (v, w) \in E \rightarrow (w, v) \in E$, otherwise it is said to be directed.*

In the remainder of this dissertation we consider *labeled* graphs and networks, that is we explicitly distinguish also between those networks that are identical up to a relabeling of nodes. We further restrict our consideration to undirected networks, notwithstanding the fact that similar findings on directed networks are equally well applicable in the engineering of networked computing systems.

2.1 Basics of Random Graph Theory

The $G(n, m)$ model

The prime abstraction of random graph theory is that of a *graph ensemble*, that is a probability space consisting of a set Ω of graphs and a measure P assigning each graph $G \in \Omega$ a probability $P(G)$. A simple graph ensemble considered in classical random graph theory is given by the the so-called $G(n, m)$ model. Here, the set of all possible undirected, labeled graphs with n vertices, m edges and no self-loops is considered, each of them being ascribed equal probability. The number of such graphs is given by the binomial coefficient $\binom{n}{m}$. Thus the associated probability measure P assigns each possible graph realization G a probability

$$P(G) = \binom{n}{m}^{-1}. \tag{2.1}$$

The $G(n, p)$ model

A slightly different model that is frequently considered in the study of random graphs is the $G(n, p)$ model. The basic idea underlying this model is a stochastic process that generates edges in an initially empty graph with n vertices. In this process, edges between all $\binom{n}{2}$ pairs of vertices are added to the graph with uniform probability p, thus eventually defining a probability space that contains all graphs with n vertices and any number of edges. In the $G(n, p)$

model, the expected number of edges for a randomly chosen graph is $p\binom{n}{2}$ and the probability $P(G)$ that a graph G with exactly m edges is created is given by

$$P(G) = p^m \cdot (1-p)^{\binom{n}{2}-m}. \tag{2.2}$$

In particular, this measure implies that - in general - graphs with different numbers of edges are assigned different probabilities. However, for $p = \frac{1}{2}$ one obtains an ensemble in which all graphs - independent from the number of edges m - are equally probable. Figure 2.1 shows a random realization of a graph generated according to the $G(n,p)$ model with $n = 200$ and $p = 0.03$.

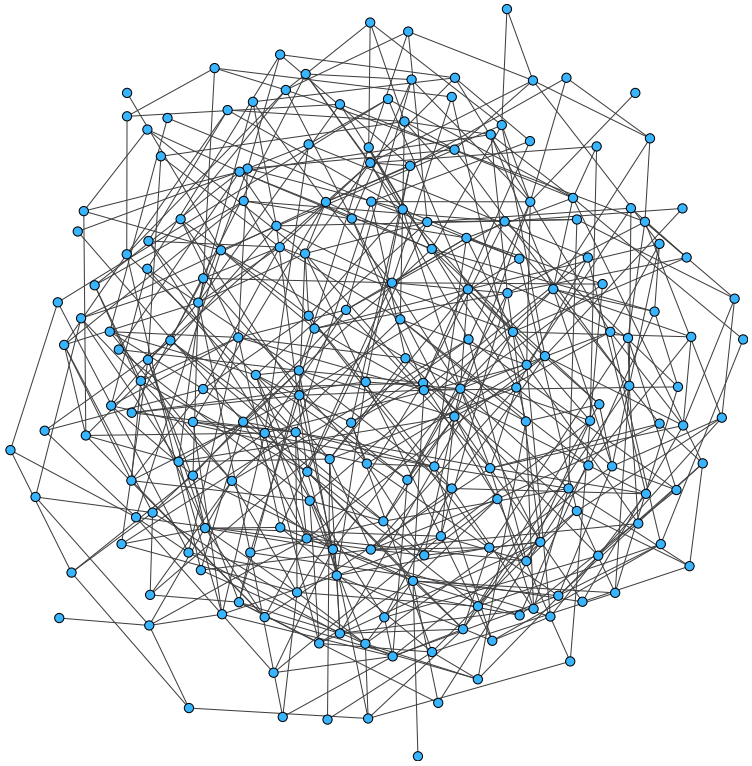


Figure 2.1: A random $G(n = 200, p = 0.03)$ graph

Arguing about Structural Properties

Defining the $G(n,m)$ and $G(n,p)$ ensembles allows to ask interesting questions about the properties of randomly chosen (or if you will randomly generated) networks. Technically, those graphs that exhibit a certain structural property like for example connectedness define a subset $\omega \subset \Omega$ of all possible graph realizations Ω. By means of the probability measure P that is defined in both models, one can argue about the probability of these subsets and thus

about the probability that a randomly chosen or generated graph exhibits a certain property. For the $G(n,p)$ and the $G(n,m)$ model, the probability of a certain graph property ω can then be expressed in terms of the parameters n, p and m respectively.

Properties Holding Almost Surely

A particularly interesting class of properties are those that hold *asymptotically almost surely* in a given graph ensemble. For an arbitrary graph property x and with $\omega_x \subseteq \Omega(n)$ denoting the subset of all graphs that exhibit property x in the probability space $\Omega(n)$ which consists of all possible networks with n nodes, we can say that property x holds *asymptotically almost surely* if and only if $lim_{n\to\infty} P(\omega_x) = 1$. In other words, the probability that a randomly drawn graph exhibits a property that holds asymptotically almost surely converges to 1 as the size of the graph goes to infinity. Such properties are interesting for a number of reasons. First of all - given that this convergence is fast enough - strong guarantees for their occurrence can be obtained if the size of a graph is sufficiently large and if the graph has been chosen at random. Furthermore, such arguments can be used to explain the occurrence of common characteristics in a variety of networks that emerge in natural and technical systems. Speaking informally, the fact that a certain property is common across different networks is not surprising if it is almost impossible to find a graph without this property.

Connectedness in Erdös/Rényi Graphs

For the graph ensembles defined by the classical models, a number of properties and the dependence of their probability on model parameters have been analyzed. Since connectedness is among the most fundamental properties of a graph, it was one of the first being studied for random graphs [Erdös and Rényi, 1959]. We say that a graph is connected if a path p exists between each possible pair of its nodes. Here, a path p between two nodes v and w is a sequence of edges $(v_1, v_2), (v_2, v_3), \ldots, (v_{l-1}, v_l)$ with $v_1 = v$ and $v_l = w$. Naturally, the number of edges m required to ensure connectedness increases as the size of the graph n is increased. Considering $m(n)$ as a function depending on the graph size, for the $G(n, m(n))$ model the number of edges

$$m(n) = \frac{n \cdot ln(n)}{2}$$

has been found to be a critical threshold. Below this threshold almost no graph in $G(n, m(n))$ is connected, while almost every graph is connected above the threshold. Analogously a size-dependent threshold function

$$p(n) = \frac{log(n)}{n} \tag{2.3}$$

Phase Transition Phenomena in Random Graphs

could be identified for the connectedness of graphs in the $G(n, p(n))$ model. The most interesting aspect of these findings is the fact that the (large graph limit for the) probability that a randomly chosen graph is connected changes abruptly from 0 to 1 as $p(n)$ and and $m(n)$ are varied. In [Cohen, 1988], it is argued that this behavior is closely related to phase transition phenomena occurring in statistical physics. Another connectivity-related property for which such a threshold behavior could be identified is the size of a graph's largest connected component. For a vertex v, the connected component C_v contains all nodes w for which a path between v and w exists. In the $G(n, m(n))$ model, the size of the largest connected components is proportional to n for almost every graph if $m(n)$ is chosen such that

$$lim_{n\to\infty} \frac{m(n)}{n} > \frac{1}{2}.$$

Diameter of Random Graphs

In this case, a path between all but a constant fraction of nodes exists and the associated connected component is called *giant connected component*. Beyond connectedness, a property that is important for the design of communication networks is a graph's diameter. Defining $l(v, w)$ as the length of the shortest path between two nodes v and w, the diameter is given by the maximum of $l(v, w)$ for any two vertices v and w. For the $G(n, p)$ model, it has been found that the diameter of almost every graph in $G(n, p)$ is d if p is chosen such that

$$\frac{p \cdot n^{d-1}}{n} \to \infty \text{ for } n \to \infty.$$

It has further been shown that the diameter of almost all graphs in the $G(n, p)$ models is virtually identical for a large range of p, which translates to the fact that adding connections beyond a certain point will most likely leave a graph's diameter unchanged.

The Small World Phenomenon

Despite the simplicity of the underlying models, findings about diameter or connectedness of random graphs have in fact been used to reason about the properties of real-world networks. A famous example is the small diameter of social networks that was observed in the late 1960s by Stanley Milgram in his now famous small-world experiment [Milgram, 1967]. Random graph theory informs us that the mere existence of such short paths is not at all surprising since almost every randomly chosen graph with a sufficiently large number of edges will exhibit a small diameter. Hence, the fact that a network has a small diameter - if taken alone - does not qualify as a sign for a sophisticated structure.

2.2 The Study of Complex Networks

Limitations of Classical Models

While first steps in studying the social graph of the United States had been taken by Stanley Milgram in the 1960s, roughly 30 years later the increasing availability of network data sets and computing resources opened unprecedented opportunities to study further properties and statistical features of networks occurring in real-world systems. Since then, the characteristics of networks stemming from data as diverse as protein-protein interactions, scientific co-authorships, movie databases, technological systems or social interactions have been investigated. Here, a number of statistical features like clustering substructures or characteristic connectivity distributions have been found to be common across largely different systems. The structures in these networks are neither completely random - in the sense that they could be explained in the framework of the classical $G(n, m)$ and $G(n, p)$ models - nor regular, which is why they are commonly called *complex networks*.

Simple Models for Complex Networks

Clearly, the fact that classical random graph models fail to reproduce complex structures that can be observed in natural or technological systems is not surprising. Links in biological, technological and social networks like those mentioned above are hardly created independently at random. The probability that a link between two particular nodes is created is rather influenced by a variety of factors like the capacity and popularity of nodes, their spatial embedding, the costs or benefits entailed by creating and maintaining links and many further aspects. Furthermore, in systems that evolve gradually, existing nodes and links are likely to influence those edges being created by subsequently entering nodes. During the last decade, a number of network models have been proposed which aim at better capturing such properties and which are able to reproduce the complex structures that can be observed in social, natural and technological networks.

Clustering Coefficient

An example for a characteristic that is common in real networks but which is unlikely to emerge in the classical random graph models is the existence of clustering substructures which can be assessed by a network's clustering coefficient as defined for example in [Watts and Strogatz, 1998]. The local clustering coefficient C_v of a node v in an undirected network is given as

$$C_v := \frac{2 \cdot |\{(x, y) \in E : (x, v) \in E \text{ or } (y, v) \in E\}|}{d_v(d_v - 1)} \tag{2.4}$$

which tells what percentage of all connections that could possibly exist between the neighbors of a node v are present in the network in question. With

this, the global average clustering coefficient of a network $G = (V, E)$ with $|V| = n$ can be defined as:

$$C(G) := \sum_{v \in V} \frac{C_v}{n} \tag{2.5}$$

The Watts/Strogatz Model

In [Watts and Strogatz, 1998] the average clustering coefficients and the diameter of networks representing movie actor relationships, the connectivity of a power grid and a neuronal network have been compared to that of random graphs. It was found that - while the small diameter of these networks can be explained by classical random graph models - the clustering coefficient of actual networks is sufficiently higher than that of random graphs. In order to capture both high clustering and small diameter, in [Watts and Strogatz, 1998] a model has been proposed that starts with a regular ring topology, each node being connected to the k closest nodes in the ring. A parameter p is then introduced which gives the probability that one endpoint of each edge is rewired to a node that is chosen uniformly at random. The parameter p can be tuned to produce graphs of increasing randomness that range between a regular ring lattice for $p = 0$ and a random graph for $p = 1$. Figure 2.2 shows a random realization of such a Watts/Strogatz network for $n = 200$ and $p = 0.1$. The main contribution of this model is the fact that - within a certain range of p - the resulting networks exhibit a small diameter while the initial clustering structure is partially retained. As such, in the probability space created by the Watts/Strogatz model two characteristics that can be found in a variety of networks encountered in reality are highly probable. It has since been found that the combination of small diameter and high clustering - the so-called "small-world" property - is typical for networks mapping for example social interactions.

Degree Distribution

Another simple statistical feature that can easily be measured for actual networks - at least if one has access to their global topology - is the distribution of vertex degrees. With d_v denoting the number of links incident to node v, the observed degree distribution P_G of a particular network realization $G = (V, E)$ is given as

$$P_G(k) := \frac{|\{v \in V : d_v = k\}|}{|V|} \text{ for } k \in \mathbb{N}. \tag{2.6}$$

Degree Distribution of Classical Random Graphs

For both the $G(n, p)$ and the $G(n, m)$ model, the degree distribution of randomly generated networks converges with high probability to a Poissonian

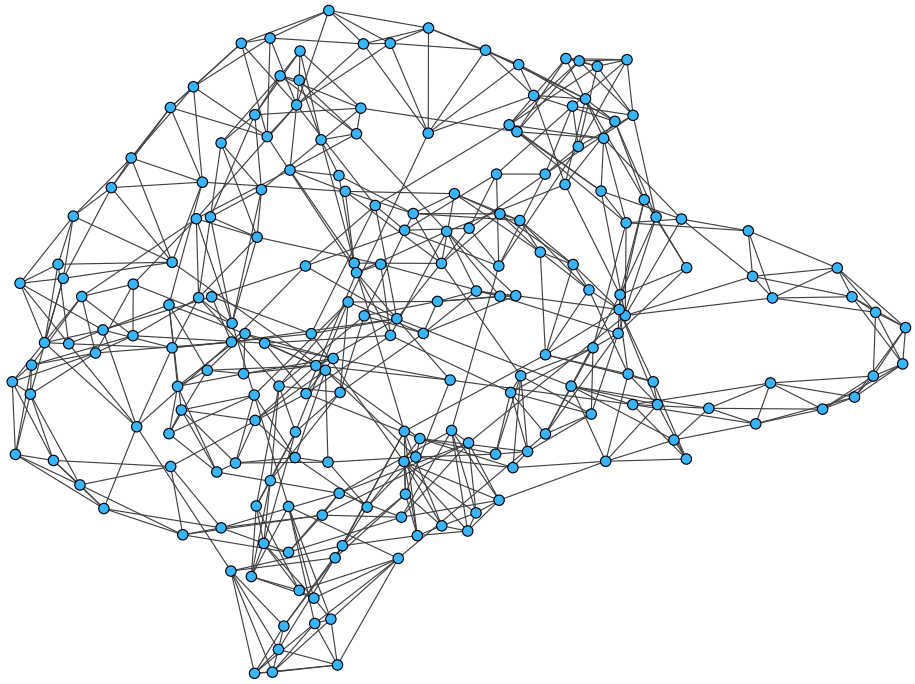


Figure 2.2: A random network with 200 nodes and 600 edges generated by the Watts/Strogatz model with parameter $p = 0.1$

distribution. For the $G(n, p)$ model, this can for example be easily understood when considering the generation of a graph as a random process consisting of $\binom{n}{2}$ steps, in each step rolling a dice whether to add an edge for a single pair of vertices. A randomly chosen vertex v participates in $n - 1$ of these Bernoulli trials, meaning that v's probability to obtain exactly k edges is given by the Binomial distribution as

$$\binom{n-1}{k} p^k \cdot (1-p)^{n-1-k}$$

Thus, in the limit of large graphs the degree distribution P_G of a random $G(n, p)$ graph converges to a Poissonian distribution $\frac{\lambda^k}{k!} e^{-\lambda}$ with the mean degree given as $\lambda = p \cdot \binom{n}{2}$. In particular, this means that the connectivity of graphs generated using the classical random graph models is - with high probability - homogeneous in the sense that nodes with significantly more or less links than the average are extremely unlikely. Since edges are rewired independently at random starting from a homogeneous ring topology, the same is true for networks generated by the Watts/Strogatz model. The fact that the connec-

tivity of nodes in both classical random graphs and Watts/Strogatz networks are rather homogeneous can easily be seen in in Figure 2.2 and Figure 2.1.

Networks with Power Law Degree Distributions

In recent years it has been observed that this homogeneity is uncommon for most examples of real-world networks. Instead, in many on these networks there exists a small number of nodes that has a significantly higher number of links than others. An example for a naturally emerging network where this heterogeneity is obvious is shown in Figure 2.3, where extremely well-connected nodes exist at the center of the network. It has been obtained by mapping the transcriptional regulatory interactions between 1333 genes of the Escherichia Coli bacterium [Gama-Castro *et al.*, 2008]. Since roughly a decade ago, the topologies of a number of heterogeneous networks in social, biological and technological systems have been studied. In these empirical studies, it has been shown that degree distributions of many of these networks - among them the World Wide Web, social networks, power grids, the Internet's router network, Peer-to-Peer overlays, co-authorship graphs, linguistic, neural and metabolic networks - seemingly follow a power law, that is

$$P_G(k) \propto k^{-\gamma} \tag{2.7}$$

for some real-valued exponent γ (see for example [Albert *et al.*, 1999; Broder *et al.*, 2000; Faloutsos *et al.*, 1999; Amaral *et al.*, 2000; Jeong *et al.*, 2000; Govindan and Tangmunarunkit, 2000; Liljeros *et al.*, 2001; Newman, 2001; Ripeanu *et al.*, 2002]). Again a number of simple stochastic network models have been proposed that aim at capturing this particular statistical feature. While we will review and classify some of these in section 2.3.5, in the following we first consider some of the interesting properties that have been attributed to power law degree distributions. Reconsidering for example the classical $G(n, m)$ model for random graphs, here we recall that - based on two fixed quantities n and m - a probability space is created which is "maximally random" in the sense that all graph realizations are equiprobable and in which Poissonian degree distributions emerge naturally. Although even this simple approach has yielded interesting results, their applicability to real-world networks with different statistical features (like for example power law degree distributions) is rather limited. Consequently suitable extensions to classical random graph theory have been studied [Molloy and Reed, 1995; Aiello *et al.*, 2000; Newman *et al.*, 2001; Chung and Lu, 2002]. Here probability spaces are considered which - rather than containing all networks with a certain, fixed number of links - consist of all possible networks that are consistent with a given predetermined sequence of degrees or expected degrees and in which all of these

Random Graphs with a Fixed Degree Distribution

"feasible" realizations are again equiprobable. Just like the classical models, this approach allows to study the probability that a randomly chosen network with a given degree distribution has certain properties. Based on their abundance in real-world systems, much of the work on random networks with fixed degree distributions has been focused on the complex structures and characteristics emerging in random power law networks.

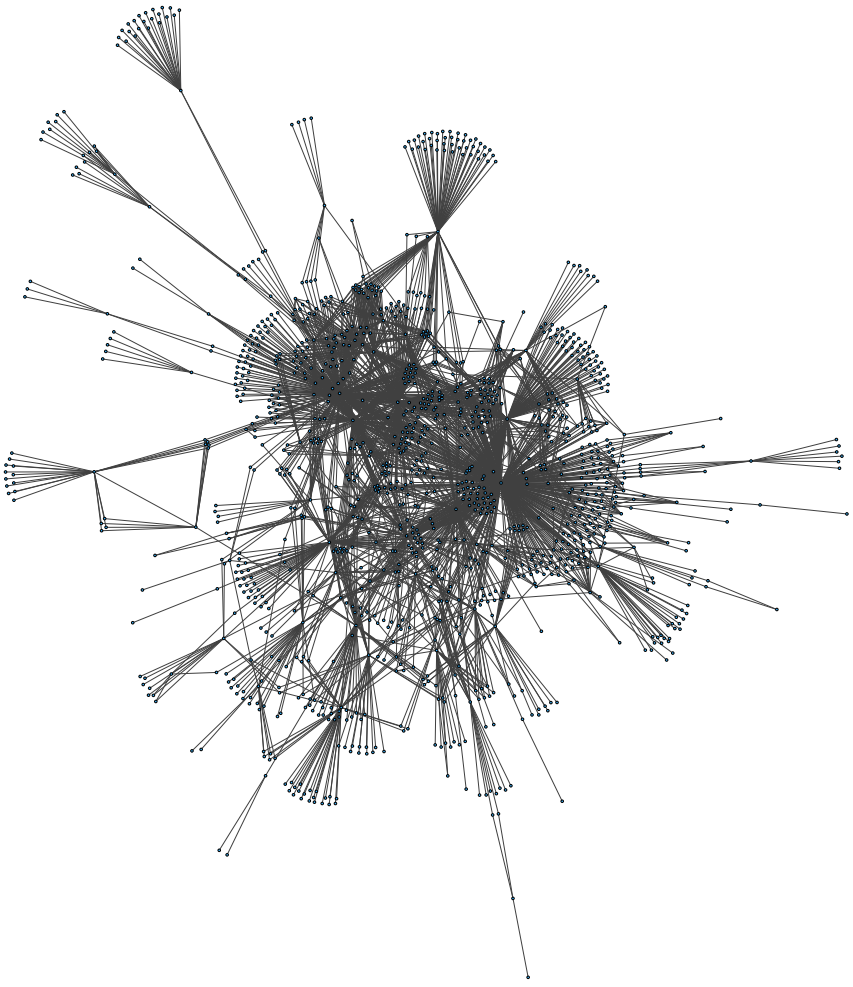


Figure 2.3: The gene regulatory network of Escherichia Coli with 1333 nodes and 2736 edges (network created using data from [Gama-Castro *et al.*, 2008])

2.3 Characteristics of Scale-Free Networks

Power Law vs. Exponential Distributions

Since in chapter 4 we will be involved with the construction, adaptation and monitoring of overlay networks whose degree distribution follows a power law, in the following section we lay the foundation for this work by reviewing a number of characteristics of so-called scale-free networks which are relevant in the design and operation of networked computing systems. Many of these characteristics can be understood intuitively when considering the properties of power law distributed quantities as opposed to - for example - quantities being distributed exponentially, that is the probability of a quantity k being proportional to $e^{-\lambda k}$ for some $\lambda \in \mathbb{R}$[1]. For a given quantity k and for some $c > 1$ we can then consider the frequency of a quantity $c \cdot k$ for both distributions. For a power law distribution as given in equation 2.7 one obtains

$$(c \cdot k)^{-\gamma} = c^{-\gamma} \cdot k^{-\gamma}$$

while for exponentially distributed quantities one comes to

$$e^{-\lambda(c \cdot k)} = \left(e^{-\lambda k}\right)^{c}.$$

Power Law Distributions and Rare Events

Thus, while the frequency of a c times larger quantity becomes quickly negligible in the case of exponentially distributed quantities, for power law distributed quantities it only decreases proportionally at a rate depending on the exponent γ. The implications of this difference become apparent when looking for example at natural phenomena. Assuming for instance $\gamma = 2.5$ and some quantity k with frequency 0.1, with $c = 100$ one obtains a frequency of 10^{-100} for the exponential as opposed to 10^{-6} for a power law distribution. We may take for instance as quantity k the magnitude of those earthquakes that occur on average once every 10 years within a certain region. If the magnitude of earthquakes were exponentially distributed, then an earthquake that is 100 times stronger will occur virtually never. It will however occur on average once every one million years if the magnitudes of earthquakes were power law distributed[2]. In short, one can say that power law distributions have no natural "scale" in the sense that the frequency of quantities away from a certain average value does not become quickly negligible. Their general shape being invariant under constant scaling they are said to be *scale-invariant* and in

[1] Please note the similarity to the tail of the Poissonian degree distribution of classical random graphs.

[2] Which seemingly they are according to the Gutenberg-Richter law [Gutenberg and Richter, 1949].

statistics they are routinely applied in the modeling of extraordinary events or catastrophes that are unlikely yet not impossible.

Power Laws in Finite, Discrete Systems. The Zipf Distribution

The reader may have noticed that we have so far used the phrase "power law distribution" in a rather intuitive and careless manner. Prior to discussing further properties of networks whose connectivity follows such a distribution, we shall thus consider a more formal definition. For this we assume a random variable X that can take discrete values from the set $\{1, \ldots, n\}$. A corresponding power law probability mass function P can then be given by the so-called Zipf distribution as follows.

$$P(X = k) = P(k) = k^{-\gamma} \cdot \left(\sum_{i=1}^{n} i^{-\gamma} \right)^{-1} \tag{2.8}$$

Here the first term determines the power law behavior with exponent $\gamma > 1$ and the latter term is the generalized, n-th harmonic number of γ which ensures the proper normalization of the probability mass function.

The Infinite Case: The Zeta Distribution

Many of the more subtle properties of power law distributions come into light as one considers very large systems. Being interested in the properties of networks at very large scales, it is justified to study the case where the maximum quantity n goes to infinity. Considering the random variable X as the degree of a node being chosen uniformly at random, for a network this means that the maximum possible degree approaches infinity. In this case, the probability mass function in equation 2.8 becomes the so-called Zeta distribution, which for $\gamma > 1$ is given as

$$P(X = k) = \frac{k^{-\gamma}}{\zeta(\gamma)} \tag{2.9}$$

with $\zeta : \mathbb{R} \to \mathbb{R}$ being the real-valued Riemann zeta function

$$\zeta(\gamma) = \sum_{i=1}^{\infty} i^{-\gamma}. \tag{2.10}$$

2.3.1 Explanatory Power of the Power Law Exponent

Moments of Power Law Distributions

In order to understand the influence of the exponent γ on some basic properties of the Zeta distribution one can study its (central) moments, the m^{th} moment M_m over x_0 being defined as:

$$M_m = \sum_{x=1}^{n} (x - x_0)^m \cdot P(X = x) \tag{2.11}$$

Thus, the first moment over $x_0 = 0$ yields the expected value $\bar{k}$ of the quantities in question, while their variance is given by the second (central) moment over $x_0 = \bar{k}$. Further interesting statistical properties like skewness and kurtosis are given by a distribution's third and fourth central moments.

Moments and Network Properties

When considering essential properties of networks with Zeta degree distribution (like for example connectedness, diameter, the resilience against random node removals or the behavior of dynamical processes like the spreading of information) one intuitively expects them to be influenced by the average degree or the degree distribution's variance and skewness. In fact, this intuition explains why the degree distribution exponent γ, although being a mere statistical parameter agnostic of any microscopic network details, significantly influences many of a network's large scale characteristics. Considering the m^{th} moment M_m over zero of the probability distribution $P(X = k)$, from equation 2.11 one obtains for the Zeta distribution:

$$M_m = \sum_{k=1}^{\infty} k^m \cdot P(k) = \sum_{k=1}^{\infty} \frac{k^{m-\gamma}}{\zeta(\gamma)} = \frac{\zeta(\gamma - m)}{\zeta(\gamma)} \tag{2.12}$$

Moments of the Zeta Distribution

Hence, for a network with Zeta degree distribution, the expected degree of a node chosen uniformly at random is

$$M_1 = E(X) = \frac{\zeta(\gamma - 1)}{\zeta(\gamma)} \tag{2.13}$$

which, in the corresponding p-series' convergence interval $\gamma \in (2, \infty)$, evaluates to:

$$M_1 = \zeta(\gamma - 1) \cdot \zeta(\gamma)^{-1} \tag{2.14}$$

Networks with diverging first Moment

For $\gamma \leq 2$ the first first moment is divergent which translates to the fact that - for an n-node power law network with exponent $\gamma \leq 2$ - the mean vertex

degree grows infinitely as $n \to \infty$. In particular, this means that the number of links necessarily grows asymptotically faster than the number of nodes, which is - considering the intended domain of overlay topologies for large networked computing systems - clearly not a desirable property. In contrast, for any $\gamma > 2$ the mean vertex degree converges to the constant $\zeta(\gamma - 1) \cdot \zeta(\gamma)^{-1}$, resulting in sparse networks in which the number of edges scales linearly with the number of nodes. Looking at the general case of the m^{th} moment M_m over zero, one finds that it converges for $\gamma > m + 1$, in this case evaluating to

$$M_m = \zeta(\gamma - m) \cdot \zeta(\gamma)^{-1} < \infty. \tag{2.15}$$

Scale-Free Networks

Considering the second central moment M_2 of a network's degree distribution, one thus finds that it is divergent for power law networks with an exponent $\gamma \leq 3$. Since the variance of a distribution can be given in terms of the second moment M_2 over zero[3], it's divergence translates to the fact that the magnitude of deviations from the expected degree increases infinitely as the size of the network grows. This formally underpins the notion of "scale-freedom" in the sense that in networks with diverging second moment the mean or expected degree does not represent a characteristic "scale" of connectivity. In the remainder of this dissertation we thus call networks with a Zeta degree distribution with finite first and infinite second moment *scale-free*[4]. From the above arguments we particularly find that - as opposed to classical random graphs - in scale-free networks nodes with hundreds or even thousands of links are likely to exist given that the network is sufficiently large.

Based, among others, on the convergence behavior of the m-th moments of the Zeta distributions, in the last couple of years a number of properties have been derived for scale-free networks. In the following we recall some of the results that have been obtained for *uncorrelated* random scale-free networks. In the language of random graph theory, these results thus hold asymptotically almost surely for a network being drawn at random from the set of all networks consistent with a particular degree distribution. Stochastic models that are suitable to generate networks with a scale-free degree distribution will be considered in section 2.3.5 and - in the context of a practical distributed construction procedure for random scale-free overlay topologies - in chapter 4.

[3] In the case of a finite expected value M_1, the variance simply evaluates to $M_2 - (M_1)^2$

[4] We will use this definition throughout this dissertation although being aware that there exist alternative and more sophisticated notions like for instance those introduced in [Li *et al.*, 2005]

2.3.2 Structural Properties

The Molloy-Reed Criterion

A number of important results about the influence of a degree distribution's moments on basic network properties are due to a result stemming from the study of random graphs with fixed degree distributions that has been presented in [Molloy and Reed, 1995]. In simple terms, the so-called Molloy-Reed criterion states that a random network whose degree distribution is such that

$$\frac{M_2}{M_1} > 2 \tag{2.16}$$

has almost surely a giant connected component, while there is almost surely none if equation 2.16 does not hold. For networks with Zeta degree distribution with exponent γ, and in the limit of large networks, the Molloy-Reed criterion evaluates (with equation 2.15) to

Giant Connected Component in Scale-Free Networks

$$Q(\gamma) := \frac{\zeta(\gamma-2)}{\zeta(\gamma-1)} > 2. \tag{2.17}$$

As has been argued for example in [Aiello *et al.*, 2000] and as indicated by the plot of $Q(\gamma)$ that is shown in Figure 2.4 this is the case for exponents below a critical point $\gamma_c \approx 3.4787$. More generally, the Molloy-Reed criterion necessarily holds if the second moment diverges and the first moment converges and thus for any exponent γ in the scale-free range between two and three [Dorogovtsev and Mendes, 2003].

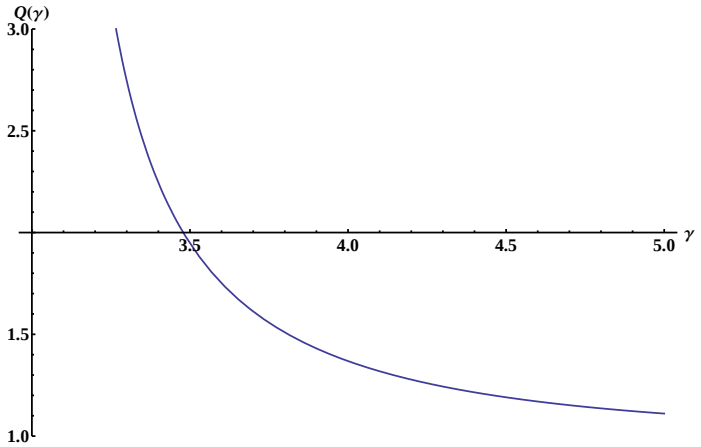


Figure 2.4: $Q(\gamma)$ from equation 2.17 in random scale-free networks

Scale-Free Networks are super-resilient

Beyond connectedness, an important property of a random network that can be analyzed by means of equation 2.17 is a network's resilience against random faults. For this one can consider a simple model in which a number of nodes and all incident edges are removed from a network uniformly at random. One can then study the number of nodes that need to be removed until the network loses connectedness or - tolerating a constant fraction of isolated nodes - at which the giant connected component disappears. Intuitively, the first moment of a network's degree distribution plays a crucial role for the resilience against random node removals since it determines the expected number of stale links that are left after a single node is removed from the network at random. By quantifying the effect of random node removals on a network's degree distribution and using the Molloy-Reed criterion, in [Cohen *et al.*, 2000] it has been shown analytically that, in random networks with a Zeta degree distribution, at least a fraction

$$r := 1 - (Q(\gamma) - 1)^{-1} \tag{2.18}$$

of all nodes needs to be removed in order to destroy the giant connected component. In particular, for scale-free networks with diverging second moment, it follows that $r \to 1$ since $Q(\gamma) \to \infty$. Hence for random power law networks with $\gamma \in (2,3)$, asymptotically a number of nodes of the order of the network's size must be removed at random to destroy the giant connected component. Contrariwise for random Erdös/Rényi graphs, removing a fraction $r \ll 1$ suffices to let the network fall apart, the exact value of r depending on the parameters n, m and p of the $G(n,m)$ and the $G(n,p)$ model.

Super-Resilient Overlay Networks

Considering an application for example in large, dynamic networked computing systems, this so-called *super-resilience* of random scale-free networks against random node removals is clearly interesting. As long as the machines in such a system fail at random, the exponent γ of the Zeta degree distribution lies between two and three and the network is sufficiently large, almost all nodes will remain connected almost surely. This favorable behavior of random scale-free overlays is actually even more pronounced when considering non-uniform failure models in which nodes with higher degrees are less likely to fail than those with small degrees. By means of appropriate construction procedures (considering e.g. an element's past behavior) or a selective protection or replication of most connected nodes, network topologies can actually be built in a way that this condition holds.

Scale-Free Networks with $\gamma \approx 2$ are super-susceptible

In short, the remarkable robustness of scale-free networks against random failures is based on the fact that - with high probability - a failure will not affect one of the few most connected nodes. Clearly the opposite is true when considering failure models in which the most connected nodes are removed preferentially. Such models can be used to evaluate a network's susceptibility against sabotage, that means scenarios in which an informed individual intentionally brings down the most connected and thus must important nodes. Again by quantifying the effect of such an attack on a network's degree distribution, in [Cohen *et al.*, 2001] the fraction r of most connected nodes that can be removed until the giant connected component is destroyed has been derived for random scale-free networks. Here it has been argued that, as a network's exponent γ approaches two, the fraction r of most connected nodes that needs to be removed to disintegrate the network goes to zero in the limit of infinite networks. One may thus call such networks *super-susceptible* against targeted attacks. In contrast, for $\gamma > 2$ a critical fraction $r_c \gg 0$ can be derived that grows gradually as γ is increased. This is illustrated in Figure 2.5, which shows randomly chosen scale-free networks with exponents 3.5 and 2.001 after random nodes (a-b) and most connected nodes (c-d) have been removed.

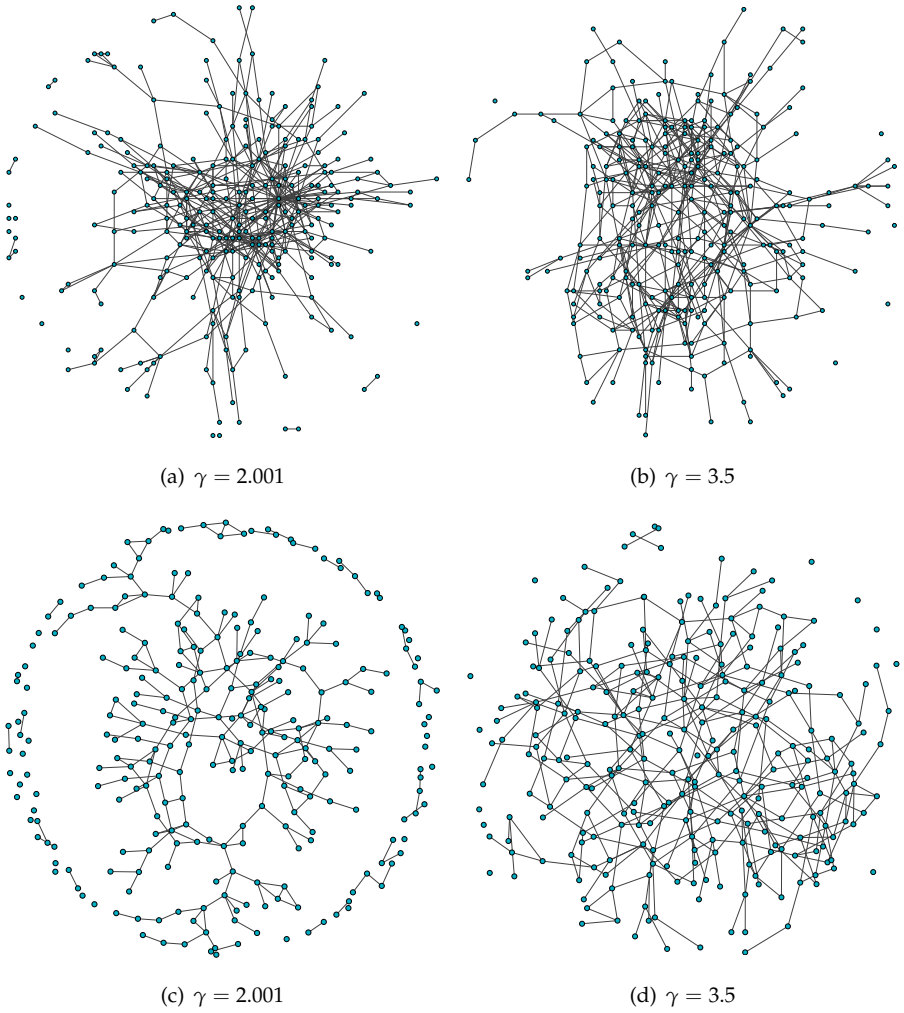
Scale-Free Networks are ultra-small

Apart from resilience against faulty constituents and their susceptibility against attacks, a number of further characteristics of random scale-free networks can be attributed to the moments - and thus the exponent - of their Zeta degree distribution. Being particularly interesting for the design of large scale overlay topologies for networked computing systems, in [Cohen and Havlin, 2003] it has been argued that with high probability the diameter of random scale-free networks with exponents $\gamma \in (2,3)$ is proportional to $ln(ln(n))$ and thus *ultra-small*. In practical terms this translates to the appealing fact that the length of a path between any two nodes in a random scale-free network may be considered virtually constant at least for any network size that can practically occur in real-world networked computing systems.

2.3.3 Routing, Search and Diffusion Phenomena

Distributed Search in Scale-Free Networks

Above we have seen that ultra-short paths exist between nodes in random scale-free networks. For practical distributed computing systems that are interconnected via network topologies that exhibit scale-free characteristics, the mere existence of such paths is rather uninteresting if there are no efficient means by which these can be found and used. Related questions have thus been studied in the context of scalable distributed search and decentralized



(a) $\gamma = 2.001$ (b) $\gamma = 3.5$

(c) $\gamma = 2.001$ (d) $\gamma = 3.5$

Figure 2.5: Effect of removing 30 random (top row) and 30 most connected (bottom row) nodes in 300 node random networks with a Zeta degree distribution with exponent $\gamma = 2.001$ (left column) and $\gamma = 3.5$ (right column). Networks have been layouted using the Kamadai-Kawai algorithm.

routing schemes. In order to locate content in an ad-hoc fashion without requiring the explicit construction of indexing structures, early Peer-to-Peer file-sharing services like GNUTELLA, whose overlay topology reportedly shared some of the features of scale-free networks [Ripeanu *et al.*, 2002], used simple flooding mechanisms by which search requests were propagated through the network until they eventually arrived at a node with access to the object in question. While the time taken by this strategy to locate a certain piece of content benefits from the existence of ultra-short paths in scale-free networks,

flooding the network with requests is clearly not a scalable solution. More efficient probabilistic search strategies for scale-free networks have been investigated throughout the last couple of years.

Efficient Distributed Search Strategies

A simple improvement over flooding strategies is the use of local schemes by which requests are forwarded only to a single neighbor that is either chosen randomly or based on some node characteristic like for example its degree. Different strategies of forwarding requests have been considered in [Adamic *et al.*, 2001], where it is assumed that each node has knowledge about the locally stored objects, its one-hop neighbors as well as the objects stored at these neighbors. If none of the current neighbors has stored the searched object, a search strategy has been evaluated in which the request is forwarded to the (unvisited) neighbor with highest degree. In scale-free networks with exponent $\gamma \in (2, 2.3)$ it has been shown in [Adamic *et al.*, 2001] that - once the random walk has reached the node with maximum degree - this simple distributed scheme approximately visits nodes in a decreasing order of their degrees. For a scale-free network with n nodes and exponent $\gamma \in (2, 2.3)$, in [Adamic *et al.*, 2001] it is further argued that with high probability the search time - that is the number of steps taken until an object is found - scales as $n^{2-\frac{4}{\gamma}}$.

Making use of Hubs

This implies that distributed search strategies which are biased towards high-degree nodes are most efficient in scale-free networks with exponents close to two. In fact this is not surprising since in these networks hubs are most pronounced and the existence of nodes with massive numbers of neighbors also results in large portions of the network being searched in a single step. Since this inevitably results in a vast majority of requests being handled by a small fraction of most connected nodes, this can however also question the scheme's usability in a number of practical systems. A number of similar search strategies have thus been suggested that try to mitigate this problem for example by a replication of biased or unbiased random walkers and objects [Lv *et al.*, 2002] or the use of parallel probabilistic broadcasts [Boykin *et al.*, 2004] of search requests. For the latter scheme, the search time in scale-free networks has been shown to scale as $log^2(n)$, while the number of required messages scales asymptotically as $n^{\frac{1}{\gamma}}$ where n denotes the number of nodes and γ is the exponent of the degree distribution.

Greedy Routing in Scale-Free Networks

Closely related to the question considered above, the provision of scalable distributed routing schemes in large, dynamic networks is a further problem of prime importance for practical networked computing systems. Here, an important question with a large number of potential practical applications is how short paths between arbitrary node pairs can reliably be found in an ad-hoc

fashion based on the local view of nodes. In [Papadopoulos *et al.*, 2010] it has recently been argued that scale-free network structures can emerge naturally as simple geometric graphs evolving between nodes embedded in hidden hyperbolic spaces. Based on earlier work on the performance of greedy routing in hyperbolic spaces that has been presented in [Kleinberg, 2007], it has further been argued in [Papadopoulos *et al.*, 2010] that, in large and dynamic scale-free networks that are suitably embedded in hyperbolic spaces, a simple greedy heuristic can be used to provide efficient distributed routing with almost sure delivery and small routing stretch[5].

Spectral Properties of Scale-Free Networks

Above we have briefly mentioned the usability of random walk strategies for the provision of distributed search in large scale networks. Whether the use of such random walk schemes is feasible in a particular network critically depends on the topology's so-called mixing properties which determine the number of steps the random walk needs to take in order to converge to its stationary distribution. While we will address this issue in more detail both analytically and empirically in chapter 4, for the moment it is sufficient to say that these properties of a network basically rely on how "well-connected" it is, that is whether there exist long shortest paths and small cuts that could prevent a fast convergence of random walks. While a general analytical treatment of these properties for example in terms of *algebraic connectivity*, *conductance*, *graph expansion* and *spectral properties* of Laplacian and stochastic matrices [Lovász, 1993; Chung, 1997; Bollobás, 1998; Hoory *et al.*, 2006] is generally complicated for large random networks, there exists a large body of evidence demonstrating that random scale-free networks are in fact with high probability "well-connected". This property of scale-free networks allows the use of random walks not only for distributed search but also for the sampling of random nodes [Zhang *et al.*, 2008] or the construction and adaptation of randomized network topologies [Scholtes, 2010]. Details of the latter scheme will be presented in chapter 4 of this dissertation.

Diffusion and Consensus Phenomena in Networks with Scale-Free Features

Apart from the convergence time of random walks, the good mixing properties of scale-free networks also influence the performance of a number of other dynamical processes which are - just like random walk schemes - essentially based on the diffusion of information and which are of prime importance for networked computing systems. One example are gossip-based algorithms which - along with random walk mechanisms - constitute an important building block for the design of dynamic, decentralized and self-

[5]Routing stretch being defined as the ratio between the length of the path discovered by the routing algorithm and the shortest path in the network.

organized computing systems. During the last years, gossiping strategies have been proposed for the provision of scalable multicast [Gupta *et al.*, 2006], the maintenance of database replicas [Demers *et al.*, 1987], the efficient computation of network-wide aggregate functions [Kempe and McSherry, 2004; Jelasity and Montresor, 2004], the sampling of random nodes [Jelasity *et al.*, 2007], the distributed collaborative filtering of information [Bakker *et al.*, 2009] or the management of overlay topologies [Jelasity *et al.*, 2009]. Another class of dynamical processes which critically depends on network topologies with good mixing properties are those resulting in self-organized consensus phenomena like flocking, synchronization, opinion formation and collective decision-making schemes frequently used in systems with distributed control. Such phenomena can again be viewed as essentially being based on the diffusion of information through the network as well as a feedback mechanism by which the diffusing information and the behavior of individual elements interact. It has been argued that the structures of scale-free networks facilitate the emergence of consensus in a number of different contexts (see [Barrat *et al.*, 2008] for a review). An example for an algorithmic scheme which is based on a self-organized synchronization mechanism and the good mixing found for example in scale-free networks has been presented in [Scholtes *et al.*, 2009; Scholtes *et al.*, 2010] and will be discussed in more detail in chapter 5 of this dissertation.

Detrimental Spreading Phenomena

Apart from using the good connectedness or expansion of random scale-free networks for distributed schemes being based on gossiping, random walk or consensus mechanisms, there also exist a number of detrimental phenomena that benefit from these properties. Examples include the proliferation of worms, viruses and malicious content in computer networks, the pollution of Peer-to-Peer file-sharing systems with corrupted data [Thommes and Coates, 2005] or the propagation of failures. Such phenomena can again be viewed as particular types of diffusion processes and relations to structural properties of the network topology upon which they operate can be studied. The fact that many technical networks are affected by these processes and the finding that many technical systems seemingly exhibit scale-free network structures underpin the importance of a sound understanding of topological features that influence the dynamics of spreading processes. Some of these issues can be addressed by means of rather simple epidemiological models like for example the SIS and the SIR models that have been introduced in [Kermack and McKendrick, 1927; Britton and Adler, 2004]. To evaluate their efficiency, one can define the epidemic threshold λ as the minimum proportion $\frac{i}{c}$ of newly in-

fected nodes i and cured nodes c per time above which an epidemic remains in the population without requiring external infections. In [Pastor-Satorras and Vespignani, 2001; Blanchard *et al.*, 2002] it has been shown for the SIS model that, in random scale-free networks, the epidemic threshold λ depends on the first and second moments of the network's degree distribution and thus the degree distribution's exponent γ in the following way:

$$\lambda = \frac{M_1}{M_2} = \frac{\zeta(\gamma - 1)}{\zeta(\gamma - 2)} \tag{2.19}$$

Again using the convergence and divergence of moments, for scale-free networks with $\gamma \in (2, 3)$ and thus a finite first and a divergent second moment one obtains $\lambda \to 0$ in the limit. This can be reformulated to random scale-free networks with exponent $\gamma \in (2, 3)$ becoming infinitely susceptible to epidemic processes in the limit of large networks. However, positively spoken it also results in an asymptotically infinite growth of the efficiency of probabilistic information spreading schemes. While many more aspects of dynamical processes in scale-free networks could be covered, here we end our summary and refer the interested reader to the excellent reviews given for example in [Boccaletti *et al.*, 2006; Barrat *et al.*, 2008].

2.3.4 Finite-Size Effects and Correlations

Applicability to Engineered and Natural Networks

Having reviewed a number of interesting findings that facilitate a substantiated reasoning about the properties of random scale-free networks and the dynamics unfolding within them, it is clearly tempting to apply these results to the large body of real-world networks for which scale-free characteristics have been identified. In fact, during the last decade a number of strong claims have been made about the properties of technical systems (like for example the Internet's alleged fragility against sabotage [Cohen *et al.*, 2001]) which are essentially based on theoretical results about random scale-free networks. However, at this point we take the opportunity to speak a word of caution regarding the applicability of these results to natural and engineered systems.

Finite-Size Effects in Scale-Free Networks

First of all we have seen that many of the findings discussed above are based on arguments about the convergence or divergence of moments of the Zeta distribution and thus hold in the limit of *infinitely large* scale-free networks. Naturally, such limit behavior is rather of theoretical interest since networks occurring in reality - although possibly being very large - are necessarily finite. If one wants to apply arguments that hold in the infinite

limit to practical, finite systems, it is important to consider the rates of convergence or divergence which determine possible deviations from the limiting behavior. An exact evaluation of these deviations with respect to a particular property or finding must be carefully considered when designing a particular system. Nevertheless, in [Dorogovtsev and Mendes, 2000; Dorogovtsev and Mendes, 2003] it has been argued that some general statements about the magnitude of these deviations in finite, random scale-free networks can be made. For the proportion

$$\frac{M_1}{M_2}$$

between the first and second moment of the degree distribution which occurs both in the epidemic threshold in equation 2.19 as well as (inversely) in the Molloy-Reed criterion leading to equations 2.17 and 2.18, it has been argued in [Dorogovtsev and Mendes, 2003] that in a finite random scale-free network the deviation from the infinite limit behavior depends on the network's degree distribution exponent. In particular, for random scale-free networks with γ close to two, finite-size effects quickly decrease as the network size increases while for networks with exponents close to three the deviations from the limiting case are much more pronounced and persist even for large networks.

Correlations in Networks

When wanting to apply the findings above to practical systems, apart from finite-size effects, another pitfall that must be avoided is the possible effect of correlations that can be introduced by construction processes by which real-world scale-free networks emerge. At this point its is important to remind the reader of the fact that - strictly speaking - all of the findings summarized above only hold for networks which are drawn from a probability space of networks in which all networks with a fixed Zeta degree distribution are assigned equal probability regardless of any further topological features. It is clearly only in this case that the probability of a network exhibiting certain qualities is completely determined by its degree distribution. Contrariwise, actual construction procedures - although being probabilistic - may preferentially produce a particular kind of network with a certain degree distribution while other networks with the same degree distribution are less likely or even occur virtually never. In real-world network topologies this can be studied by considering - apart from node degrees - further statistical properties and testing for correlations between them. Examples for correlations that are frequently considered in the study of complex networks are the clustering coefficient or the distribution of degrees of nearest-neighbor nodes. While in the infinite limit, such correlations are naturally absent for random scale-free networks, when want-

ing to reason about the properties of real-world networks it is crucially important to explicitly regard the potential effects of any correlations that may be due to the construction processes by which they emerge. Alternatively, when wanting to rely on the rich body of theoretical results on random and thus uncorrelated scale-free networks in the engineering of networked computing systems, it is necessary to explicitly consider randomized construction procedures that eliminate correlations. Practical aspects of such an approach which explicitly employs randomness to form scale-free overlay topologies in which construction-dependent correlations are absent will be considered in more detail in chapter 4.

2.3.5 Evolution of Scale-Free Structures

Emergence of Power Law Distributions

The properties of scale-free networks and their apparent abundance in natural, technical and social systems have resulted in significant interest in possible mechanism by which they emerge. Referring to the scale-invariant properties of Zeta distributions that have been discussed above, certain aspects of this question can be related to earlier studies of mechanisms by which scale-invariant, fractal features in physics and biology emerge. In particular, similarities between the evolution of scale-free networks and non-equilibrium processes resulting in fractal tempo-spatial structures [Mandelbrot, 1982; Bak *et al.*, 1987; Vicsek, 1992] point to surprising relations between non-equilibrium statistical physics and the study of random networks. While in the remainder of this section we focus on models for scale-free networks, excellent interdisciplinary reviews on the question how scale-invariant distributions arise in more general contexts can be found for example in [Mitzenmacher, 2002; Caldarelli, 2007]. Due to the vast number of different models that have been proposed during the last couple of years, here we can only review a small selection of models that are relevant to questions being considered later in this dissertation.

Growth-Based Models: Preferential Attachment

A rather natural way to model the emergence of networks in practical systems is by considering their temporal evolution. An important class of such models are those that incorporate a growth process in terms of nodes or links being added to the network one by one. A simple growth mechanism by which scale-free networks emerge is the *preferential attachment* scheme that has been studied in [Barabási and Albert, 1999]. In this model, nodes enter the network one at a time, each new node drawing k links to existing nodes. The probability p that one of these k new links is drawn to an existing node with degree d is given as

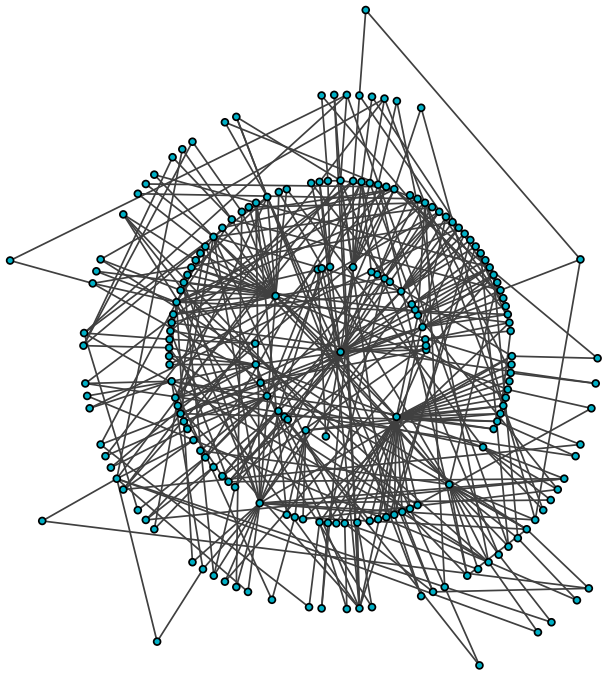


Figure 2.6: A network with 200 nodes and 384 edges generated by the Barabási/Albert model

$$p = \frac{d}{2m}$$

where m is the current number of edges in the network. An example network resulting from this construction procedure is shown in Figure 2.6.

As shown by a mean-field approximation in [Barabási and Albert, 1999] and more rigorously later in [Bollobás *et al.*, 2001], this model generates networks with a power law degree distribution with an exponent $\gamma = 3$. In [Dorogovtsev *et al.*, 2000b], a more general model has been considered in which the probability that a new node draws a link to an existing node with degree k is given by

General Linear Preferences

$$p = \frac{d + A}{\sum_k (k + A)}$$

where A is some positive constant which influences the strength of the preference. For this it has been shown that the resulting networks have power law degree distributions with the exponent γ depending on the particular choice of A. While preference functions are rather theoretic constructions, different

mechanisms have been considered that may underlie this preference in practical systems. A simple explanation for the emergence of degree-based preferences are random walk schemes, in which joining nodes follow a random path starting from an arbitrary initial node. Depending on the random walk strategy and length, as well as the exact mechanism by which links are created, it has been argued that such schemes can give rise to networks with power law degree distributions [Vazquez, 2000; Saramaeki and Kaski, 2004]. Another class of models that can be viewed as producing degree-based linking preferences are those incorporations copying mechanisms. Here, joining nodes copy a fraction of links of existing nodes that are chosen at random. Depending on the exact type of copying mechanism as well as parameters like the fraction of links to copy, such models have been shown to be able to produce power law networks with exponents in the scale-free range [Dorogovtsev *et al.*, 2000a; Vazquez *et al.*, 2001]. All the models mentioned so far are based on some kind of stochastic process. Contrariwise, in [Barabási *et al.*, 2001], a simple deterministic procedure has been considered which resembles iterative construction procedures of fractal geometric structures. With $\gamma = 1 + \frac{ln(3)}{ln(2)} \approx 2.585$, the degree distribution exponent of the networks resulting from this scheme has again been found to be in the scale-free range.

Copying, Random Walk and Deterministic Models

Non-Equilibrium Models for Scale-Free Networks

All of the models considered above - stochastic or deterministic - involve a growth process in which the different age of nodes drives the heterogeneity of the resulting degree distributions. In simple words, "older" nodes have more chances to receive links from joining nodes and are thus more likely to become hubs. If the preference of linking to old and well-connected nodes is large enough to compensate for the increase in options that results from the network's growth, hubs can experience virtually unlimited growth in connectivity by means of a self-energizing process that is frequently paraphrased as a "rich-get-richer" or Matthew effect. In the language of physics, the growth mechanism that drives the evolution of scale-free structures in such models can in fact be viewed in terms of a *non-equilibrium process* by which the system *self-organizes* into a state with scale-invariant characteristics and complex structural features. Such non-equilibrium processes can intuitively be related to the self-organized formation of different kinds of complex features in natural systems, which is why it seems rather natural to study network models that explicitly incorporate growth. At the same time, there exist a number of situations where such growth-based mechanism appear to be inadequate. First of all, as argued for example in [Dorogovtsev and Mendes, 2003] growth processes inevitably introduce correlations. To give a simple example, in the preferential attach-

Benefits of Equilibrium Models

ment model mentioned above it is clearly impossible that a node v_{last} that joins the network in the last step of the iterative construction process could possibly be a highly connected hub since - apart from the k links it has drawn by itself - it had no chance to acquire any further connections. Contrariwise, in the probability spaces that underlie the results about random scale-free networks, realizations in which v_{last} is a hub are just as likely as any other configuration which is why one needs to take care when wanting to apply theoretical findings on networks resulting from non-equilibrium growth processes. Apart from these rather subtle and technical difficulties, we can further clearly imagine systems in which a correlation between a node's age and its likelihood of becoming highly connected is not desirable. We may further want to consider the evolution or adaptation of scale-free structures in steady-state regimes in which growth is absent. In the language of physics such conditions can be viewed as *equilibrium situations*.

Equilibrium Models for Scale-Free Networks

During the last couple of years, different kinds of *equilibrium models* have been considered which are able to produce scale-free networks. Since there is no growth in these models, one needs to come up with alternative mechanisms by which the nodes' heterogeneous connectivity is introduced. A natural way to accomplish this is by explicitly ascribing (numeric) node characteristics that represent fitness, capacity, popularity or other heterogeneously properties. In [Caldarelli *et al.*, 2002a] it has been shown that when nodes v_i are assigned characteristics $p(v_i)$ according to a sufficiently skewed distribution p and when links are created between pairs of vertices v_i, v_j with probability $p(v_i) \cdot p(v_j)$ the resulting networks will have power law degree distributions. [Baiesi and Manna, 2003] introduced a model where a local rewiring of links favored connections between nodes whose degrees differ as much as possible. In a certain parameter range, this dynamics eventually resulted in networks with power law degree distributions. In [Goh *et al.*, 2001] and similarly in [Chung and Lu, 2002] a model has been considered in which nodes are assigned Zipf-distributed node weights and links are being created with probabilities proportional to the product of these weights. In this model random, uncorrelated power law networks evolve with the degree distribution exponent depending on the Zipf distribution assigning initial node weights. Since it provides the basis for the distributed adaptation scheme introduced in chapter 4, this model will be considered in more detail later. Many more equilibrium and non-equilibrium models producing scale-free networks exists and for a more exhaustive presentation we refere the reader to [Dorogovtsev and Mendes, 2003; Boccaletti *et al.*, 2006; Caldarelli, 2007].

2.4 Complex Networks - An Equilibrium Statistical Mechanics Perspective

Statistical Mechanics and Complex Networks

Considering networks as an (admittedly gross) abstraction for arbitrary systems consisting of interacting elements, it is not surprising that their study has raised significant interest in a number of scientific disciplines. In particular, in the study of the characteristics of complex networks a number of surprising analogies to systems being considered in statistical physics have been uncovered. We have already briefly mentioned that some properties of networks resulting from the simple growth-based model considered in [Barabási and Albert, 1999] resemble scale-invariant, fractal geometrical features that can be found in a variety of natural systems. The observation of mere statistical similarities may result in these analogies to be considered rather superficial and sketchy. However it has since been argued for example in [Albert and Barabási, 2002; Berg and Lassig, 2002; Dorogovtsev *et al.*, 2003; Farkas *et al.*, 2004; Garlaschelli *et al.*, 2006] that there exist surprisingly substantial analogies between both equilibrium and non-equilibrium statistical mechanics and complex networks. These and related works have facilitated to study the emergence of complex structures in networks in the well-established framework of statistical mechanics, thus allowing to argue in terms of *particle ensembles* and macroscopic *thermodynamic quantities* like energy, volume, chemical potential or temperature. In the following, we introduce some of the most basic notions of equilibrium statistical mechanics only in so far as required to motivate some of the relations between overlay topologies and thermodynamic systems that will be pointed out in more detail in the subsequent chapters 3 and 4.

Isolated Equilibrium Systems

Prior to establishing these analogies we first introduce some key abstractions being used in the study of many-particle systems in the field of statistical mechanics. For this we start by considering a system S that consists of N particles and that is completely isolated from its environment. In Figure 2.7 such a system S is shown along with its surrounding. Although being impossible in practice, in a thought experiment one can precisely define the state of the system S at a certain point in time by specifying a vector containing positions and velocities of all N particles. Such a precise, microscopic description is called a *microstate* of the corresponding particle system. Clearly, for any practical purpose a microstate description is neither possible nor informative since one is rather interested in a system's bulk properties than in precise particle positions and velocities.

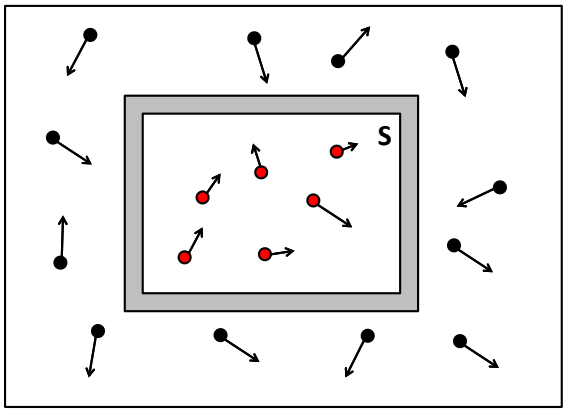


Figure 2.7: Microcanonical ensemble

Macrostates of Particle Systems

These complexities can be reduced when limiting oneself to a description in terms of aggregate, macroscopic parameters like volume, energy, particle number or temperature. Such a description of a thermodynamic system is called a *macrostate* and can be thought of as a set that subsumes all possible microstate realizations of a particle system that are consistent with certain fixed macroscopic parameters. For the completely isolated system S shown in 2.7, the possible microstates are determined by the number of particles N, the volume V of the system (determining the number of possible particle positions) and the total energy E (influencing the possible assignments of particle velocities).

Probability Measure

Technically, a macrostate defined in terms of the macroscopic parameters N, V and E subsumes a set of possible microstates that are consistent with these fixed quantities. In order to reason about particle systems in terms of statistics, such a macrostate can be turned into a probability space by defining a measure which assigns each possible microstate r a probability P_r. To understand what this means, one can view a particle system as a dynamical system. In general particles will move and collide, hence changing positions, velocities and transferring energy between each other. The probability P_r can then be interpreted as the probability to find the system - in any given moment - in microstate r.

Equilibrium Postulate

Here the notion of a *thermodynamic equilibrium* comes into play. While one could imagine many different ways of assigning a microstate r a probability P_r, here we restrict ourselves to equilibrium systems in which all microstates with identical energy are ascribed equal probability. In physical systems, the second law of thermodynamics ensures that any isolated system will eventually reach such a state. Due to all microstates being equally likely, this state maximizes entropy and minimizes the amount of knowledge we have about the system's

detailed internals. In this equilibrium situation, for a macrostate consisting of $\Omega(N, V, E)$ possible microstates, the probability to find the system in a particular microstate r is given by:

$$P_r(N, V, E) = \frac{1}{\Omega(N, V, E)} \tag{2.20}$$

Statistical Ensembles and Random Networks

In statistical mechanics, the probability space defined by a macrostate and an associated probability measure is called a *statistical ensemble*. In particular, above we have considered the situation of a completely isolated system in thermodynamic equilibrium for which the resulting statistical ensemble is called the *micro-canonical* ensemble. At this point a first analogy between particle systems and the study of classical random graphs can be identified. Just like the $G(n, m)$ model defines a set of all possible networks with a fixed number n of nodes and a fixed number m of edges in which all networks have probability $\binom{n}{m}^{-1}$, the above ensemble is a collection of all possible configurations of particle positions and velocities with fixed aggregate quantities N, V and E, each of the feasible configurations being assigned equal probability.

Equilibrium Systems with Energy Exchange

Real-world particle systems are hardly ever perfectly isolated. Usually systems can exchange energy or even particles with their surrounding. In a first step, one can relax the condition of a perfectly isolated system by allowing an exchange of energy with its environment. Such a situation is shown in Figure 2.8. For this, one assumes an impermeable border at which particles can exchange energy for example due to particle collisions. For the moment, we do however assume that no particle can enter or leave the system. When particles of the environment and the interior system S collide at the border (as depicted in Figure 2.8), momentum can be transferred between particles in- and outside the system and thus energy is allowed to leave or enter the particle system S. We thus further assume that the system is in a thermal equilibrium with is environment in the sense that the inflow of energy is balanced with the outflow.

Canonical Ensemble

Just like for the micro-canonical ensemble one can now again consider thermodynamic parameters that determine the macrostate of such a particle system. In the situation depicted above, both particle number N and volume V are - just like in the case of a completely isolated systems - fixed. In contrast, the total energy E of the system can now vary over time. However since a balanced in- and outflow of energy is assumed, one can reasonably define a time-invariant average energy around which the total energy fluctuates. In physical systems, both the average energy as well as the degree of fluctuation

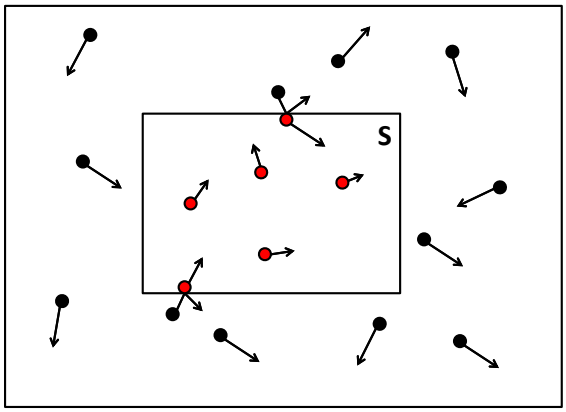


Figure 2.8: Canonical ensemble

around the average value are given by the temperature T, thus completing the definition of a macrostate of what is called the *canonical* ensemble.

Boltzmann Distribution

What is left for the definition of a probability space is to assign a probability $P_r(N, V, T)$ to each microstate r. Different from the case of the micro-canonical ensemble, such a macrostate contains microstates with all possible energies. We have argued above that, in a state of equilibrium, all microstates with a certain total energy are equiprobable. We have further argued that we assume a time-invariant average energy. By mere combinatorial arguments about the number of possible microstate realizations with a certain total energy, it is possible to derive the Boltzmann distribution. It gives the probability $P_r(N, V, T)$ to find the system in a microstate r with energy E_r as[6]

$$P_r(N, V, T) = \frac{1}{Z(N, V, T)} \cdot e^{-\frac{E_r}{T}} \tag{2.21}$$

Here, $Z(N, V, T)$ is a so-called partition function which normalizes the probability distribution to 1, that is:

$$Z(N, V, T) = \sum_r e^{\frac{-E_r}{T}} \tag{2.22}$$

Equilibrium Systems with Energy and Particle Exchange

We finally consider a situation in which a system is allowed to exchange both energy and particles with its environment. Such a situation is depicted in Figure 2.9. Just like in the canonical ensemble, we again assume an equilibrium situation in which the in- and outflows of both energy and particles

[6]Usually, equation 2.21 also involves the Boltzmann constant k_B which gives a unit-dependent proportionality of temperature and energy. Here we are not interested in any particular units. Assuming for example Planck natural units allows us to omit this constant.

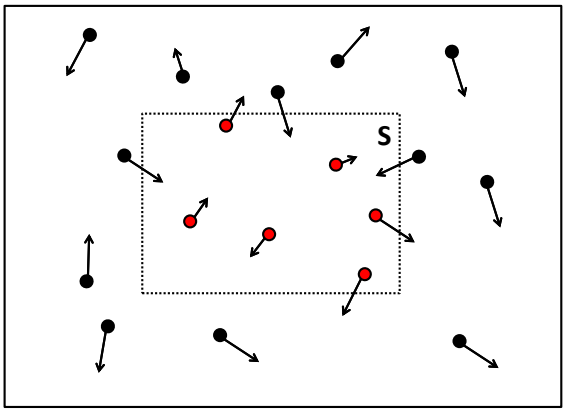


Figure 2.9: Grand canonical ensemble

are balanced. In addition to the fixed temperature T we further assume that the average number of particles in the system remains constant by fixing the so-called chemical potential μ.

Grand Canonical Ensemble

A macrostate of this so-called *grand canonical* ensemble suitable to describe the physical situation depicted above is determined by the volume V of the considered system and its temperature T. Since particles are allowed to enter or leave the system, a macrostate now consists of all possible microstates with arbitrary particle numbers and with arbitrary energies. Similar like above, the probability P_r of a particular microstate r depends on its energy E_r and its number of particles N_r. It can again be derived analytically and is given by

$$P_r(V,T,\mu) = \frac{1}{Z(V,T,\mu)} \cdot e^{-\frac{E_r - \mu \cdot N_r}{T}} \tag{2.23}$$

where $Z(V,T,\mu)$ is again a partition function that normalizes the probability distribution to 1, that is:

$$Z(V,T,\mu) = \sum_r e^{-\frac{E_r - \mu \cdot N_r}{T}} \tag{2.24}$$

Following arguments put forth in [Dorogovtsev *et al.*, 2003; Dorogovtsev and Mendes, 2003; Farkas *et al.*, 2004; Garlaschelli *et al.*, 2006; Waclaw, 2007], in the following we demonstrate how this equilibrium statistical mechanics perspective can be applied to random networks and thus - as we will argue in chapter 3 - to probabilistically structured overlay topologies. For this, we recall that the $G(n,p)$ model defines a probability space that contains all networks

Grand Canonical Ensemble of Random Graphs

with n nodes and any number of edges, the number of possible graphs depending on its size n. This reminds us of the grand canonical ensemble where a macrostate contains all microstates with a fixed volume V and an arbitrary number of particles, the number of possible microstates and with it their probabilities depending on the volume of the system. We further observe that average numbers of both particles (in the statistical ensemble) and edges (in the random graph model) are determined by the chemical potential μ respective the edge probability p. For a full analogy between statistical mechanics and random networks we merely lack definitions of a network's energy and temperature. This can easily be resolved by defining a trivial energy function which assigns a constant energy α to each edge, thus assigning energy $E_G = m \cdot \alpha$ to a graph G with m edges. Referring to equation 2.2 and assuming a network G with exactly m edges, this definition allows to reformulate the probability

$$P_G(n,p) = p^m \cdot (1-p)^{\binom{n}{2}-m}$$

that is given by the $G(n,p)$ model, in terms of the probability $P_G(n,\mu,T)$ of the grand canonical ensemble. As a first step, we can substitute the thermodynamic quantities in equation 2.23 by their respective network analogies.

$$P_G(n,\mu,T) := \frac{1}{Z(n,\mu,T)} \cdot e^{-\frac{E_G - \mu \cdot m_G}{T}} = \frac{1}{Z(n,\mu,T)} \cdot e^{\frac{m_G(\mu-\alpha)}{T}}$$

Based on equation 2.24, for the normalizing partition function $Z(n,\mu,T)$ we then obtain:

$$Z(n,\mu,T) := \sum_{G' \in G(n,p)} e^{-\frac{E_{G'} - \mu \cdot m_{G'}}{T}} = \sum_{G' \in G(n,p)} e^{\frac{m_{G'}(\mu-\alpha)}{T}}$$

In the latter equation, the summation goes over all possible networks with n nodes. Alternatively, we may consider the partition function as summing over all possible edge numbers. The maximum number of edges in a loop-free, labeled n-node network is $\binom{n}{2}$. When changing the summation from running over all networks to running over all possible edge numbers, for each number of edges k we further need to multiply the number of networks with n nodes that have exactly k edges which is given by $\binom{\binom{n}{2}}{k}$. This allows us to write for the partition function

$$Z(n,\mu,T) = \sum_{k=0}^{\binom{n}{2}} e^{\frac{k(\mu-\alpha)}{T}} \cdot \binom{\binom{n}{2}}{k} = \sum_{k=0}^{\binom{n}{2}} \left(e^{\frac{\mu-\alpha}{T}}\right)^k \cdot \binom{\binom{n}{2}}{k} = \left(1 + e^{\frac{\mu-\alpha}{T}}\right)^{\binom{n}{2}}$$

where the latter equality simply results from the binomial theorem. For a graph G with $m_G =: m$ edges we can then calculate $P_G(n,\mu,T)$ as

$$P_G(n,\mu,T) = \frac{e^{\frac{m(\mu-\alpha)}{T}}}{\left(1+e^{\frac{\mu-\alpha}{T}}\right)^{\binom{n}{2}}} = \frac{\left(e^{\frac{(\mu-\alpha)}{T}}\right)^m}{\left(1+e^{\frac{\mu-\alpha}{T}}\right)^{\binom{n}{2}}}.$$

Doing some further calculations

$$\begin{aligned} P_G(n,\mu,T) &= \left(e^{\frac{(\mu-\alpha)}{T}}\right)^m \cdot \left(1+e^{\frac{\mu-\alpha}{T}}\right)^{-\binom{n}{2}} \\ &= \left(e^{\frac{(\mu-\alpha)}{T}}\right)^m \cdot \left(1+e^{\frac{\mu-\alpha}{T}}\right)^{-\binom{n}{2}+m} \cdot \left(1+e^{\frac{\mu-\alpha}{T}}\right)^{-m} \\ &= \left(\frac{1}{1+e^{\frac{\alpha-\mu}{T}}}\right)^m \cdot \left(1-\frac{1}{1+e^{\frac{\alpha-\mu}{T}}}\right)^{\binom{n}{2}-m} \end{aligned}$$

yields the probability in the form $P_G(n,p)$ of the $G(n,p)$ model. From this one can immediately extract the relation

$$p = \frac{1}{1+e^{\frac{\alpha-\mu}{T}}} \tag{2.25}$$

between the per-edge probability p in the $G(n,p)$ model and the per-edge energy α and temperature T in the grand canonical ensemble of the corresponding equilibrium particle system.

Thermodynamics of the Classical $G(n,p)$ Model

With this reformulation of the classical $G(n,p)$ model in terms of the microstate probability of the grand canonical ensemble, interesting analogies between random graphs and equilibrium particle systems become apparent. For any fixed, constant edge energy α and chemical potential μ, temperature plays the role of the edge probability p in the $G(n,p)$ model. Here, two different cases can be identified: As the temperature T tends to infinity, the per-edge probability p tends to $\frac{1}{2}$, thus assigning equal probability to all possible networks, independent of their edge number and thus energy. This corresponds to the thermodynamic situation of infinite temperature, in which a system can equally well occupy states with arbitrary energy levels. A further interesting case is $T \to 0$. Here, depending on whether $\alpha > \mu$ or $\alpha \leq \mu$, the probability p converges to zero or one respectively. This corresponds to the zero-temperature limit in thermodynamics. When μ (which can be associated with the energy available per potential edge) is greater or equal than the energy α "required" per edge, only the fully connected graph is accessible while for $\mu < \alpha$ only the empty graph is accessible. Any finite temperature results in a random graph ensemble with a particular value for $p \notin \{0, \frac{1}{2}, 1\}$.

Edge Energies in Networks

Clearly, reformulating the $G(n, p)$ model in terms of the grand canonical ensemble is merely a demonstration of the analogies between random networks and thermodynamic systems. Thus, rather than providing deeper insight this analogy merely yields a different perspective on the structure of networks. The more interesting aspects of these analogies come into light when assigning edges (v, w) non-constant energies according for example to an energy function $e : V \times V \rightarrow \mathbb{R}$. A network's energy E_G can then again be given as the sum of edge energies which are in this case however not necessarily identical for all edges. From an argumentation analogous to the one yielding equation 2.25, one can then express the probability that an edge between two vertices v and w occurs as follows:

$$p_{(v,w)} = \frac{1}{1 + e^{\frac{e(v,w)-\mu}{T}}} \tag{2.26}$$

A Thermodynamic Perspective on Network Structures

In this equilibrium perspective, classical Erdös/Rényi random graphs in the $G(n, p)$ ensemble can be seen as merely one particular class of networks that is generated by assigning identical energies to all edges. By assigning edge energies that capture for example node characteristics, the synergetic potential existing between two nodes or the "stress" created by a connection, probability spaces can be created in which probabilities are assigned to graphs in much more sophisticated ways. Furthermore, critical points in the energy or temperature parameter space at which the properties of the resulting networks suddenly change can naturally be related to phase transition phenomena occurring in physical systems. In chapter 4 we will demonstrate how this equilibrium statistical mechanics perspective can be practically applied in the management and adaptation of probabilistic, scale-free overlay topologies. Here, based on the results put forth in this chapter, changes in the energy landscape and local, randomized edge rewirings will be used in order to adapt macroscopic network qualities in a distributed fashion. Prior to presenting this scheme in more detail, in the following chapter 3 we will discuss the potential of applying a "thermodynamic" perspective on random networks in the management of large scale overlay topologies in a more general fashion.

Chapter 3

Overlay Networks - A Thermodynamic Perspective

> A theory is the more impressive the greater the simplicity of its premises is, the more different kinds of things it relates, and the more extended is its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content concerning which I am convinced that, within the framework of applicability of its basic concepts, it will never be overthrown.
>
> ALBERT EINSTEIN, "AUTOBIOGRAPHICAL NOTES", 1949

A Thermodynamic Perspective on Large Scale Overlays

In the last chapter we have introduced some fundamentals of random graph theory and the study of complex networks. While this excursus may appear rather theoretical, in this chapter we discuss how a reasoning in terms of random graphs and complex networks can facilitate the practical design and operation of large dynamic networked computing systems making use of overlay topologies. For this, we briefly introduce and illustrate the commonly employed taxonomy of *highly structured*, *unstructured* and *probabilistically or loosely structured* overlays and summarize their advantages and disadvantages. We then discuss the potential benefits of taking a *thermodynamic perspective on the management of overlay topologies* which is inspired by the analogies between complex networks and systems considered in statistical mechanics that have been discussed in the previous chapter. In short, the key motivation for this approach is the finding that - in the face of highly dynamic participants - maintaining overlay networks with simple deterministic structures can be complicated and costly while *the formation of topologies with complex,*

probabilistic structures is often simpler, more efficient and, in fact, natural. While some related arguments have been put forth earlier in [Scholtes *et al.*, 2008a; Scholtes, 2010], here we present them in a broader context and in the light of the findings that have been summarized in the previous chapter.

3.1 Structured and Unstructured Overlays

Overlay Topologies - Highly Structured Approaches

In section 1.3 we have introduced the basic motivation for the use of overlay networks and have commented on their importance for the deployment of novel services and Peer-to-Peer systems on top of existing network infrastructures like the Internet. During the last couple of years, the design and analysis of scalable overlay topologies and distributed algorithms operating upon them has received significant attention. Different approaches have been considered to efficiently provide a set of base services like distributed search, the key-based retrieval of decentrally stored data, efficient application-layer routing or scalable content distribution. In many of the commonly deployed distributed services, *highly structured overlay networks* are being used in which virtual connections between the participating machines or applications are created in a globally consistent and controlled fashion to form a predetermined and deterministic topology. This allows to optimize both the structure of the overlay as well as the distributed algorithms operating upon them with respect to the particular algorithmic task being addressed.

The Lookup Problem

To exemplify this approach along with its advantages and disadvantages, in the following we give a brief (and thus necessarily rough and incomplete) overview of the highly structured overlay CHORD [Stoica *et al.*, 2001], which addresses the scalable lookup of data items that are stored decentrally on the participating machines. Solving this so-called *lookup problem* efficiently, is a problem of prime importance for a variety of distributed applications [Balakrishnan *et al.*, 2003]. While CHORD is one of the most prominent highly structured overlays that address this issue, a number of other approaches have been proposed [Rowstron and Druschel, 2001; Ratnasamy *et al.*, 2001; Aberer, 2001; Hildrum *et al.*, 2002; Maymounkov and Mazieres, 2002]. The main issue in looking up data in large distributed systems is to efficiently locate a machine that is responsible for a certain data item under the restriction that each of the participants has only a limited view of the network. This view is usually restricted to a small, ideally constant, number of nodes to which it maintains virtual connections in the overlay.

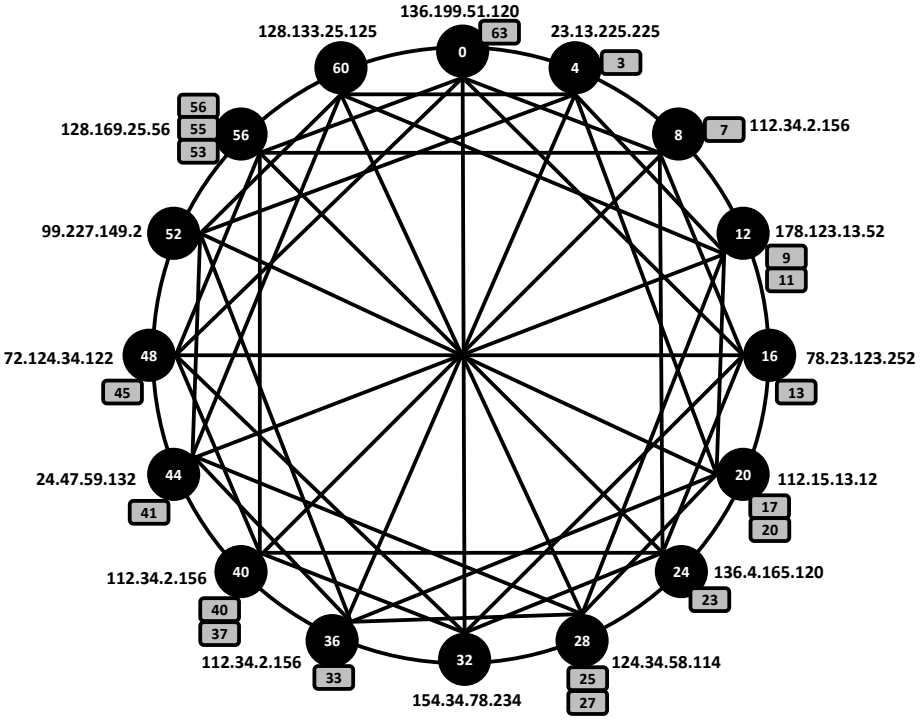


Figure 3.1: Example for a highly structured overlay topology created by the CHORD Protocol

Looking up Data in Highly Structured Systems: CHORD *in a Nutshell*

Figure 3.1 shows an overlay network built by the CHORD protocol to efficiently solve this problem. Black nodes represent machines participating in the distributed storage and retrieval of data items. By means of a consistent hashing scheme [Karger *et al.*, 1997], unique identifiers in the range $[0, 2^m - 1]$ for some suitably large m are assigned to each of these nodes based on the machines' IP addresses. A participant with ID i creates and maintains virtual connections to all machines whose IDs are closest to the positions $i + 2^j$ modulo 2^m in the identifier space (for $j = 0, 1, \ldots, m - 1$). Data items that shall be stored are then mapped onto the participating machines by means of the same hashing scheme. In particular, each participating node is responsible for the storage of all keys that fall in the range between its own ID and the next smallest node ID in the system. In Figure 3.1 the keys of data items are depicted as boxes and are arranged next to the nodes that are responsible for their storage. When a data item with a certain key needs to be retrieved or stored, a node uses a simple greedy mechanism, sending the request to the neighbor whose identifier is closest but smaller than the key in question. By means of this sim-

ple scheme and the structure of the overlay, a lookup in CHORD takes $O(logn)$ steps where n is the number of participating machines.

Drawbacks of Structured Overlays

A key feature of protocols like CHORD is that they give rise to ordered, highly structured overlay topologies that are in part inspired by linked data structures, like for instance search trees, that are commonly used to address related problems in non-distributed computing. While distributed algorithms operating on such highly regular structures can achieve good and predictable performance, their application in a distributed context is complicated significantly by the dynamics present in real-world computing systems. Considering for example Peer-to-Peer systems, nodes in the overlay topology can represent highly heterogeneous and unreliable machines that are contributed by participating users. Joining and exiting participants, crashing or misbehaving machines and communication errors result in a continuous fluctuation of nodes and links in the topology, an effect that is commonly called churn. To countervail the loss of structure and the uncertainties that are due to these fluctuations, maintenance mechanisms need to be employed which continuously repair the deterministic structure of the topology. In CHORD, a stabilization protocol continuously checks neighbors, removes broken links and reorganizes the topology accordingly, thus maintaining the distributed indexing structure that facilitates efficient lookups. As has been argued for example in [Balakrishnan *et al.*, 2003], the overhead entailed by such maintenance routines can dominate the overall performance in real-world deployments, that is *the cost of maintaining a highly structured topology potentially exceeds the cost entailed by actual data queries*. Apart from these operational costs, a further drawback is that correctly implementing maintenance protocols is non-trivial due to the possibility of multiple concurrent node joins and failures. Besides the costs and complexities of structure maintenance, there are further aspects that question the use of highly structured topologies in a number of real-world scenarios. Their efficiency often comes at the price of imposing a *tight coupling between the network structure and distributed algorithms*. Its topology being tailored to facilitate fast key-based requests, CHORD cannot efficiently support complex query types like range, partial match or fuzzy queries. Furthermore, the rigidness of highly structured systems makes it hard to adapt their network topology to the heterogeneous and possibly dynamic resources of machines. In particular, due to the homogeneous distribution of keys and connectivity, CHORD implicitly assumes that the characteristics of participating machines are more or less equal and time-invariant, which is hardly the case in reality.

Overlay Topologies - Unstructured Approaches

A simple idea that allows to circumvent many of the above problems is to dismiss the tight control over the overlay topology that is being applied in highly structured systems. Simple protocols can be used instead which neither construct nor maintain a particular topology, but rather form links in an uncontrolled, ad hoc fashion. The interest in such systems has largely been triggered by a first generation of *unstructured overlays* used for example in the GNUTELLA file-sharing application. Being used mainly for the illegal exchange of copyrighted content, GNUTELLA can be seen as one of the first truly decentralized Peer-to-Peer systems that has been deployed at a global scale in the Internet. In order to simplify its implementation and circumvent the complexities entailed by dynamic and heterogeneous participants, a maximally simple protocol has been used. Machines of users that enter the system create links to existing participants in a rather random and uncontrolled fashion, thus creating an overlay topology whose structure is largely influenced by the behavior of users, the spreading of information about existing nodes with popular content, the reliability of machines or the distribution and dynamics of resources. Apart from being connected - that is allowing each pair of participating machines to communicate with each other - no further assumptions about the detailed structure of the network or the placement of content were made. An example for such an unstructured overlay, as it could possible have emerged for example from the protocol underlying early versions of GNUTELLA, can be seen in Figure 3.2.

Advantages and Disadvantages of Unstructured Overlays

The use of such an unstructured topology clearly circumvents the complex and costly maintenance algorithms of highly structured systems. In an unstructured system, the fluctuation that is due to joining and exiting participants cannot destroy any sophisticated structure and thus very simple maintenance mechanisms that sustain or restore connectedness are usually sufficient. Furthermore, the flexible structure of the overlay facilitates adaptation mechanism by which for example the number of virtual connections maintained by a particular node can be matched to its resources or reliability. However, apart from simplifying matters, dismissing control about the overlay's structure also results in the fact that there are no cues that could be utilized by distributed search schemes. Knowing nothing whatsoever about the structure of the topology and the placement of data items leaves exhaustive search as the only option. In early versions of GNUTELLA a simple flooding strategy was thus used by which search requests were propagated through the network up

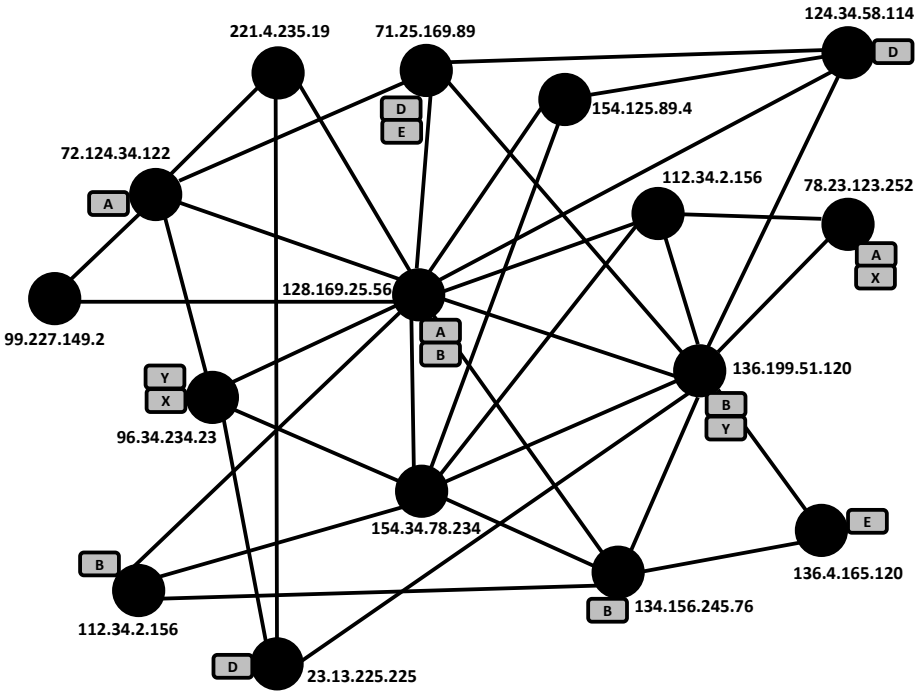


Figure 3.2: Example for an unstructured overlay topology as created for example by the GNUTELLA protocol

to a maximum depth of seven hops[1]. The small diameter of the overlay resulting from GNUTELLA's connection scheme (as predicted for example even by the maximally simple classical random graph models) and its deployment at a global scale quickly resulted in significant scalability issues. The mere signaling traffic resulting from the propagation of search requests at times reportedly amounted to a significant fraction of global Internet traffic [Ripeanu *et al.*, 2002].

[1]See `http://rfc-gnutella.sourceforge.net/developer/stable/index.html#t3-2-7`, accessed on October 18th, 2010.

3.2 Thermodynamically Structured Overlays

The Middle-Ground: Loosely Structured Overlays

Summarizing the above arguments, one can conclude that highly and rigidly structured overlays can achieve optimal and predictable performance at the price of potentially complex and costly construction and maintenance protocols and a lack of adaptiveness. On the contrary, unstructured overlays sacrifice this optimum performance in favor of a simplified management of the overlay, a decoupling of the topology's structure and algorithmic schemes and the facilitation of adaptive mechanisms. Since each of the two approaches provides advantageous aspects, in recent years a middle ground of so-called *loosely or probabilistically structured* overlay topologies is increasingly being explored. It is largely inspired by the observation that even the simple, explicitly or implicitly stochastic topology construction mechanisms used by unstructured systems can give rise to topologies which - although not being deterministic - exhibit non-random, complex characteristics that facilitate the development of efficient distributed algorithms. An example for this can be found in the GNUTELLA system, where the observation of complex characteristics in the emerging overlay has resulted in vastly improved distributed strategies which have since replaced flooding-based search [Adamic *et al.*, 2001; Ripeanu *et al.*, 2002; Chawathe *et al.*, 2003]. In the last couple of years, the approach of employing simple, probabilistic protocols by which overlay topologies with advantageous, complex characteristics emerge is increasingly being acknowledged in research and a number systems have been proposed that make use of probabilistic, loosely structured overlays [Ganesh *et al.*, 2003; Sarshar and Roychowdhury, 2004; Voulgaris *et al.*, 2005; Sarshar and Roychowdhury, 2005; Lua *et al.*, 2005; Shahabi and Banaei-Kashani, 2005; Tian *et al.*, 2005; Sandberg, 2006; Vishnumurthy and Francis, 2006; Terpstra *et al.*, 2007; Bustamante and Qiao, 2008; Halim *et al.*, 2008].

Harnessing Complex Structures in Overlay Topologies

One of the most interesting aspects of both unstructured and loosely structured systems is that *the emerging overlay topologies are often neither completely random nor deterministic*. As such, they constitute a promising *middle ground between highly structured and unstructured overlays* that seems to be best suited for the design of large, dynamic and adaptive distributed systems. The structural properties of such network topologies, their influence on dynamical processes and the question how they emerge from simple, local and stochastic construction procedures can furthermore be analyzed in the conceptual framework of random graphs and complex networks. It has thus been acknowledged that the tools and findings that have been accumulated in complex network science

Towards Thermodynamically Structured Overlays

during the last decade play a crucial role in the engineering of simple protocols that create robust overlay networks with predictable characteristics [van Steen, 2004; van Steen, 2010]. In the remainder of this chapter we further argue that the analogies between complex networks and equilibrium as well as non-equilibrium particle systems that have been put forth in the previous chapter facilitate *a thermodynamic perspective on the self-organized formation of structures in overlay networks, the macroscopic reasoning about their properties, the relation between the stochastic dynamics of individual nodes and aggregate network qualities, as well as on the adaptation of their qualities to dynamically changing conditions*. The resulting idea of using *thermodynamic analogies* in the design of networked systems resembles - though rather in spirit than in terms of the underlying theoretic foundation - the *thermodynamic system design* that has earlier been advocated in [Kubiatowicz, 2003] for the engineering of reliable distributed systems. Here, the relation between the resulting, inherently probabilistic statements and systems designed in a more traditional, deterministic fashion has been summarized as follows:

> *"Those uncomfortable with probabilistic arguments should consider that traditional systems fail under many circumstances. Thermodynamic, self-organized systems can provide strong guarantees." [Kubiatowicz, 2003]*

Thermodynamically Structured Overlays

An important objective of this dissertation is to further an *interdisciplinary perspective* on large dynamic networked systems that is based on recent results of complex networks science and the statistical mechanics' view on networks. A particular focus will be laid upon the question how findings from the complex networks community and analogies to thermodynamic systems can be applied in a constructive fashion in the engineering of what one may call *thermodynamically structured overlays*. In the remainder of this chapter, a number of promising aspects and arguments that justify this notion are presented.

3.2.1 The Structuredness of Overlays: An Entropy Perspective

Entropy of Network Ensembles

When studying the formation and maintenance of structures in overlay networks, applying the concept of *entropy* as introduced in classical thermodynamics [Clausius, 1864], statistical mechanics [Boltzmann, 1896] and information theory [Shannon, 1948] can actually be insightful. In the context of networks, a number of different notions of *graph or network entropy* have been considered (see for example [Simonyi, 1995; Ji *et al.*, 2008; Bianconi, 2009]). For

a statistical ensemble Ω of networks and a probability measure P, it seems natural to consider the following definition of *structural entropy* $H(\Omega, P)$

$$H(\Omega, P) := \sum_{G \in \Omega} P(G) \cdot log\left(\frac{1}{P(G)}\right) \tag{3.1}$$

Entropy of Highly Structured and Unstructured Overlays

as it directly relates to the entropy of a statistical ensemble in statistical mechanics. The structural entropy can be viewed as capturing the degree of information which one has about the detailed structure of a network drawn from a statistical ensemble or - in other terms - the degree of uncertainty introduced by the stochastic processes constructing a topology. The structural entropy is maximized for a process or ensemble which constructs every possible network with equal probability $P \equiv \frac{1}{|\Omega|}$, thus yielding $H(\Omega, P) = log(|\Omega|)$.[2] In this case, no statements about the detailed structure of a network stemming from the corresponding construction process can be made. This situation of *maximum ignorance* can intuitively mapped to *completely unstructured overlays* which make no assumptions about the network topology emerging from a totally uncontrolled construction process. The opposite is true for systems that use *highly structured overlays* and thus tightly control the topology construction process. The probability measure of the statistical ensemble corresponding to this situation tends to a delta function, meaning that only one particular network topology can occur. The entropy of such a statistical ensemble tends to a minimum of $H(\Omega, P) = 0$ which means that we have complete information about the detailed structure of network resulting from such a process.

The Middle-Ground: Overlay Networks with Complex Structures

The structural entropy of networks with complex, probabilistic structures ranges between the two extremes of minimum and maximum structural entropy. In the framework of statistical ensembles, this means that - while not being completely deterministic - networks with certain features are much more likely to emerge than other realizations. As such, the structural entropy gives an intuitive *measure for the structuredness or orderliness of an overlay* resulting from a particular topology construction or maintenance scheme or. It may further be seen as a *measure for the amount of ignorance or information* one has about the emerging structures. In this perspective one may further view machine crashes, communication errors or churn as processes that decrease the orderliness of an overlay and thus increase the structural entropy of the underlying network ensemble. The fact that such effects inevitably occur in real-world computing systems resembles - though informally - the increase of entropy in

[2] One easily sees that this corresponds to the entropy of the micro canonical ensemble of equilibrium statistical mechanics and the $G(n, m)$ model of classical random graphs.

physical systems that is due to the second law of thermodynamics. Similarly, protocols and algorithms that maintain or restore ordered structures in overlay networks can be viewed as decreasing the structural entropy of a network ensemble at the expense of requiring communication and computation. In this view, one may argue that low entropy systems facilitate the use of efficient distributed algorithms while high entropy systems are usually easier to maintain. Network topologies with complex structures apparently offer a compromise between these conflicting goals.

3.2.2 Macro-Level Arguments: Thermodynamic Guarantees

Thermodynamic Guarantees in Probabilistically Structured Overlays

When using distributed protocols that create loosely or probabilistically structured overlays, one is typically interested in creating topologies with certain qualities like for instance being connected, having small diameter and average shortest path lengths, facilitating gossiping and information spreading schemes or being robust against failing nodes. We have seen in chapter 2 that a reasoning in terms of statistical ensembles allows to assess the probability that networks being created by a certain statistical model exhibit such properties. We have also seen that strong probabilistic guarantees can be derived for properties that hold asymptotically almost surely, given that the system in question is sufficiently large and the convergence sufficiently fast. Furthermore, the parameters which define the statistical ensemble from which a network is drawn and upon which these probabilistic arguments depend are often rather simple aggregate quantities like for instance a number of edges being added to a network at random or the distribution of node degrees in the network. In a sense, this *macroscopic reasoning* about the collective properties of random networks resembles the study of relations between measurable and observable bulk material properties like for instance volume, pressure, temperature, heat capacity or compressibility in *classical thermodynamics*. A more formal justification for this analogy can actually be found when considering the relations between the parameters of the classical $G(n, p)$ random graph model and the macroscopic quantities defining the grand canonical ensemble of equilibrium systems that have been discussed in section 2.4. As such it appears justified to call probabilistic guarantees about network properties that can be related to mere aggregate quantities of a statistical ensemble *thermodynamic guarantees*.

Thermodynamic Guarantees

Macroscopic Arguments in Large Scale Computing Systems

In the engineering of overlay topologies for large and dynamic computing systems, an active and meaningful use of such thermodynamic guarantees can provide a number of benefits. First of all it allows to employ *a macro-level reasoning about the properties of very large network structures* that is agnostic of the

exact, microscopic details of the connectivity structure and rather considers aggregate or average quantities of statistical ensembles. Given that the system is sufficiently large, *de facto guarantees about properties that are relevant to the robustness, functioning and performance of a distributed system can thus be made* even though allowing a certain degree of uncertainty and thus flexibility at the level of individual nodes or connections. In order to be able to employ these arguments, one needs to ensure that the parameters of the statistical ensemble from which a topology is drawn fall within a certain range. However this is often much simpler than correctly implementing highly sophisticated construction procedures. Furthermore, based on the statistical nature of the underlying arguments and again resembling arguments about thermodynamic systems, thermodynamic guarantees *tend to become stronger as the size of the system increases*. This is opposed to many deterministically structured systems in which a sound reasoning about their properties often becomes increasingly complicated as they grow larger.

3.2.3 From Micro- to Macro-Level: A Statistical Mechanics Perspective

Statistical Mechanics of Overlay Networks

The arguments put forth above suggest that a macro-level perspective on probabilistically structured overlay topologies possibly allows to derive strong, thermodynamic guarantees about their collective properties. While this can clearly simplify the design and analysis of large networked computing systems, an important question is how a suitable macroscopic description can be derived from the actual protocols, algorithms or behavioral patterns which govern the dynamics of individual machines, users or applications. Seeking again analogies in the domain of physics, this task is closely related to the field of *statistical mechanics* which is involved with relating macroscopic, bulk material properties to the statistics of microscopic - mainly mechanical - interactions between individual atoms and molecules. As has been presented in section 2.4, the crucial idea which allows to cope with systems consisting of an intractable number of interacting elements is to suitably subsume all possible *microstates* of a system in a *macrostate* with reasonably assigned statistical weights.

Large Scale Networked Systems: Relating the Micro- and Macro-Level

In the context of overlay networks, this translates to the question how a graph ensemble with suitably assigned probabilities can be derived from the dynamics of the underlying topology construction process. During the last couple of years, in the study of complex networks a number of methods originating from the field of statistical mechanics have been used for this purpose.

In some simple cases, probabilities of individual network realizations can directly be inferred from equilibrium arguments about the (stochastic) dynamical process by which the topology evolves. A simpler approach which is often applied when dealing with more complex situations is to directly calculate aggregate or average quantities of interest like average degree, degree distribution or clustering coefficient for the resulting statistical ensemble based on a stochastic model of the dynamics of individual elements. For this purpose, statistical mechanics tools like mean field theory or master equations are frequently being used (see for example [Barabási and Albert, 1999; Barabási *et al.*, 1999; Newman *et al.*, 2000; Dorogovtsev *et al.*, 2000b; Albert and Barabási, 2002; Dorogovtsev and Mendes, 2003; Farkas *et al.*, 2004; Palla *et al.*, 2004]). An application of such methods to overlay networks allows to relate a stochastic model of the local behavior that governs the topology construction process and the dynamics of participating machines to a macroscopic description in terms of a statistical ensemble which can then be used to reasonably argue about the collective properties of the resulting network topologies.

Limitations of Macroscopic Arguments

Whether or not it is possible to apply such arguments in a particular setting clearly depends on how well a description in terms of the resulting statistical ensemble captures the processes that construct a network topology. In particular, all arguments from section 2.3.4 about the potential influence of correlations apply and the imprudent use of overly simple stochastic models in the description of networks resulting from highly sophisticated processes can lead to erroneous results and wrong statements. This is a common critique from the engineering community and it will be addressed in somewhat more detail in the context of scale-free networks in chapter 4. In fact, this critique mainly applies when trying to model highly sophisticated and optimized engineered systems by means of maximally simple models giving rise to complex structures or behavior as they are frequently considered in the natural sciences[3].

Designing Predictable Systems by Randomization

However, rather than using the findings that have been summarized in chapter 2 in the modeling of systems which are given *a priori*, one may also consider the alternative approach of *explicitly designing networked computing systems along simple stochastic models* that give rise to networks with interesting, complex characteristics. This approach is particularly facilitated in the domain of Peer-to-Peer systems and overlay networks, where the formation and deletion of (virtual) connections can be controlled by applications and are freed from

[3]Differences in the reasoning about highly organized, engineered systems on the one hand and large systems consisting of randomly behaving elements on the other hand may be viewed in terms of what has been called in [Weaver, 1948] *disorganized vs. organized complexity*. Some further arguments can be found in the recent article [Alderson and Doyle, 2010] as well as in chapter 4.

many of the tight constraints that exist at lower network layers. Here, explicitly and targetedly randomized protocols, message exchanges or connection strategies can be used to enforce particular, well studied statistical ensembles of complex networks for which thermodynamic guarantees on relevant properties are known to hold and for which correlations are absent or minimal. In such systems, a sensible introduction of randomness can actually result in systems that allow to *make reliable statements about global, emergent properties* as well as about their relations to the dynamics of individual elements. In chapter 4 we will consider some practical aspects of this approach, thus proposing a distributed protocol which creates random network topologies that are drawn from an equilibrium statistical ensemble of scale-free networks with a tunable degree distribution exponent.

3.2.4 Structure Formation in Equilibrium and Non-Equilibrium Overlays

Equilibrium and Non-Equilibrium Regimes in Networked Systems

The arguments presented above suggest that methods and findings from random graph theory and complex networks science can have rather practical implications on the design and analysis of protocols that give rise to overlays with probabilistic structures. Apart from this, an interesting aspect of the analogies between complex networks and physical systems is that they allow to view the formation of network structures both from an equilibrium and a non-equilibrium perspective. When dealing with networked computing systems, one easily finds situations in which - depending on the dynamics that governs the system - either the one or the other perspective is required.

Structure Formation in Non-Equilibrium Overlays

In most real-world computing systems phases can be identified in which the number of connected devices or users grows or decays. One obvious reason for systems that grow over time is the addition of resources due to a growing demand of computational power or bandwidth or a growing user base. In the context of Peer-to-Peer systems, where the dynamics of resources is closely coupled to user behavior, periodic phases of growth and decay may further occur based on the circadian rhythm of the participants. In such phases, networked systems can naturally be related to non-equilibrium physical systems which are characterized by an in- or outflow of energy or particles. Taking a *non-equilibrium statistical mechanics perspective* on the formation of structures or - based on the notions discussed above - the decrease of structural entropy in overlay networks points at interesting relations to self-organization phenomena occurring in physical and biological systems. As such, *the non-equilibrium*

statistical mechanics perspective on the emergence of complex network structures constitutes an important foundation for the meaningful application of biologically inspired self-organization and self-adaptation schemes in overlay networks.

Crystallization of Overlays in Equilibrium Systems

Although taking this non-equilibrium perspective is appealing, one can clearly think of a number of situations in which networked computing systems rather reside in a state of equilibrium. One can for example imagine a Peer-to-Peer system whose participants randomly join and leave at constant and balanced rates and in which the protocol being applied by the nodes maintains a certain average number of randomly created connections. Similarly, even in systems without any dynamics in terms of joining or leaving participants, link fluctuations can be introduced deliberately for example by a sporadic random rewiring of overlay connections. Such situations naturally correspond to particle systems residing in thermodynamic equilibrium and the random link or node fluctuations which are due to the dynamics of the system may be viewed in analogy to temperature-dependent fluctuations in equilibrium particle systems. *In such equilibrium systems the formation of network structures can be viewed as a stochastic optimization process* by which - based on a reasonably assigned energy landscape and temperature-dependent random fluctuations - optimal or near-optimal topologies can emerge in a self-organized fashion. This perspective closely resembles well known simulated annealing strategies [Kirkpatrick *et al.*, 1983] which are used routinely for heuristic optimization in large combinatorial problems.

3.2.5 Adaptation in Thermodynamically Structured Overlays

Statistical Mechanics - A Constructive Approach

Above we have pointed out some promising aspects of employing a statistical mechanics and thermodynamic perspective in the management of large scale overlay networks. These perspectives are in part motivated by analogies between particle systems and overlay topologies. In some respects, these analogies are necessarily incomplete. While the behavior of atoms and molecules is ruled by the laws of physics, computing devices can actively influence their behavior for example in response to a change in internal or environmental conditions. In equilibrium systems, energy landscapes can often be adjusted at will in order to give rise to overlay topologies with certain qualities. Likewise, connection protocols and linking preferences can often be tuned to change the structures formed in growing non-equilibrium systems. In the underlying theoretical framework of statistical ensembles, such systems can be viewed as intentionally switching between topologies drawn from different statistical ensembles defined by different aggregate parameters.

Harnessing Critical Points: Adaptation by Phase Transitions

Reconsidering for example the influence of such parameters on *thermodynamic guarantees*, an interesting aspect is the existence of critical points in the parameter space of statistical ensembles above or below which certain of a network's structural properties quickly fade while others quickly emerge. In chapter 2, we have encountered a number of examples for this kind of threshold behavior, most of which can be related to the critical point that has been derived in [Molloy and Reed, 1995]. When wanting to adapt the characteristics of an overlay, such critical points are particularly interesting since here even a slight change of parameters can entail a drastic change of a network's collective properties. Referring to earlier arguments that suggest a close relation of such threshold phenomena to phase transitions in thermodynamic systems, one can argue that *phase transitions can be harnessed to quickly adapt the qualities of thermodynamically structured overlays.* Such an approach can, for example, be used if one wishes to reach a state in which thermodynamic guarantees on the performance of certain dynamical processes can be derived, thus possibly forfeiting guarantees on other properties that may momentarily seem less important. In order to achieve such an adaptation at the macro-level of an overlay network, individual participants are in general required to adapt their local, micro-level behavior, protocols or linking preferences in a particular way. The question which micro-level changes are necessary to achieve a certain change of aggregate or statistical parameters at the macro-level can again be answered by means of the principles that have been put forth in section 3.2.3. In chapter 4, we will demonstrate this by presenting a practical scheme that can be used to trigger phase transitions in probabilistically structured, scale-free overlays based on the change of a random walk bias that is being used by nodes to sample endpoints of overlay connections.

Adaptation at the Micro- and Macro-Level

Complex Adaptive Overlay Topologies

All these possibilities are opened up mainly by the fact that - in contrast to many other networked systems - *connections in overlay networks are virtual rather than physical constructs and thus easily alterable*. This flexibility facilitates a constructive approach to the findings that have been obtained in the complex networks and statistical physics community. Their targeted use constitutes an important cornerstone in the engineering of *overlay topologies with complex adaptive structures*. Since the mechanisms by which the formation and adaptation of these structures is achieved can directly be related to corresponding self-organization principles in equilibrium and non-equilibrium systems in nature, they can further be seen as a step towards *organic computing systems* as envisioned for example in [Müller-Schloer *et al.*, 2004; Schmeck, 2005].

Chapter 4

Managing Scale-Free Overlay Networks

> In nature, nothing is perfect and everything is perfect. Trees can be contorted, bent in weird ways, and they're still beautiful.
>
> ALICE WALKER

Creating and Adapting Random Scale-Free Overlays

In section 3.2.3 we have mentioned that one possible approach towards a *thermodynamic management of large scale overlays* involves the application of explicitly randomized schemes which actively enforce well studied statistical ensembles of complex networks with advantageous properties. Existing findings about the networks being drawn from such statistical ensembles can then be applied, possibly resulting in strong macroscopic, *thermodynamic guarantees* for the qualities of overlay networks emerging from distributed stochastic construction procedures. Based on the findings presented in section 2.3, one may further argue that the characteristics of random scale-free networks appear interesting for a targeted use in large scale overlay networks. In this chapter we turn to some practical questions that occur when wanting to implement such systems. First and foremost we introduce a distributed, stochastic protocol that is able to transform arbitrary overlay networks into random scale-free topologies with an adjustable degree distribution exponent. This distributed scheme has originally been published in [Scholtes, 2010] and will be presented in more detail in section 4.2. Here we further argue that the proposed protocol is a practicable approach towards the idea of *actively using phase transition phenomena for a targeted adaptation of macroscopic network qualities*. In particular, the targeted change of the degree distribution exponent of scale-free network topologies

facilitates the adaptation of properties which influence the performance of dynamical processes like distributed search, information dissemination as well as the resilience against faults and attacks.

Monitoring Macroscopic Properties of Scale-Free Overlays

Apart from considering systems in which the participating machines have complete control over the topology construction process, one may further study the question how a macroscopic reasoning as laid out in section 3.2 can be applied to systems where the processes by which the overlay emerges cannot easily be assessed or influenced. Examples include systems in which the structure of the overlay is due to the dynamics of users, collaborative or social aspects as well as technological constraints. Although the network topologies emerging in such systems are clearly not random, the measurement of aggregate statistical parameters can still - at least to a certain extent - facilitate a macroscopic reasoning about the properties of the network in question. Again considering the management of overlay topologies with scale-free structures, in section 4.3 of this chapter we propose a simple monitoring protocol by which each node can obtain knowledge about the network's degree distribution in a fully distributed, decentralized and probabilistic fashion. As such, the protocol - which has originally been published in [Scholtes *et al.*, 2008a] - allows to reliably measure aggregate statistical information about the network without requiring a central instance to collect a complete histogram of node degrees.

4.1 Scale-Free Structures in Computer Networks - A Critical Appraisal

Scale-Free Networks and Computing Systems

Prior to discussing these rather practical aspects of managing overlay topologies with complex characteristics, it deems appropriate to make some remarks about both the enthusiasm and the criticism that has accompanied the application of findings about scale-free structures in the domain of networked computing systems. Following the observation of power law degree distributions in a number of social, biological but also technological systems in the late 1990s, a number of theoretic findings have been obtained about the properties of scale-free networks (some of them being reviewed in section 2.3.2). While it seems alluring to apply these results to networked computing systems with power law degree distributions, such a reasoning bears a number of pitfalls. Some of these pitfalls are actually due to methodical errors that can occur when mapping the topologies of very large, dynamic networked computing systems. As has been argued for example in [Lakhina *et al.*, 2003;

Problems Associated with Data Acquisition and Analysis

Stutzbach *et al.*, 2006] imprudently sampling a subset of nodes in large networks can introduce significant biases with respect to the observed degree distributions. When using tools like for instance *traceroute*[1] to map Internet topologies, it has been argued that the choice of source points from which *traceroute* measurements are being started are crucial for the sampling quality [Achlioptas *et al.*, 2005]. In [Stutzbach *et al.*, 2008; Stutzbach *et al.*, 2009] it has further been laid out that the use of slow crawlers which progressively map Peer-to-Peer overlays while they continue to evolve can lead to wrong conclusions about their connectivity distribution. In particular, in [Achlioptas *et al.*, 2005; Stutzbach *et al.*, 2008] it has been suggested that biases stemming from mere sampling processes question the validity of earlier observations of power law degree distributions in the GNUTELLA overlay network and the Internet's router topology. Apart from problems that are due to the complex data acquisition process in very large networked systems like the Internet or Peer-to-Peer systems, a further potential error source is associated with fitting procedures commonly used both to identify power laws in empirical data and to infer the exponent of the distribution. As laid out in [Willinger and Doyle, 2002; Clauset *et al.*, 2007] and as will be summarized in more detail in the corresponding section 4.3 of this chapter, the mere statistical properties of highly skewed distributions complicate a reliable fitting and thus require caution and experience in the data analysis process.

Criticism of Simple Models for Complex Engineered Systems

Even if the alleged power laws for the connectivity of engineered computing systems like those mentioned above hold, one needs to be cautious when *applying arguments about random networks to highly optimized, technological systems like the Internet*. The evolution of such systems is often subject to detailed and sophisticated planning and optimization and thus far from being random. As a result, properties that hold asymptotically almost surely for a network drawn from a statistical ensemble of random scale-free networks can very well not hold for the router topology of the Internet or a particular Peer-to-Peer overlay, even though their degree distribution follows a power law. This is because sophisticated construction processes may selectively produce network topologies that are unlikely to occur at random. In the past, a hasty transfer of findings about random scale-free networks to technological systems with alleged power law connectivity has raised the concerns of network engineers. For instance [Willinger and Doyle, 2002; Alderson *et al.*, 2005; Doyle *et al.*, 2005] argue that earlier claims about the

[1] *traceroute* is a simple program which makes use of the Time-to-live header in IP packets to discover all routers on a route between the source and an arbitrary destination machine.

Internet's alleged susceptibility against targeted attacks [Albert *et al.*, 2000; Cohen *et al.*, 2001] - even though correct for random scale-free networks - are wrong for the actual topology of Internet routers. This can easily be understood when considering that in a random scale-free network, the most connected hub nodes lie - with high probability - in the network's center. Consequently, they play a crucial role for the network's connectedness and selectively removing them can rapidly disintegrate a system (as shown in Figure 2.5). Contrariwise, the purpose of the most connected routers in the Internet is to aggregate large numbers of low bandwidth links at the network's edge and connect them to a high speed core of routers with few, high bandwidth connections.

Reconciling Engineered and Complex Networks

Recently, the potential problems which occur when imprudently transferring results for random scale-free networks to the Internet topology have evolved into a more general critique of using complex networks methods in the modeling of engineered networked computing systems [Alderson, 2008]. As argued for instance in a recent critical review article [Alderson and Doyle, 2010], much of this criticism boils down to the difference between, what has been termed by [Weaver, 1948], *organized and disorganized complexity*. Here, disorganized complexity is viewed as emerging from simple, random interactions between a sufficiently large number of elements whereas organized complexity is the result of engineering principles, architectural decisions, multi-layered organization, compromises between conflicting goals and sophisticated, non-random interactions. Reconsidering the modeling of the structure and the evolution of the Internet, in [Carlson and Doyle, 1999; Fabrikant *et al.*, 2002; Li *et al.*, 2005] first steps have been taken towards blending the views of statistical physics and complex networks (as summarized for instance in [Pastor-Satorras and Vespignani, 2004]) with those of engineering. Nevertheless, it is an important yet unresolved question exactly which aspects of sophisticated, engineered systems that exhibit *organized complexity* can be tackled in the framework of simple stochastic models. For this, domain-specific information like technological and economical constraints as well as geographic, political and social aspects probably need to be incorporated.

Complex Networks and Peer-to-Peer Systems

Having warned the reader about these pitfalls, it is important to note that nevertheless a number of aspects (like growth, preferential attachment, copying mechanisms, recursive construction schemes or highly heterogeneous node fitness) of simple stochastic models that produce scale-free networks[2] can be found in a number of technological networked systems. There clearly is growth in the World Wide Web and linking strategies are likely to resemble preferen-

[2] See for example the corresponding models that have been summarized in section 2.3.5.

tial attachment or random walk schemes. Ranging from simple mobile phones, notebooks, workstations to high performance servers, the capabilities of devices that are connected to Internet-based systems are clearly highly heterogeneous, thus resembling fitness based models. Furthermore, protocols underlying Peer-to-Peer systems like GNUTELLA can be related rather directly to copying and random walk schemes known to give rise to power law degree distributions.

Scale-Free Overlay Networks: A Constructive Approach

Picking up on arguments put forth in chapter 3, here we further argue that, in addition to the modeling of technological systems, the large body of literature on complex networks facilitates a *constructive approach* towards overlay networks with interesting, complex features. This approach is rendered possible by the fact that overlays are mere virtual, application-level constructs, being built on top of lower network layers which allow connections to be established between arbitrary machines irrespective of the underlying physical topology. In such a setting, where links between nodes can be created and deleted at will based on application needs, explicitly and targetedly randomized protocols can be used to *create overlays for which a reasoning in terms of random scale-free networks is valid* and which facilitate the application of the thermodynamic perspective that has been presented in chapter 3[3]. In the following section we introduce such a randomized membership protocol which gives rise to uncorrelated, random scale-free overlays with tunable degree distribution exponent. Based on the notion of actively using phase transition phenomena, we further discuss the possibility to actively adjust the degree distribution exponent and thus adapt the collective properties of the overlay network. While maintaining the overall power law structure of the overlay, in critical situations such a scheme allows for example to attenuate the heterogeneity of nodes, by this sacrificing properties like ultra-small diameter or extreme robustness against random node failures in favor of better attack resilience.

Application Scenarios

Possible application scenarios involve probabilistically structured Peer-to-Peer systems making use of gossip-based protocols, epidemic algorithms, information spreading schemes, self-organized synchronization protocols (see for instance chapter 5) or other types of probabilistic distributed algorithms that are known to operate efficiently in scale-free networks. Based on the properties reviewed in section 2.3, the structure of the overlay is particularly suitable for systems being comprised of highly heterogeneous participants or requiring extreme robustness against random node failures for instance in situ-

[3] In the terminology of [Weaver, 1948] this may be viewed as deliberately *constructing overlays with disorganized complexity.*

ations of massive churn. The existence of massively connected hub nodes in a random scale-free topology could further be used in the development of control strategies for information spreading, synchronization, distributed consensus or collective decision-making processes. Due to these promising aspects, during the last couple of years the use of overlay topologies with scale-free structures has been advocated in a number of contexts [Lu and Fang, 2005; Fraigniaud *et al.*, 2005; Qi and Yu, 2008; Ktari *et al.*, 2009; Ferretti and D'Angelo, 2010]. Closest to the scenarios addressed in the following section, [Sarshar and Roychowdhury, 2004] consider the creation of scale-free overlay structures in dynamic Peer-to-Peer systems by means of a protocol which compensates link losses that are due to nodes dropping out of the system uniformly at random. Here, joining nodes as well as nodes that have lost a link due to a failing neighbor establish connections according to a preferential attachment rule. For this protocol and nodes leaving the network uniformly at random, it has been shown that a power law overlay emerges with the degree distribution exponent depending on the number of links being created as compensation for a lost connection. However, it has not been considered how preferential attachment can be implemented in practice efficiently in a way that it gives rise to uncorrelated networks. Furthermore, the proposed compensation mechanism is not suitable to adapt the degree distribution exponent of an existing scale-free overlay with a static set of participants or a fixed number of connections. Referring to a general lack of practicable local algorithmic schemes for the creation of scale-free overlays, in the recent article [Guclu and Yuksel, 2009] a distributed algorithm has been considered. The proposed scheme is again a variation of the preferential attachment model and differs from the original scheme insofar as hard limits on the maximum degree of nodes are imposed. Furthermore, a distributed protocol for the construction process has been presented. For this, connection targets are chosen from a restricted *horizon* that is obtained by sampling the subgraph starting at a bootstrap node up to a predetermined depth. For this scheme, the effects of different connectivity limits and horizon sizes on the performance of flooding based search and the distribution of node degrees have been studied. It remains unclear, to what extend results about uncorrelated random scale-free networks can be applied to the topologies emerging from this scheme and how the degree distribution exponent can be tuned. In some ways similar to the work presented in this chapter, in [Keyani *et al.*, 2002] a distributed scheme has been proposed which allows to monitor the rate of attacks on the most connected nodes in an overlay topology with scale-free features. In the same article, a node recovery strategy has

Related Work

been proposed which drives the connectivity distribution of an overlay topology from a Zipf to an exponential distribution and which thus reduces the risk of the network being disintegrated. However, a mechanism to recreate a scale-free overlay topology has not been considered. As such, the protocol discussed in the following section can be seen as complementary to the work presented in [Keyani *et al.*, 2002]. In particular, the distributed attack detection mechanism could be used in the context of our scheme in order to decide when to make a transition to a topological phase with reduced attack susceptibility.

4.2 Distributed Creation and Adaptation of Random Scale-Free Overlays

Assumptions

In what follows, we discuss a probabilistic and fully distributed rewiring protocol that gives rise to random, uncorrelated scale-free overlays. Prior to giving an in-depth description of the protocol, we first summarize some preliminaries and restrictions which will simplify its analysis. Some of these restrictions will be weakened in later sections, for others we will discuss how they can be met in practical systems. First and foremost, the protocol is targeted at transforming an existing overlay topology into a network whose degree distribution follows a power law with tunable exponent. As initial situation, we assume an arbitrary connected network topology which may have evolved from some arbitrary bootstrap process. For the mere functioning of the proposed scheme, apart from being connected no further qualities of the initial topology are required. However, as we shall see later, spectral properties of the initial overlay influence the efficiency of the rewiring scheme in terms of how many message exchanges are required to create a scale-free overlay with sufficiently random structures. We further assume that each of the n nodes in this initial overlay is identified by some numeric identifier $i \in \{1, \ldots, n\}$. For the sake of simplicity, in the following we assume that each of the n nodes is assigned a unique identifier. However in sufficiently large systems, per-node quantities i that are chosen uniformly at random - and which are thus not necessarily unique - can be used instead and in the practical evaluation of the protocol such a random assignment has been applied.

Equilibrium Scale-Free Overlays

Although we have argued in section 3.2.4 that both equilibrium and non-equilibrium situations are likely to occur in distributed systems making use of overlay topologies, in this section we *focus on equilibrium situations* for a number of reasons. First of all, suitable equilibrium models are naturally capable of producing uncorrelated, random networks that facilitate a reasoning in terms of the literature on random (equilibrium) scale-free network ensembles defined by a given degree sequence or degree distribution. Apart from the mere construction of scale-free structures, we further intend to particularly address the *adaptation of scale-free overlays* in situations where the set of participating machines is either static or resides in a dynamic equilibrium state, that is nodes randomly joining and leaving the system at roughly balanced rates. While we will consider extensions for such dynamic situations later in this section, for the sake of simplifying the analytical derivation of the proposed protocol in the following presentation we will restrict ourselves to a static situation without node fluctuations.

Equilibrium Model for Random Scale-Free Networks

Prior to considering a practical protocol, we first need a suitable equilibrium model that is capable of generating random power law networks with tunable degree distribution exponent. For this, the simple model which has been discussed for instance in [Goh *et al.*, 2001; Lee *et al.*, 2005], and with slightly different assumptions in [Chung and Lu, 2002], can be used. Here it is assumed that each node with ID $i \in \{1, \ldots, n\}$ is assigned a weight

$$w_i = i^{-\alpha} \tag{4.1}$$

for some parameter α in the range $(0, 1)$. Then, m edges are created between pairs of nodes (i, j) which are chosen at random with probabilities p_i and p_j respectively. These probabilities are given by the normalized weights

$$p_i = \frac{w_i}{\sum_{k=1}^{n} w_k}. \tag{4.2}$$

It is easy to see that such a construction procedure gives rise to a random network ensemble in which the expected degree of each node i is fixed and determined by the weights w_i and the number of generated edges m. For the case where the creation of multiple edges between a single pair of nodes as well as the construction of self-loops are being prevented - for example by simply skipping their creation - it has been argued in [Lee *et al.*, 2005] that the weight defined above produces a random statistical ensemble of power law networks with degree distribution

$$P_G(k) \propto k^{-(1+\frac{1}{\alpha})}$$

Hence, for $\alpha \to 0$, the model yields a power law network with degree distribution exponent $\gamma \to \infty$, while for $\alpha \to 1$ the exponent γ converges to two. It thus provides a simple numeric parameter that can be adjusted to effectuate (scale-free) power law networks with arbitrary degree distribution exponent γ in the range $(2, \infty)$.

Rewiring Scheme

In order to apply this model in a practical overlay network, a distributed mechanism is required which creates edges between nodes i and j which are chosen according to the probability given in equation 4.2. For this, we assume that we start with a random, connected overlay topology consisting of n nodes and m edges. In practice, this initial topology may emerge by means of an arbitrary bootstrapping method that connects joining nodes to existing participants either deterministically or at random. In order to transform this initial topology into a random power law network, one can *view the above model as a rewiring scheme in which edges are gradually replaced rather than being created*. The rewiring scheme must then ensure that edges between node pairs emerge with the desired node-dependent probabilities.

Sampling Nodes in Peer-to-Peer Systems

For this, a node initiating the rewiring of an edge must be able to sample two new endpoints for the edge being rewired according to the probability measure given in equation 4.2. Here we are interested only in fully distributed mechanisms. In the context of large dynamic Peer-to-Peer systems, different mechanisms have been proposed. In [Jelasity *et al.*, 2007], a gossip-based mechanism has been presented which can be used to provide nodes with a fixed-size list of random node samples. In [Gkantsidis *et al.*, 2006; Zhong *et al.*, 2008], the use of random walks for non-uniform random sampling in Peer-to-Peer systems is studied and analytical arguments for their convergence behavior are given. Here, we intend to apply the same approach to our rewiring scheme, assuming that nodes initiating the rewiring of an edge sample two new endpoints by means of two random walks through the current network topology. For a simple unbiased random walk, the probability $\pi_i(l)$ to find the walker after l random walk steps at node i converges to

Sampling by Random Walks

$$\pi_i(l) \to \frac{d_i}{N \cdot \bar{d}} \; (l \to \infty)$$

Metropolis-Hastings Algorithm

where $\bar{d}$ is the average node degree of the network. However here we rather wish to sample nodes according to the probabilities given in equation 4.2. For this it is necessary to introduce a bias in the selection of a random walk's next

random target that influences the transition probabilities accordingly. Considering a random walk in a connected overlay topology $G(V, E)$ as Markov chain with state space V and stationary distribution π, this bias can be configured by means of a Metropolis-Hastings chain [Metropolis *et al.*, 1953; Hastings, 1970; Azar *et al.*, 1992] in such a way that a desired stationary distribution π holds. In general, this can be achieved by introducing a bias as shown in the following transition matrix T:

$$T_{ij} = \begin{cases} \frac{1}{d_i} \min\left\{ \frac{\pi_j}{\pi_i} \frac{d_i}{d_j}, 1 \right\} & (i,j) \in E, i \neq j \\ 1 - \frac{1}{d_i} \sum_{(k,i) \in E} P_{ik} & i = j \\ 0 & (i,j) \notin E \end{cases} \tag{4.3}$$

Existence of Stationary Distribution

Here d_i denotes the current degree of node $i \in V$ and an entry T_{ij} gives the probability that a random walk residing at node i moves to node j. The fact that this transition matrix has stationary distribution π follows from the reversibility of the underlying Markov chain, as well as from its irreducibility (assuming a connected network topology) and aperiodicity (self-loops are possible). Under these restrictions, the Markov chain convergence theorem ensures that the probability $\pi_i(l)$ to find a random walker that has been started in an arbitrary node resides at node i after l steps converges to π as l goes to infinity.

Desired Stationary Distribution

From this, one can easily configure a random walk bias that results in a stationary distribution suitable to sample nodes in a way that - after all edges of the initial topology have been rewired - a random scale-free network with degree distribution exponent γ emerges. From the probability p_i in equation 4.2 and the fact that it gives rise to a power law network with degree distribution exponent $1 + \frac{1}{\alpha}$, we obtain the following required stationary distribution:

$$\pi_i^{\gamma} = \frac{i^{\frac{-1}{\gamma-1}}}{\sum_{k=1}^{n} k^{\frac{-1}{\gamma-1}}} \tag{4.4}$$

Random Walk Bias

Equation 4.3 and $\frac{\pi_j}{\pi_i} = \left(\frac{i}{j}\right)^{\frac{1}{\gamma-1}}$ yields the following transition matrix P:

$$P_{ij} = \begin{cases} \frac{1}{d_i} \min\left\{ \left(\frac{i}{j}\right)^{\frac{1}{\gamma-1}} \frac{d_i}{d_j}, 1 \right\} & (i,j) \in E, i \neq j \\ 1 - \frac{1}{d_i} \sum_{(k,i) \in E} P_{ik} & i = j \\ 0 & (i,j) \notin E \end{cases} \tag{4.5}$$

Thus, a random walk with the above bias can be used to sample endpoints of edges and thus perform rewiring operations that effectuate a random power law network with a particular degree distribution exponent.

4.2.1 Bounding the Random Walk Length

Convergence of Random Walk Schemes

Since our goal is to practically apply the above sampling strategy in a distributed rewiring scheme, an important question that needs to be answered is how many steps a random walk with the above bias needs to take before the probability $\pi_i(l)$ to find it in a node i after l steps is sufficiently close to the desired stationary limit π_i. In the rewiring protocol that will be presented in section 4.2.2, this translates to the number of messages that need to be exchanged for a single rewiring operation. To assess the convergence behavior of the random walk, one first needs to give a formal definition of when two probability distributions π and π' shall be considered *sufficiently close*. For this we use the usual definition of the *total variation distance* D which - for two probability measures π and π' and a finite state space V - can be defined as follows:

Total Variation Distance

$$D(\pi', \pi) = \frac{1}{2} \sum_{v \in V} |\pi'(v) - \pi(v)|$$

Bounding the Walk Length

The configuration of the random walk bias according to equation 4.5 and the Markov convergence theorem ensure that $D(\pi(l), \pi) \to 0$ for $l \to \infty$. For an arbitrarily chosen total variation distance $\epsilon > 0$ we can then assess the number of steps l our random walk needs to take until $D(\pi(l), \pi) \leq \epsilon$. In order to bound the minimally required number of steps l, arguments put forth in [Sinclair, 1992] can be used. Here it is argued that an upper bound for l is given by

$$l \leq \frac{ln\left(\frac{1}{\pi_s \epsilon}\right)}{1 - |\lambda_2(P)|}$$

Bounding the Eigengap of Stochastic Matrices

where π is the stationary distribution of the Markov chain, s is the node in which the random walk is started and $\lambda_{max}(P) = max\{\lambda_2, |\lambda_{|V|-1}|\}$ is the eigenvalue with the second largest absolute value in the descending sequence of scalar *eigenvalues* $1 = \lambda_1(P) \geq \lambda_2(P) \geq \ldots \geq \lambda_{|V|}(P) \geq -1$ which satisfy

$$P \cdot x_i = \lambda_i(P) \cdot x_i \tag{4.6}$$

for a corresponding *eigenvector* x_i. Thus, finding an upper bound for the number of random walk steps involves to find a lower bound for the second

largest eigenvalue $\lambda_2(P)$ of the transition matrix. Unfortunately, obtaining good bounds for the eigenvalues of stochastic matrices is a non-trivial task. Nevertheless, based on the canonical path approach introduced in [Diaconis and Stroock, 1991; Sinclair, 1992], analytical arguments concerning the convergence behavior of random walks with Zipf stationary distributions have been put forth in [Zhong *et al.*, 2005; Zhong *et al.*, 2008]. In the following we apply these arguments to the particular random walk strategy considered in this section. In [Zhong *et al.*, 2008] it has been argued that, if the stationary distribution π is a Zipf-like distribution, a lower bound for the so-called *eigenvalue gap* $1 - |\lambda_2(P)|$ can be given as

$$1 - |\lambda_2(P)| \geq \frac{\pi_{min}}{D \cdot d_{max}}.$$

Bound for Zipf Stationary Distributions

Here D denotes the diameter of the network topology upon which the random walk operates, π_{min} is the minimum probability ascribed to any vertex by the stationary distribution and d_{max} is the maximum degree of any vertex in the network. Thus, for the special case of Zipf stationary distributions, an asymptotic upper bound for the random walk length l required to achieve a total variation distance smaller than ϵ is given as [Sinclair, 1992; Zhong *et al.*, 2008]:

$$l \leq ln\left(\frac{1}{\pi_s \epsilon}\right) \cdot \frac{D \cdot d_{max}}{\pi_{min}} \tag{4.7}$$

For a random walk strategy configured to eventually effectuate a degree distribution exponent γ and thus stationary distribution π^γ, for the inverse stationary probability of the starting node s, the following bound holds:

$$\frac{1}{\pi_s^\gamma} = s^{\frac{1}{\gamma-1}} \cdot \sum_{k=1}^{n} k^{\frac{-1}{\gamma-1}} \leq s^{\frac{1}{\gamma-1}} \cdot \sum_{k=1}^{n} 1 = n \cdot s^{\frac{1}{\gamma-1}}$$

Bounding π_{min}

While this holds for arbitrary $\gamma \in [2, \infty)$ and starting nodes s, for the special case of node n we can give a better bound by observing that - due to the increasing skewness - node n is ascribed minimal probability for $\gamma = 2$, that is for $\gamma \in [2, \infty)$

$$\pi_{min}^\gamma \geq \pi_{min}^{\gamma=2}$$

holds. With this, we can bound the inverse minimal probability by considering the logarithmic growth of the harmonic series, so that

$$\frac{1}{\pi_{min}^{\gamma}} \leq \frac{1}{\pi_{min}^{\gamma=2}} = n \cdot \sum_{k=1}^{n} \frac{1}{k} = n \cdot H_n = n \cdot (ln(n) + \tau + r_n) \tag{4.8}$$

Bounds for Power Law Networks

where $\tau \approx 0.5772$ denotes the Euler-Mascheroni constant and $r_n \to 0$ in the limit of large n. We can now consider a bound for the case where the initial topology is an n node power law network with initial exponent γ_i, the rewiring protocol being applied to change the exponent to γ. In this case, diameter and maximum degree of the initial topology can - with high probability - be bounded as $O(ln(n))$ and $O(n^{\frac{1}{\gamma_i}})$ respectively [Zhong *et al.*, 2008]. Combining these arguments with the bounds given in equation 4.8 and 4.7, an asymptotic upper bound for the minimal length l of a random walk started in node s that is required to achieve total variation distance smaller than ϵ can then be given as follows:

$$l = O\left(ln\left(\frac{n \cdot s^{\frac{1}{\gamma-1}}}{\epsilon}\right) \cdot ln(n)^2 \cdot n^{1+\frac{1}{\gamma_i}}\right) \tag{4.9}$$

Tightness of Bounding Techniques

This theoretic bound scales worse than linear with the network size n. However the underlying bounding technique is not necessarily tight, that is the actual convergence behavior of a random walk can be considerably better. Since at present, obtaining tight upper bounds for the convergence of Markov chains in complex network topologies is an open research issue, in section 4.2.3 we present simulations that have been performed to derive practicable random walk lengths empirically. As will be argued later, the results of these simulations suggest that the proposed adaptation scheme can be practically implemented with reasonable random walk lengths. Although these results suggest that the analytical bound derived above are not tight and thus uninformative with respect to the actual performance in practice, they can nevertheless be helpful when studying in how far the parameters being associated with the random walk bias, the initial network topology or the targeted topology influence the number of steps required for the sampling of random edges. From equation 4.9 one can for example infer that the upper bound for the minimal random walk length will generally be higher when wanting to effectuate highly skewed scale-free networks with exponents close to two.

4.2.2 Protocol Definition

Protocol Overview

The equilibrium model and the sampling strategy laid out above suggest a rewiring protocol that consists of the following three basic operations: (1) In periodic intervals, a node a selects an edge to a random neighbor b that has not yet been rewired. (2) A random walk with the bias given in the transition matrix shown in equation 4.5 is started to sample two nodes x and y with probabilities proportional to π_x and π_y respectively. (3) The edge (v, w) is replaced by the edge (x, y) and the latter is marked as having resulted from a rewiring operation. After all m edges of the overlay have been rewired, a power law overlay is eventually obtained whose exponent depends on the particular choice of the random walk bias defined in equation 4.5. In the algorithms 1 - 4, we give a detailed algorithmic description of the protocol. In these algorithms, d_v denotes the degree of node v, i_v is the ID of node v and *self* denotes the node at which the code is being executed. We further assume that nodes have information about the IDs and the degrees of their nearest neighbors.

Rewiring Procedure

Protocol Main Loop

The detailed algorithm of the main program loop that is responsible for initiating random walks is shown in algorithm 1. Rewiring operations are initiated by nodes in regular intervals only for those edges that have not yet been rewired. By this means, at most m rewirings are performed where m is the number of edges in the initial random network topology. The number of rewiring operations and thus message transfers taking place within a certain time interval can be adjusted by choosing an appropriate (network-size dependent) *delay* time during which a node is inactive. When a node with an unmarked edge wakes up, a rewiring operation is initiated. In order to prevent both endpoints of an edge to initiate rewiring operations for the same edge, rewirings are only started by the node with higher degree or - if the degrees are equal - by the node with the smaller ID. As we shall see later in section 4.2.3, the choice of letting a rewiring be initiated by the better connected endpoint can improve the performance of the scheme. To find the endpoints of a new edge by which the previously unmarked edge shall be replaced, a node initiates a biased random walk through the overlay (lines $6 - 11$). In order to prevent nodes from disconnecting themselves from the network we further assume that only edges from nodes with at least two neighbors are rewired.

Algorithm 1: Main Loop of Adaptation Protocol

```
1:  loop
2:    Sleep(delay)
3:    if neighbors.Count > marked.Count then
4:      n = RandomUnmarkedNeighbor()
5:      if d_n > 1 && d_self > 1 &&
        (d_self > d_n || (d_self = d_n && i_self < i_n)) then
6:        {Initiate random walk}
7:        msg.Hops ← 0
8:        msg.a ← self
9:        msg.b ← n
10:       msg.target ← null
11:       Send({walk, msg}, n)
12:     end if
13:   end if
14: end loop
```

Biased Random Walk Strategy

When a node v receives a random walk message, it needs to ensure that the message is forwarded with the bias given in equation 4.5. In algorithm 2, this is done in lines $14-21$. Comparing the algorithm with the stochastic matrix P defined in equation 4.5, here we select a neighbor uniformly at random and draw a random value uniformly in the interval $[0,1]$ that indicates whether a transition along this edge occurs or whether the random walk stays in the current node. Considering that the probability of a neighbor j to be chosen (uniformly at random) from the list of neighbors of node i is $\frac{1}{d_i}$, the fact that this strategy yields the same transition probabilities as given in equation 4.5 can be seen by simple multiplication. It is however significantly easier to implement in practice than directly drawing the next random walk step with the probabilities given in equation 4.5.

Sampling Edges

Different schemes can be imagined by which the two endpoints v and w of a new edge (v,w) are being sampled. The node initiating the rewiring could for example start one random walk for each endpoint of the new edge, collect the target nodes of both walks and connect them to each other. In order to simplify the implementation, in algorithms 1 and 2 we consider that both endpoints of the new edge are being sampled in a single random walk of length $2 \cdot l$. For this we assume that after l steps, the node at which the random walk currently resides is stored in a field *target* of the message being forwarded. By this, all

Algorithm 2: Node receives $\{walk, msg\}$

1: $msg.Hops \leftarrow msg.Hops + 1$
2: **if** $msg.Hops = l$ **then**
3: {Store Endpoint}
4: $msg.target \leftarrow self$
5: **else if** $msg.Hops = 2l$ **then**
6: {Rewire}
7: **if** $!neighbors.Contains(msg.target)$ && $msg.target \neq self$ **then**
8: $Send(\{disconnect, msg.a\}, msg.b)$
9: $Send(\{disconnect, msg.b\}, msg.a)$
10: $Send(\{connect, self\}, msg.target)$
11: $Send(\{connect, msg.target\}, self)$
12: **end if**
13: **else**
14: $n \leftarrow self.RandomNeighbor$
15: **if** $random.Next() \leq \frac{d_{self}}{d_n} \left(\frac{i_{self}}{i_n}\right)^{\frac{1}{\gamma a - 1}}$ **then**
16: {Forward Random Walk}
17: $Send(\{walk, msg\}, n)$
18: **else**
19: {Self-Loop}
20: $Send(\{walk, msg\}, self)$
21: **end if**
22: **end if**

information related to a rewiring operation is stored in the random walk message and the node at which the random walk arrives after $2 \cdot l$ steps can initiate the rewiring.

Rewiring Operation

The node at which the random walk eventually resides creates a connection to the *target* node stored in the message while initiating the deletion of the edge between node a that has started the random walk and its neighbor b. As can be seen in algorithm 4 a disconnection requires - apart from removing the edge - no further action at the side of the node from which the edge is removed. As shown in algorithm 3, both endpoints of the newly created edge mutually mark each other in order to prevent the edge from being rewired in future invocations of the protocol. We emphasize that this is to prevent unnecessary

Algorithm 3: Node receives $\{connect, y\}$

1: $neighbors.Add(y)$
2: $marked.Add(y)$

Algorithm 4: Node receives $\{disconnect, b\}$

1: $neighbors.Remove(b)$

rewiring operations and thus message exchanges rather than being required for the functioning of the protocol. A schematic illustration of a single rewiring operation is shown in Figure 4.1. Here, the node labeled with 0 initiates a random walk of length $l = 3$, thus eventually replacing edge e by e'.

Number and Size of Message Transfers

Concluding the description of the proposed protocol, we consider the size and number of messages that need to be sent across the network. Sampling the two endpoints of the new edge requires at most $2 \cdot l$ messages [4], where l is the number of steps taken by a single random walk to sample a node with a probability sufficiently close to the stationary distribution π. Once both endpoints of the new edge have been sampled, the rewiring requires two messages to disconnect nodes a and b and one message to connect to the node *target* that has been stored in the random walk message. Since the IDs of the initial node, its neighbors and the intermediate *target*, as well as the current hop count need to be stored in the random walk message, the required number of bits for a message is logarithmic in the number n of nodes in the system. Hence, the number of bits that need to be transferred per rewiring operation is $O(l \cdot log(n))$. Since one rewiring operation is performed for each of the m edges in the initial overlay topology, the total number of bits that needs to be transferred in order to transform it into a random scale-free network with the desired exponent is $O(m \cdot l \cdot log(n))$. We further assume that each node stores one additional bit per neighbor, indicating whether the edge to this neighbor has previously been rewired or not.

Communication Cost of Protocol

[4] *At most* $2 \cdot l$ since self-loops are allowed to ensure aperiodicity of the underlying Markov chain. While a self-loop is considered as one step of the random walk, it does not entail a message exchange.

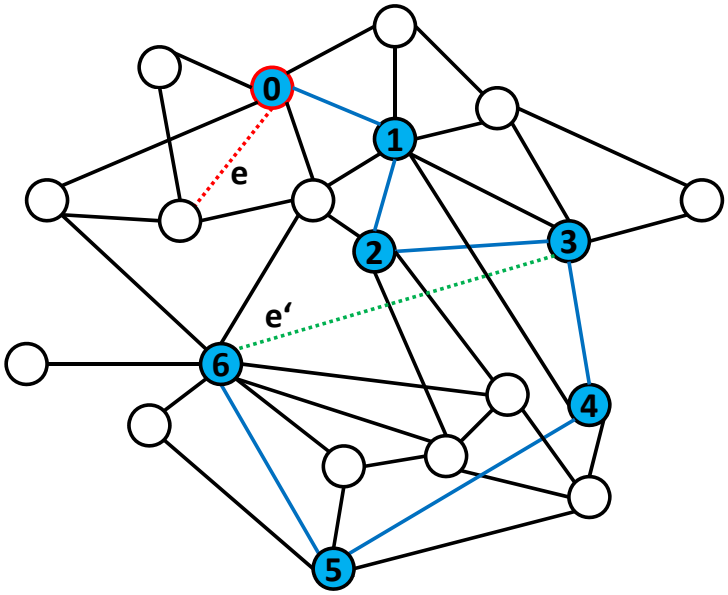


Figure 4.1: Example sequence of random sampling and rewiring of edge e initiated by node 0 by means of a random walk with length $2 \cdot l = 6$

Join and Leave Procedure

Dynamic Equilibrium States

So far, we have considered a protocol which is suitable for the gradual rewiring of existing edges in a *static topology* while neglecting the impact of nodes which dynamically join or leave the system. However, it is rather simple to extend this protocol in a way that it handles *node fluctuations in dynamic equilibrium states* where the size of the network remains roughly constant with nodes joining and leaving the system randomly at roughly balanced rates. In this case, a simple join procedure as shown in algorithms 5 and 6 is sufficient. The node wishing to join the overlay chooses its own ID for instance by drawing it uniformly at random from the range $[1, n]$ with n being a fixed maximum size of the network. It then creates connections to k existing nodes in the overlay, which are obtained by an arbitrary bootstrapping process[5]. We further assume that newly created edges are labeled as *unmarked*, thus being subject to future rewiring operations according to the protocol presented above. Nodes leaving the system do not require any particular handling, except for the removal of stale links as shown in algorithm 7.

Algorithm 5: Node joins the system

```
1: {Draw random ID}
2: i_self ← Random(1, n)
3: for i ← 0; i < k; i ← i + 1 do
4:     x ← ArbitraryBootstrapNode()
5:     {Create link to bootstrap node}
6:     Send({join, self}, x)
7:     neighbors.Add(x)
8: end for
```

Algorithm 6: Node receives $\{join, w\}$

```
1: neighbors.Add(w)
```

Algorithm 7: Node finds that neighbor w left the system

```
1: neighbors.Remove(w)
2: if marked.Contains(w) then
3:     marked.Remove(w)
4: end if
```

4.2.3 Experimental Results

Overview of Evaluation

Having given a description of the rewiring protocol as well as some analytical arguments about its convergence behavior, in this section we present simulation results that have been obtained using a prototypical implementation of the proposed scheme in the simulation environment TOPGEN [Scholtes *et al.*, 2008b]. The evaluation is split up in two parts. In a first step we seek to establish, by simulation, a practicable lower bound for the minimally required random walk length l. We further study the influence of the initiating node's degree on the convergence time of a random walk. Based on these results, in a second step we then simulate the rewiring protocol and study its influence on a network's degree distribution.

[5] Commonly used bootstrapping mechanisms include for instance the maintenance of a web accessible resource which contains an updated list of randomly chosen nodes that are currently online or the use of a hard-coded list of explicitly managed bootstrap nodes.

Random Walk Length

Random Walk Convergence Experiments

In section 4.2.1, we have considered theoretic asymptotic upper bounds for the required number of steps l in an n node power law network. In the following, these theoretic results will be complemented by an experimental study of the random walk's convergence behavior. With this we intend to derive a random walk length that provides a reasonable trade-off between the imposed number of messages and the resulting total variation distance. Since it is crucial for the scalability of the proposed rewiring scheme, we further investigate how the minimally required random walk length changes as the network size is varied. The results presented in the following have been obtained as follows. In each simulation run a number R of random walks was started from a randomly chosen node in a randomly generated scale-free network[6]. In each of these R simulated random walks, a hit counter was increased in the node at which the random walk resided in the l-th step. After R random walks had been simulated, the total variation distance was computed based on the observed hit frequencies and the stationary distribution expected for the chosen random walk bias. Depending on the network size and the minimum probability π_{min} of the expected stationary distribution, the number of simulated random walks R was chosen in a range between 10^6 and 10^8. In particular, it was chosen such that nodes with minimum stationary probability π_{min} were expected to be hit sufficiently often to reasonably argue about the total variation distance. The above procedure was then repeated for ten randomly generated scale-free networks and different starting nodes and the average total variation distance was computed.

Random Walk Convergence Results

Figure 4.2 shows the random walk length l minimally required for the average total variation distance to fall below $\epsilon = 0.05$. Results are shown for different network sizes and for random walks configured to effectuate - when used in the rewiring protocol - random scale-free topologies with degree distribution exponents 2.1, 2.5 and 3.5. Rather than the linear scaling behavior suggested by the theoretical upper bound presented in section 4.2.1, the observed required length l rather scales in a sub-linear fashion. The observation that the actual convergence behavior is significantly better than the theoretical upper bound is consistent with observations made in [Zhang *et al.*, 2008] and indicates that the rewiring scheme can be implemented efficiently in practice. Informally, this observed fast convergence can be attributed to the small diam-

[6]For the sake of simplicity, here we have used networks generated by the preferential attachment scheme presented in [Barabási and Albert, 1999] and described in section 2.3.5

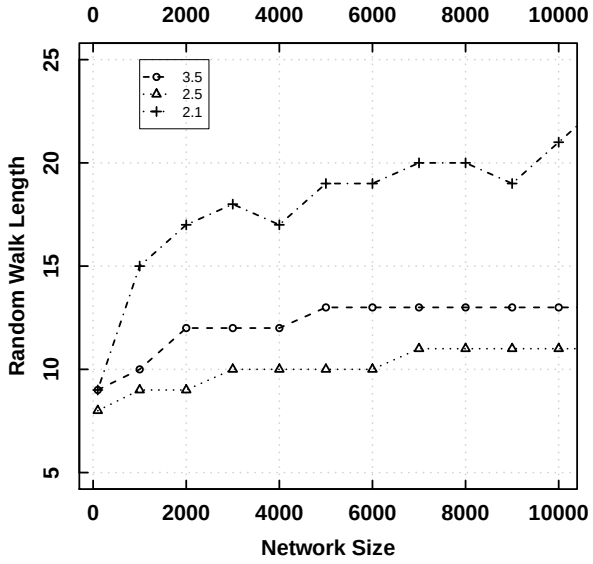


Figure 4.2: Minimum random walk length l required to achieve $D(\pi(l), \pi) \leq 0.05$ in Barabási-Albert networks with random walk biases configured to effectuate exponents 2.1, 2.5 and 3.5 (Lines are drawn to guide the eye)

eter of scale-free networks and the absence of network cuts with small capacity[7]. While no experimental results are presented for further network topologies, these arguments suggest that a similar scaling behavior of the minimally required random walk length is likely to hold for other network topologies that are known to have good expansion characteristics.

Dependence of Convergence on Node Degree

In networks with highly heterogeneous connectivity, a further interesting question is how the choice of the starting node of a random walk influences the total variation distance that can be achieved by a fixed random walk length. To investigate this, a number of scale-free networks with 1000 nodes and 5000 edges were created using a preferential attachment scheme and a large number of random walks configured for a targeted exponent $\gamma = 3$ were started *from each* node of the network[8]. The frequency with which nodes were the final target of a random walk was recorded and the total variation distance to the expected stationary distribution was computed for each series of random walks starting at a particular node. Figure 4.3 shows the correlation between the degree of the node at which a random walk started and the total varia-

[7] A network cut being defined as a partition of the node set, its capacity being the number of edges connecting both partitions.

[8] Here *large* again means sufficiently large to reasonably compute the total variation distance.

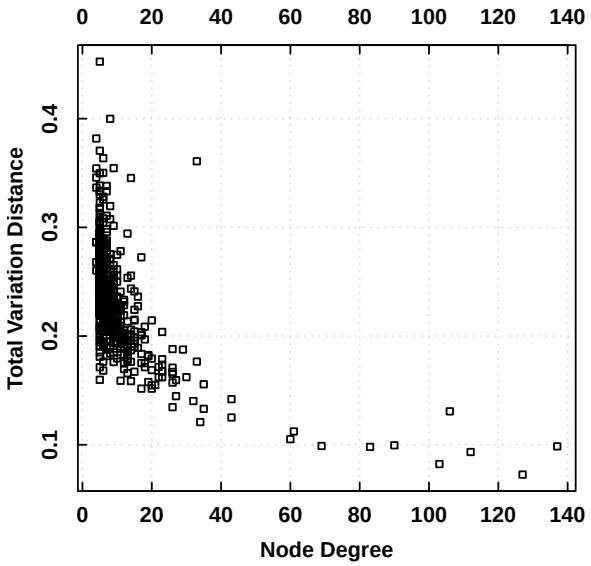


Figure 4.3: Correlation between degree of starting node and average $D(\pi(l), \pi)$ in 1000 node Barabási-Albert networks with $\gamma = 3$ and $l = 5$

tion distance that was achieved[9]. This result underpins the intuition that random walk sampling schemes being started at highly connected nodes converge faster. This is because a random walk starting at a node with high degree can potentially reach a large number of nodes even in the first step. In the extremal case where the starting node has connections to all other nodes, even a single step is sufficient. In the protocol presented in 4.2.2, this justifies the choice that rewiring operations for an edge (i, j) are initiated by the node with higher degree. In fact, in network topologies with highly skewed degree distributions this results in a large number of rewiring operations being initiated by high degree nodes, thus facilitating the use of short random walk lengths.

[9]Please note that in these experiments there was no correlation between the (random) assignment of node IDs and the degree in the initial topology.

Transformation of Degree Distribution

Description of Experiments

We now turn to the question of how the proposed rewiring protocol influences the degree distribution of a network topology. All results presented in the following figures have been obtained for networks consisting of 5000 nodes and roughly 25000 edges. Initial topologies upon which the protocol was started were created by means of the Barabási-Albert preferential attachment model as well as the Erdös/Rényi model for classical random graphs. Based on the results presented in the previous section, a random walk length $l = 20$ was chosen for the following experiments. In each simulation run, the protocol presented in algorithm 4.2.2 was applied by all nodes in a network topology with initially unmarked edges until all edges had been rewired (and thus marked). The *delay* interval between individual rewiring iterations was chosen such that - on average - a single rewiring took place per time unit. However, this choice has been made merely for illustration purposes since it allows to trace the evolution of a network topology as links are being rewired progressively. Clearly, performing more than one per rewiring at a time would be more appropriate in actual applications of the protocol. When using a *delay* value so that one rewiring is expected to take place per unit of simulated time, for the chosen network size an adaptation cycle is expected to be completed after roughly 25000 time steps.

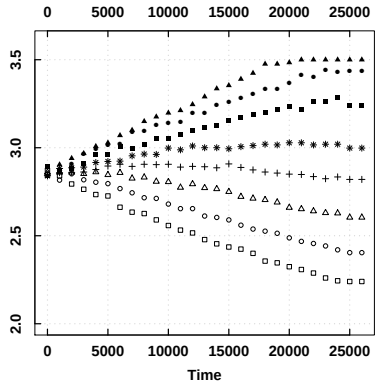
Measuring and Fitting Methodology

The degree distribution of the network topology was computed each 200 time units and a fit to the current degree distribution exponent was performed. For this, an R implementation of the maximum likelihood power law fit procedure described in [Clauset *et al.*, 2007] was used. This procedure yields the fitted degree distribution exponent γ_f that holds with maximum likelihood, the minimum network degree d_{min} above which the fit holds, as well as the Kolmogorov-Smirnov (KS) statistic D. In general, better fits result in smaller values of D, thus allowing to evaluate whether the "power law nature" of the degree distribution is strengthened or fades away under the application of the rewiring scheme. All results are averages of at least five independent applications of the protocol on randomly chosen network realizations of identical size. For more details on the maximum likelihood fitting procedure, we refer the reader to the description in the following section 4.3.1.

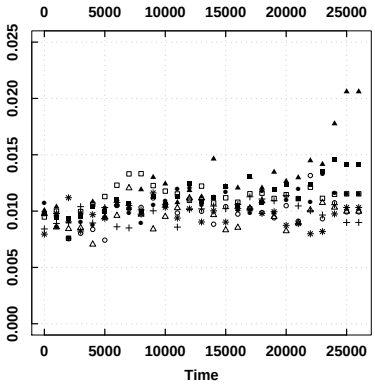
Static Topologies

In the following, we first consider a simple static situation in which no nodes enter or leave the initial network topology. Figures 4.4(a) and 4.5(a) show the effect of the proposed protocol on the degree distribution of a network that was initially created by the Barabási/Albert (BA) and the Erdös/Rényi (ER) model respectively. For BA networks, the average fitted ex-

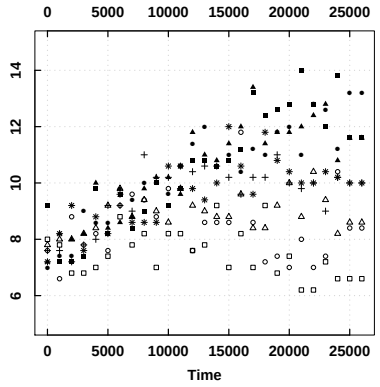
ponent γ_f of the initial topology was on average 2.9. For ER networks the used fitting procedure yielded 3.5 with an at least 10-fold value of the KS-statistic which reflects the fact that the initial degree distribution does not follow a power law. Figures 4.4(a) and 4.5(a) confirm that the protocol does lead to an adaptation of the degree distribution exponent of the overlay. In particular, the evolution of the Kolmogorov-Smirnov statistic D that is shown in Figure 4.4(b) demonstrates that the scale-free characteristic of Barabási-Albert scale-free networks is preserved. In Figure 4.6, log-log plots of log-binned degree distributions are shown for networks before and after the adaptation targeting at different exponents.



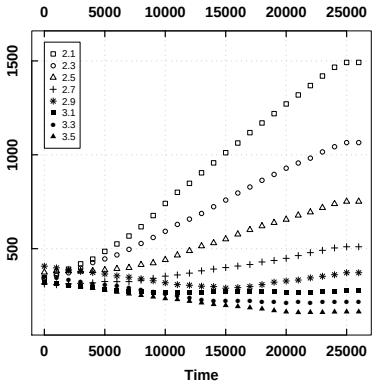
(a) Average fitted exponent γ_f



(b) Average Kolmogorov-Smirnov statistic D

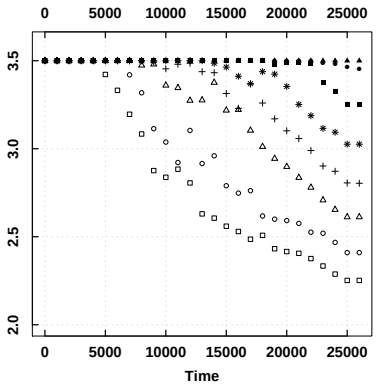


(c) Average minimum degree d_{min} above which fit holds

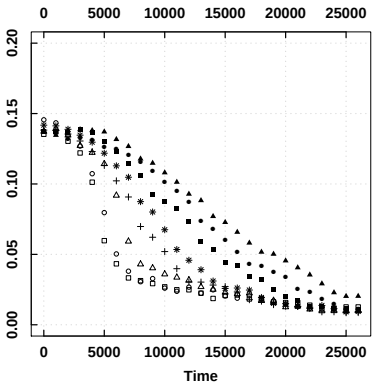


(d) Average maximum degree

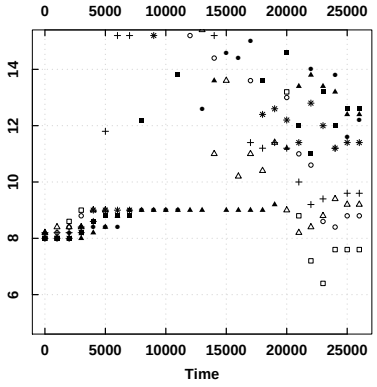
Figure 4.4: Time Evolution of 5000 node Barabási/Albert networks during adaptation runs with $\gamma \in [2.1, 3.5]$ and $l = 20$



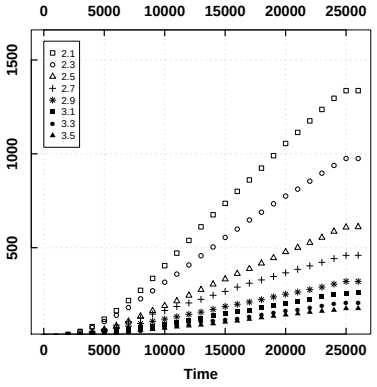
(a) Average fitted exponent γ_f



(b) Average Kolmogorov-Smirnov statistic D



(c) Average minimum degree d_{min} above which fit holds

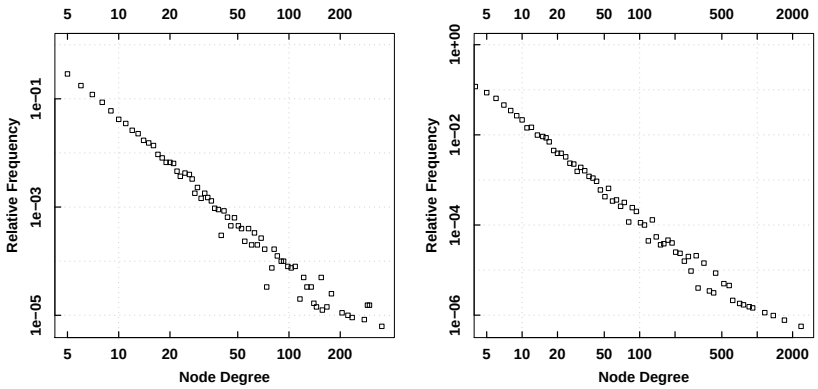


(d) Average maximum degree

Figure 4.5: Time Evolution of 5000 node Erdös/Rényi networks during adaptation runs with $\gamma \in [2.1, 3.5]$ and $l = 20$

For Erdös/Rényi networks, the roughly 10-fold decrease of the KS-statistic D that can be seen in Figure 4.5(b) indicates the emergence of scale-free characteristics, that is the power law fit to the degree distribution becomes more reliable. Plots of the degree distribution before and after the rewiring that are shown in Figure 4.7 underpin this assumption. In Figures 4.4(d) and 4.5(d), the evolution of the average maximum degree is shown for BA and ER networks. The results are consistent with the maximum degree expected in networks of the given size and with different degree distribution exponents. In Figure 4.8, the average fit parameters for the network topology eventually reached after the adaptation are shown. The results demonstrate that - as expected from the underlying theoretical model - the protocol can be applied to transform arbi-

trary initial topologies into random scale-free networks whose degree distribution follows a power law with an exponent reasonably close to the intended value.



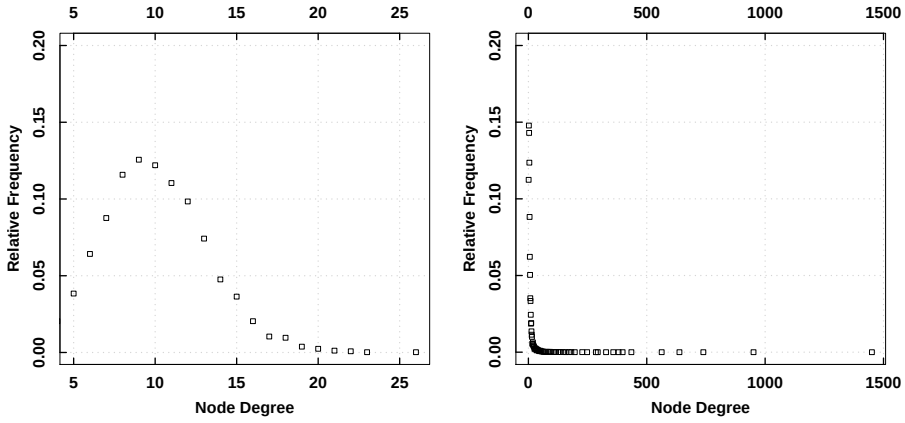
(a) Before Adaptation

(b) $\gamma = 2.1$

(c) $\gamma = 2.9$

(d) $\gamma = 3.5$

Figure 4.6: Log-Log plots of Log-binned degree distributions of 10000 node Barabási/Albert network before (a) and after adaptation with different target exponents (b-d)



(a) Before Adaptation (b) $\gamma = 2.1$

Figure 4.7: Log-binned degree distributions of 10000 node Erdös/Rényi network before (a) and after (b) adaptation with target exponent $\gamma = 2.1$

	γ_t	2.1	2.3	2.5	2.7	2.9	3.1	3.3	3.5
BA	γ_f	2.24	2.40	2.60	2.82	2.99	3.24	3.44	3.5
	D	0.012	0.01	0.01	0.01	0.01	0.01	0.012	0.02
	d_{min}	6.6	8.4	8.6	10	10	11.6	13.2	11.6
ER	γ_f	2.252	2.41	2.61	2.80	3.03	3.25	3.45	3.5
	D	0.013	0.01	0.01	0.009	0.009	0.01	0.012	0.02
	d_{min}	7.6	8.8	9.2	9.6	11.4	12.6	12.2	12.4

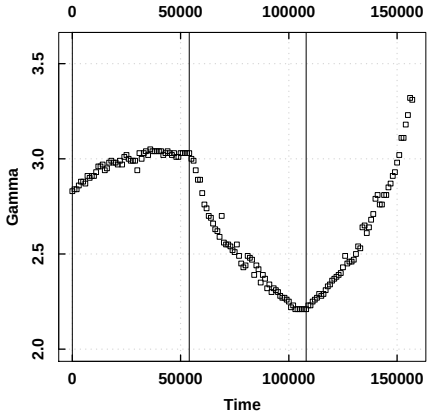
Figure 4.8: Average fitted exponent γ_f, Kolmogorov-Smirnov statistic D and minimum degree d_{min} above which the fit holds after adaptation with targeted exponents $\gamma_t \in [2.1, 3.5]$ for 5000 node Erdös/Rényi (ER) and Barabási/Albert (BA) networks with roughly 25000 edges

Multiple Subsequent Adaptations

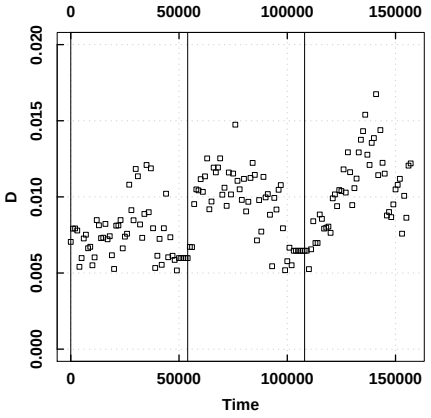
Above we have presented results for one *construction or adaptation cycle*, that means each connection being rewired once in order to perform a single transformation of the topology. Here we consider a situation where three subsequent adaptation cycles were initiated targeting at different degree distribution exponents. In Figure 4.9, results are shown for an initial Barabási/Albert network topology with 10000 nodes and roughly 50000 edges. The chosen random walk length of $l = 22$ was again consistent with the values found above. In Figure 4.9, time steps in which adaption cycles were initiated are indicated by vertical lines. The target degree distribution exponents were 2.9, 2.1 and 3.5 respectively. Again the results indicate that the proposed scheme is suitable to achieve the desired adaptation. Furthermore, Figure 4.9(b) shows how the Kolmogorov-Smirnov statistic and thus the reliability of the power law fit temporarily fades during the adaptation while it is being restored near the ends of a cycle. To illustrate the effect of the adaptation protocol on the heterogeneity of the network structure, in Figures 4.11(a) and 4.11(b) a 200 node network is shown after completion of two adaptation cycle targeting at $\gamma = 4$ and $\gamma = 2.1$.

Dynamic Equilibrium Topologies

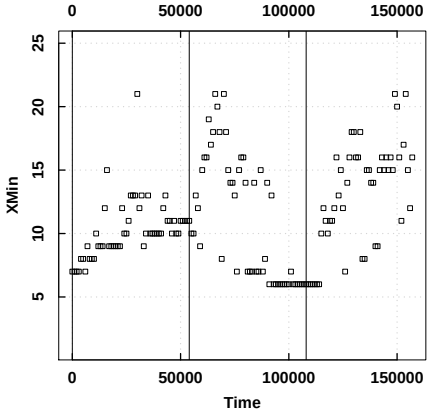
So far, only static topologies with a fixed set of nodes have been considered, that is no nodes were entering or leaving the topology. In the following, we present some preliminary results for dynamic topologies in which nodes join and leave the overlay uniformly at random at balanced rates, thus constituting a dynamic equilibrium state for a system with roughly fixed size. In the following experiments, a Barabási/Albert network with 5000 nodes and roughly 25000 edges has been used as initial topology. Results for Erdös/Rényi networks have shown to be identical which is why we have not included them here. In each simulated time step, a single node was removed uniformly at random while a single other node joined the network. Nodes joining the network were connected to $k = 5$ random bootstrap nodes according to the algorithm depicted in algorithm 5. Stale links left by failing nodes were removed immediately from the network. The delay parameter has been set to a value that was sufficiently small to compensate for the constant creation of unmarked links by joining nodes. Hence, the average number of links being rewired (and thus being marked) within a certain time interval was roughly equal to the average number of random (unmarked) edges created in that same time interval by joining nodes. Figure 4.10 shows the averaged results of simulations targeting degree distribution exponents $\gamma_t \in [2.1, 3.5]$. In particular, Figure 4.10(a) suggests that the degree distribution exponent approaches the targeted value as node fluctuations drive the gradual rewiring of links.



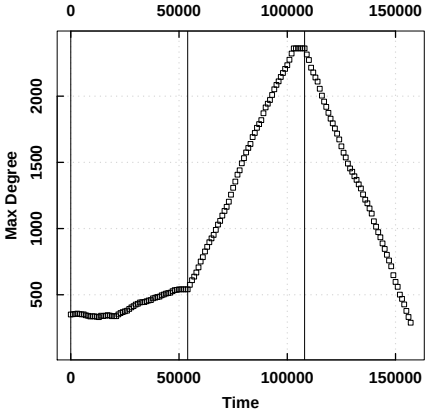
(a) Average fitted exponent γ_f



(b) Average Kolmogorov-Smirnov statistic D

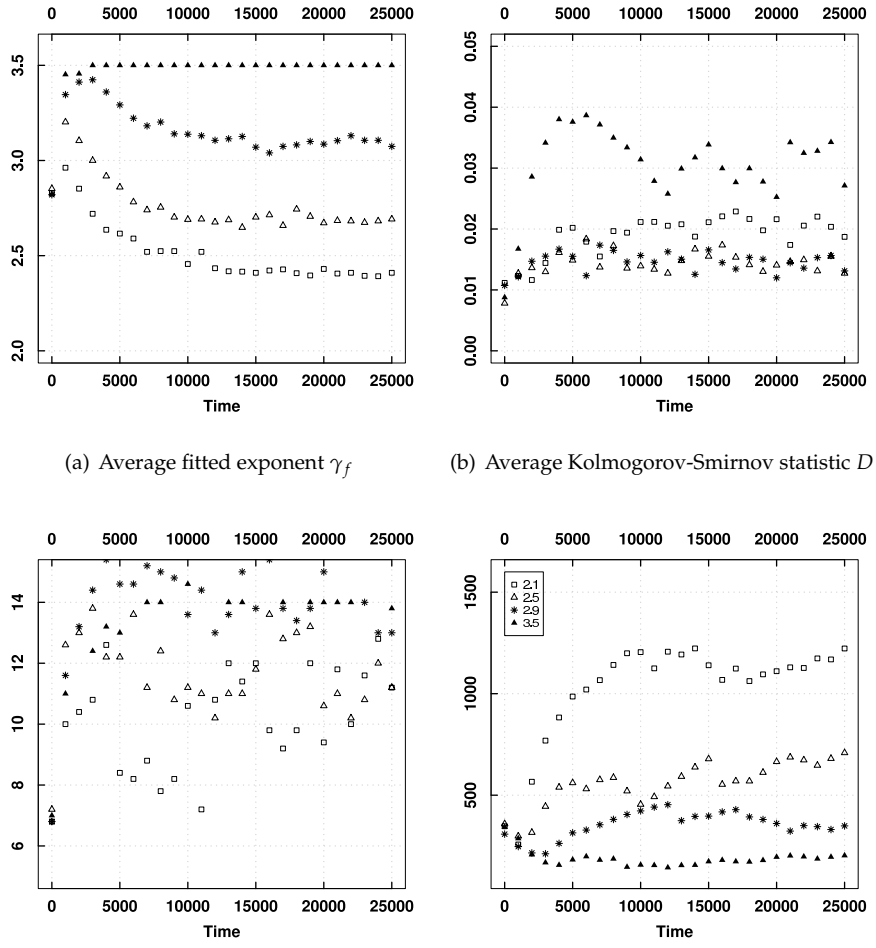


(c) Average minimum degree d_{min}



(d) Average maximum degree

Figure 4.9: Time Evolution of 10000 node Barabási/Albert networks during multiple adaptation cycles with $\gamma_0 = 2.9$, $\gamma_{54000} = 2.1$ and $\gamma_{108000} = 3.5$. Start times of adaptation cycles are indicated by vertical lines.



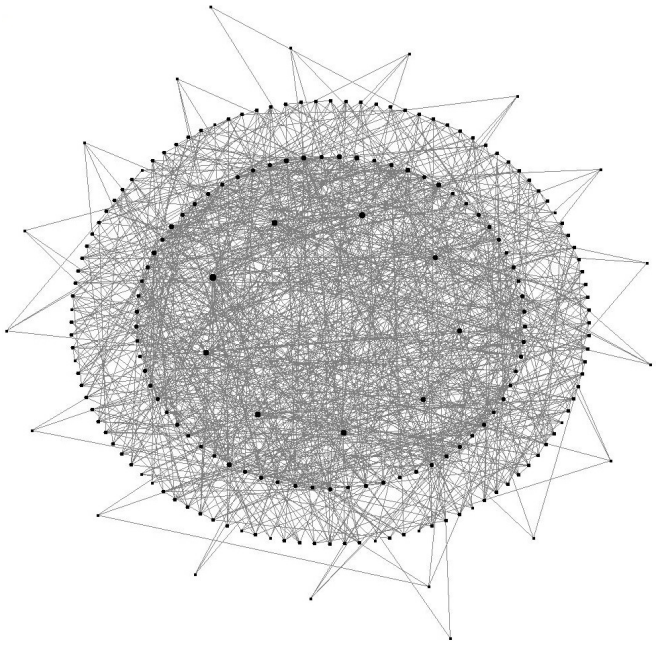
(a) Average fitted exponent γ_f

(b) Average Kolmogorov-Smirnov statistic D

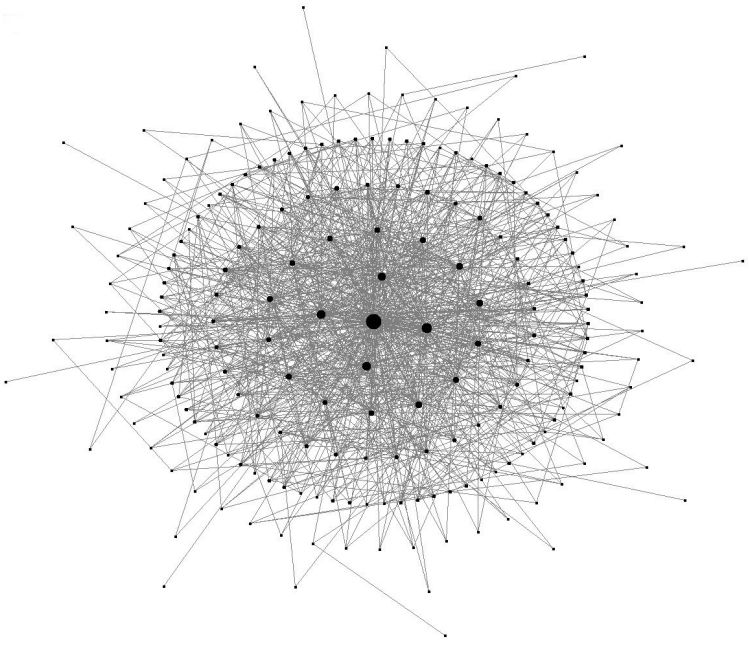
(c) Average minimum degree d_{min}

(d) Average maximum degree

Figure 4.10: Time evolution of dynamic 5000 node Barabási/Albert networks with random, uniform churn, $\gamma_t \in [2.1, 3.5]$ and $l = 20$



(a) $\gamma = 4$



(b) $\gamma = 2.1$

Figure 4.11: Network with 200 nodes and 1000 edges after adaptation targeting at different exponents γ

4.2.4 Summary and Perspectives

Equilibrium Statistical Mechanics Perspective of Rewiring Protocol

Prior to summarizing some contributions, open issues and threats to validity, we intend to relate the proposed protocol to the *thermodynamic and statistical mechanics perspective* that has been laid out in section 3.2. Assuming an overlay with a fixed set of participants, the rewiring protocol can actually be viewed as giving rise to a *canonical ensemble of random scale-free overlay networks* which is defined by three fixed, aggregate quantities n (number of nodes), $k \cdot n$ (number of edges) and γ (distribution of node degrees). The rewiring of all edges results in drawing a new realization from the associated statistical ensemble. In the physical view, this corresponds to a random rearrangement of particles within a given volume. Here, energy landscape and temperature determine the probability with which each possible particle arrangement occurs. Similarly, in the *equilibrium overlay* the rewiring of connections results in a random reconfiguration of edges within the space of all $\frac{n \cdot (n-1)}{2}$ pairs of vertices. Here, the degree distribution exponent γ determines which of the configurations are accessible and how statistical weights are assigned to particular realizations of the overlay. A more general way of looking at the probabilistic approach underlying the proposed protocol is that, instead of statically fixing a particular network structure, by means of a random rewiring scheme a huge number of different network topologies are allowed to emerge. By setting a particular bias in the random walk sampling scheme, probabilities in the resulting statistical ensemble can then be reallocated sensibly such that overlay topologies with advantageous characteristics emerge almost surely.

Macro-Level Guarantees, Micro-Macro Reasoning and Phase Transitions

Based on the existing literature on random networks with a given power law distribution, collective properties of the randomized overlay topology resulting from the rewiring protocol can further be related to aggregate, *thermodynamic parameters*, namely the number of nodes, the number of edges as well as the degree distribution exponent. In this sense, the proposed protocol facilitates the distributed construction of probabilistically structured overlays for which *thermodynamic guarantees* can be given. Based on the equilibrium model presented in [Lee *et al.*, 2005] as well as the configuration of stochastic matrices according to the Metropolis-Hastings method [Metropolis *et al.*, 1953], it is possible to analytically relate the bias used by individual nodes in the forwarding of random walk messages to the macroscopic degree distribution exponent. Furthermore, the protocol allows to specifically influence this forwarding bias in a meaningful way. At certain critical points in this parameter space, phase transitions occur and the collective properties of the overlay change abruptly.

Application Scenarios

A particular point at which such a transition occurs in power law overlays is when the degree distribution exponent leaves the scale-free phase between two and three. As the exponent increases beyond the critical point of three, the variance of the distribution becomes finite and super-hubs with unbounded connectivity are quickly becoming unlikely. In practice, the decision for a power law overlay topology with a particular skewness entails a balancing of possibly contradicting objectives. Scale-free overlays in which extremely powerful machines take the role of super-hubs facilitate efficient, probabilistic distributed search, a fast spreading of information or the fast formation of consensus. However, this comes at the price of imposing significant load on these machines as well as an increased susceptibility against attacks or failures of a small set of super-hubs. Changing the degree distribution exponent thus allows to rebalance load to a broader basis of less pronounced hubs. In critical situations, for instance when super-hubs are found to be under attack, rewiring operations that lead to a less skewed degree distribution can mitigate the risk that the structure of the overlay is destroyed while decreasing the performance of probabilistic search or content dissemination schemes. In this sense, the proposed protocol allows to fine-tune the heterogeneity in connectivity while retaining the overall power law structure of a network. Furthermore, we have argued that the proposed protocol is suitable to transform arbitrary connected topologies into random scale-free networks given that the expansion properties of the initial topology provide sufficiently fast convergence of the random walk sampling strategy. In Barabási-Albert and Erdös/Rényi networks, the random walk length required to provide sampling probabilities that are acceptably close to the stationary limit are found to be significantly smaller than the theoretical upper bounds. Based on the observed superior convergence behavior of random walks being started in hub nodes, the performance of the protocol benefits from the fact that rewiring operations are preferentially started by high degree nodes. Based on the results presented in this article, we thus argue that the proposed protocol is a simple and practicable approach to create and adapt thermodynamically structured overlays for large scale Peer-to-Peer systems.

Correlations and Random Walk Length

We conclude this section by summarizing some open issues that have not been considered so far. An important aspect in any practical application of the proposed rewiring scheme is the fact that any reasonably efficient implementation requires to accept moderate total variation distances. While this allows to keep the random walk length and thus the message overhead in an acceptable range, it poses a limit to the randomness of the resulting network

topology. Although small total variation distances suggest that the resulting correlations and thus the deviation of properties from those of truly random networks are rather moderate, a further investigation of these effects is an open issue. Regarding the minimally required random walk length, we have argued that the protocol seems to benefit from the fact that rewiring operations are preferentially started by high degree nodes. This is due to the observed correlation between the degree of the initiating node and the total variation distance achieved by a random walk. It thus seems to be reasonable to choose the length l for each random walk individually based on the degree of the node initiating it. This promises to further reduce the message overhead. A further potential improvement is the use of two-stage random walks which - in a first stage - move to highly connected nodes, and then - in a second stage - switch their bias to sample nodes according to the desired stationary distribution. However, none of these possible extensions has been considered in detail in this dissertation. Furthermore, although we have argued that simulations are a reasonable approach to establish empirical bounds on the required random walk length, the range of network sizes and topologies considered so far is fairly limited. A study of the protocol's performance in further network topologies must thus be considered future work.

Controlling and Initiating Rewirings

The reader may further observe that a rather grossly simplified abstraction of practical computing systems has been used. We thus need to comment on some issues that necessarily arise in practical systems. First, when wanting to create a power law topology with a particular exponent, all nodes must be aware of this value and forward random walk rewiring requests accordingly. So far we have not considered how the formation or adaptation of the overlay topology is initiated and how a value for the desired exponent can be found consistently by all participating nodes. Apart from using a centralized control instance, for this purpose the use of (probabilistic) distributed consensus protocols may be considered. However this issue is not addressed in this dissertation. While some preliminary results on network topologies with dynamic participants have been presented, a major open issue of this work is the fact that simulations have only been performed for the rather special situation of dynamic equilibrium states with nodes failing uniformly at random. An evaluation of the impact of further realistic (non-uniform) churn models and rates is thus crucial. Considering an application in practical systems, a further potential problem that could not be considered so far is the impact of message losses and how to properly handle them in the protocol. However, in [Zhong *et al.*, 2008], some ideas on handling message losses in random walk sampling have

Application in Systems with Dynamic Constituents

been presented which may also be applicable in the context of the rewiring protocol considered in this section.

Respecting Peer Capacities

Finally, in the rather simple scheme described in this section, we have considered nodes being assigned (random) identifiers that influence rewiring operations and hence their connectivity in the resulting overlay topology. As a result, the number of links that any given node will eventually acquire is determined by mere chance. However, in most practical systems, capacities of participating machines (like for instance communication bandwidth or processing power) are non-homogeneous and thus should be taken into account in the construction of the overlay. Several approaches may be considered to resolve this issue. While the simplest possibility would be an assignment of IDs to nodes in which highest capacity nodes are assigned smallest IDs, such a strategy may not be feasible in large dynamic systems. One possible approach to introduce a correlation between connectivity and node capacities may be to combine a initial random assignment with a distributed probabilistic mechanism by which nodes are allowed to reasonably swap their IDs according to capacity differences. When considering systems with a sufficiently skewed distribution of capacities, another adjustment of the protocol may be considered based on the models that have been presented in [Caldarelli *et al.*, 2002b; Servedio and Caldarelli, 2004]. Here an equilibrium model with hidden vertex intrinsic variables has been studied which is similar to the simple one used in this section. Apart from a trivial assignment of Zipf-distributed weights, the emergence of power law and scale-free networks has been observed for a number of further sufficiently skewed distribution of vertex variables. Based on the observation of highly skewed node characteristics in practical Peer-to-Peer systems (see for example [Stutzbach and Rejaie, 2006]), this proposes a simple extension of the sampling scheme underlying the rewiring protocol studied in this section. For this, one can use as weights w_i in equation 4.1 a simple numerical value that resembles a node's local characteristic like for example its bandwidth or expected uptime. The transition matrix must then be adjusted accordingly, replacing $\left(\frac{i}{j}\right)^{\frac{1}{\gamma-1}}$ in equation 4.5 by $\frac{w_j}{w_i}$. The overlay resulting from the proposed rewiring and link sampling scheme will then be a power law topology whose exponent depends on the initial distribution of capacities and in which nodes with the highest capacities have the most connections. In particular, this means that one loses the ability to effectuate an overlay with a particular degree distribution exponent. Nevertheless, an adaptation may still be possible when considering particular capacity distributions and a suitable transformation of the local capacity value in the assignment of local weights.

Further Perspectives

Clearly, the creation and adaptation of random scale-free structures is merely one possible scenario in the thermodynamic management of large, dynamic overlays. In fact, rather than being limited to the construction of scale-free structures, a combination of the proposed rewiring and random walk sampling scheme with the equilibrium statistical mechanics and complex networks perspective on overlays opens up a number of further perspectives. A simple extension would be to use the rewiring mechanism not only to switch between different statistical ensembles for scale-free networks but between completely different statistical ensembles. A transformation between random Erdös/Rényi overlays and scale-free topologies can for example simply be achieved by switching between highly skewed and uniform statistical weights in the sampling process. Furthermore, apart from these rather simple networks, more complex overlay topologies with particular clustering or community substructures may be desirable. For this, it appears promising to consider further, possibly non purely local, energy concepts like for instance those summarized in [Farkas *et al.*, 2004]. In the light of these limitations and open issues, the scheme proposed in this section may thus be seen as a mere demonstration of the *thermodynamic perspective* on the management of large, dynamic overlay topologies whose self-organization and self-adaptation capabilities can be analyzed exactly in the framework of complex networks and random graph theory.

4.3 Monitoring Scale-Free Overlays

Distributed Monitoring of Aggregate Quantities

The rewiring protocol that has been considered in the previous section constitutes one possible building block for an *equilibrium statistical mechanics* approach to the management of large, dynamic overlay networks. In chapter 3 we have argued that a meaningful, randomized construction of overlays with particular statistical parameters can facilitate macroscopic statements about a number of aggregate qualities which are - although rather basic - nevertheless important for the design of reliable and scalable distributed systems. In this section we consider a different yet related question. Assuming a given network topology, we study how aggregate quantities that facilitate such a macro-level reasoning in terms of statistical ensembles can be derived efficiently and in a distributed fashion. Constituting a complementary aspect for the distributed management of network topologies with scale-free characteristics, we particularly address the question how individual nodes can infer knowledge about power law nature and exponent of a network's degree distribution by means of a probabilistic and fully decentralized monitoring protocol.

Monitoring the Degree Distribution

Clearly, when wanting to use such a distributed monitoring scheme to make statements about networked systems resulting from sophisticated construction procedures, all the cautionary remarks that have been put forth in section 4.1 need to be taken into account. Nevertheless, one can imagine a number of scenarios where measurable aggregate quantities can be be viewed as parameters of a statistical ensemble that represents the actual network construction process reasonably well. One example are networked systems whose evolution is known to follow well-studied stochastic models like growth and preferential attachment schemes, link duplication or random walk mechanisms (see for instance the models mentioned in section 2.3.5). Still details of these models (like for instance the preferences of added participants) which determine the exact shape of the connectivity distribution and influence network qualities relevant to networked computing systems and distributed algorithms may not be easily assessable. Another example can be found in the extension to the rewiring protocol that has briefly been mentioned in section 4.2.4. Based on the equilibrium model for scale-free networks that has been discussed in [Caldarelli *et al.*, 2002b], node capacity depending fitness values can be used as statistical weights in the rewiring protocol. Different from the simpler scheme considered in section 4.2, here the degree distribution of the resulting overlay network cannot easily be derived from a simple tunable protocol parameter. The exact shape of the distribution rather depends on possibly time-variant

characteristics of participating machines. Finally, even in global scale network infrastructures like the Internet, a measurement of the degree distribution can be meaningful if suitably combined with measurements of further characteristics which are adequate to assess basic network qualities. On the other hand, commonly used approaches like the (near-)complete mapping of large network topologies or the sampling of a small subset of nodes can be problematic. Moreover, due to the statistical properties of highly skewed distributions, one needs to take care with respect to the used fitting procedures. In the following, we address the question how nodes in power law networks can reason about the shape of the degree distribution in a fully decentralized and statistically sound way.

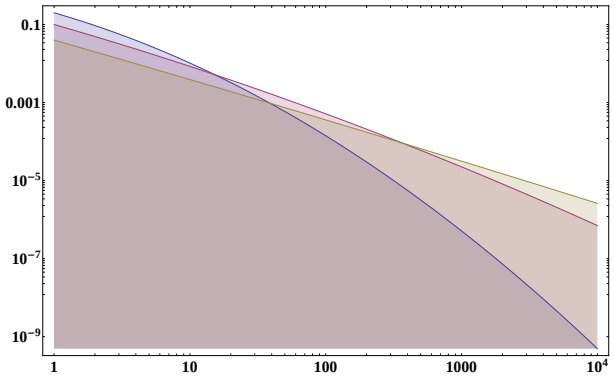
4.3.1 Fitting to Power Law Distributions

Fitting to Skewed Distributions

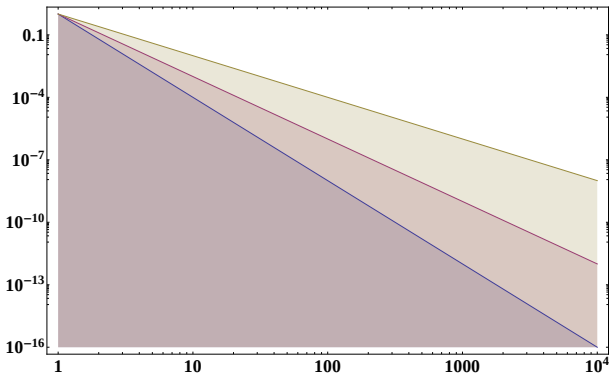
A first step towards a distributed measuring of the degree distribution shape in power law networks is to find an *objective and automatable procedure to fit a power law function to a set of empirical data*. For highly skewed distributions like power laws, the provision of reliable fitting procedures is an active area of research in statistics. A frequent source of errors when using graphical or least squares fitting procedures is that log-normal distributed data can easily be mistaken for power laws when plotting them on a double logarithmic axis. This can be seen in Figure 4.12, showing double logarithmic plots of three log-normal cumulative distribution functions for three different standard deviations as opposed to three power law functions with different exponents. This is aggravated by the fact that, many power law distributions in real-world systems exhibit exponential cutoffs due to finite-size effects, thus making power law and log-normal distributed data hardly distinguishable visually.

Maximum Likelihood Estimation of Exponents

Notwithstanding the fact that a trained eye and the proper use of sample binning and plotting can often identify power laws and rule out alternative distributions, here we are interested in an objective and automatable approach that is suitable to be used in a distributed algorithm. Currently, one of the most reliable fitting methods for power law distributed data is based on the well-known Maximum Likelihood Estimation (MLE) technique and has been presented [Clauset *et al.*, 2007; Bauke, 2007]. Given a sample of data to which a power law function shall be fitted, it can be used to obtain the most likely value for the exponent γ as well as a measure for the goodness of this estimation. Referring to [Clauset *et al.*, 2007] for a full motivation and description of the procedure, here we briefly summarize those aspects of the approach that are



(a) Three log-normal distributions with $\mu = 0$ and $\sigma = 2$, 4 and 10



(b) Three power law distributions with different exponents

Figure 4.12: Comparison of log-normal and power law CDFs in double logarithmic plots

essential for the application in a decentralized fitting scheme. The application of the MLE method to networks with power law degree distributions results in the equation [Clauset *et al.*, 2007]

$$\frac{\partial_\gamma \xi(\gamma, d_{min})}{\xi(\gamma, d_{min})} = -\frac{1}{|V|} \sum_{v \in V} ln(d_v) \tag{4.10}$$

A Simplified Estimation

where $\xi(\gamma, a)$ is the Hurwitz-Zeta function, ∂_γ is the partial derivative with respect to γ and d_v denotes the degree of a vertex v in the network. For a known lower bound d_{min} above which the power law holds, a numerical solution of this equation for γ then yields the maximally likely degree distribution exponent. For the scenarios we address, a numerical solution of this equation can be too costly especially when considering light-weight network nodes

with restricted computational power. Fortunately, the authors of [Clauset *et al.*, 2007] give a much simpler yet less accurate estimate for the exponent γ which does not involve the Hurwitz-Zeta function. For the degree distribution of a network $G = (V, E)$ it is given by

$$\gamma = 1 + |V| \cdot \left[\sum_{v \in V} ln \left(\frac{d_v}{d_{min} - \frac{1}{2}} \right) \right]^{-1}$$

which can be reformulated by application of basic log rules as follows:

$$\gamma = 1 + \frac{|V|}{\sum_{v \in V} ln(d_v) - |V| \cdot ln \left(d_{min} - \frac{1}{2} \right)} \tag{4.11}$$

Estimating the Minimum Degree

In [Clauset *et al.*, 2007] it has been argued that for $d_{min} > 6$ the accuracy loss of this simpler estimation of the exponent is less than 1 %. All these equations contain a parameter d_{min} which represents the minimum degree above which the degree distribution exhibits power law behavior. Most real networks exhibit such a lower bound (for example due to minimum connectivity constraints) and estimating it is often non-trivial. Underestimating d_{min} will result in a large number of non power law samples being included in the fit while an overestimation can result in a significant fraction of valid data not being considered in the fit. Both significantly reduces the fit reliability. In [Clauset *et al.*, 2007], a simple method to estimate the minimum degree above which the power law fit holds is given based on the Kolmogorov-Smirnov distance. For a random variable taking discrete, finite values - the Kolmogorov-Smirnov distance D between two cumulative distribution functions F and F' can be defined as follows:

The Kolmogorov-Smirnov Distance

$$D := max_x \left| F'(x) - F(x) \right|$$

If we assume that F_y denotes the cumulative distribution function (CDF) of a fitted power law with an estimated $d_{min} = y$, then the Kolmogorov-Smirnov distance between the observed and the fitted CDF F_y can be given as follows:

$$D(y) := max_{x \geq y} \left| \frac{|\{v \in V : d_v \geq x\}|}{|\{v \in V : d_v \geq y\}|} - F_y(X \geq x) \right| \tag{4.12}$$

The optimum value of d_{min} can then be obtained by minimizing the function $D(y)$, thus yielding the value d_{min} for which the Kolmogorov-Smirnov distance between the empirical and the fitted cumulative distribution function

is minimal. In summary, these equations can be used to derive the parameters γ and d_{min} for which a power law distribution has maximum likelihood to produce the observed node degrees d_v.

4.3.2 Gossip-based Fitting

Distributed Maximum Likelihood Estimation

In their current form, an application of the fitting method summarized above requires a histogram of the degree of all nodes in the network. Clearly collecting such a histogram is not desirable in large scale networked computing systems. Here, we thus study how one can, in a distributed and probabilistic fashion, progressively calculate estimates for the exponent γ, the minimum degree d_{min} above which an assumed power law holds and the resulting Kolmogorov-Smirnov distance D. For this, we recall that in equation 4.11 merely aggregate values are required, namely the number of nodes $|V|$ in the system as well as a sum of logarithmized node degrees. Hence, obtaining these aggregate information in a scalable and distributed fashion is sufficient to perform a Maximum Likelihood Estimation of the degree distribution exponent.

Gossip-based Aggregation

One particularly appealing possibility to obtain such aggregate quantities in large dynamic network topologies is by means of the probabilistic, gossip-based aggregation scheme that has been presented in [Jelasity *et al.*, 2005]. Here, it has been shown that aggregate functions *min*, *max*, *sum*, *product*, *average* as well as higher moments of per-node numerical attributes can be calculated efficiently by means of a simple probabilistic protocol. While referring the reader to [Jelasity *et al.*, 2005] for a more detailed presentation, for this nodes need to periodically exchange a local approximation of aggregate values with a neighbor chosen uniformly at random. These local approximations are then updated according to a function which depends on the aggregate function that shall be computed. With a suitably chosen update function, local approximations will converge towards the actual global aggregates and the variance of the set of approximate values will decrease upon each update. However, just like for the random walk sampling scheme that has been considered in section 4.2, the interesting question here is how fast this convergence is in a particular network topology. In [Jelasity *et al.*, 2005], both analytical and empirical arguments for the convergence behavior of the protocol have been presented. For random $G(n, m)$ and Barabási/Albert networks, the convergence has been found to be close to the optimum of a random graphs, requiring no more than 20 per-node information exchanges to achieve a variance $< 10^{-9}$ in network topologies with one million nodes. These results are consistent with the findings about diameter, average path lengths, expansion and random walk con-

Convergence of Gossip-based Aggregation

Algorithm 8: Main Loop of Monitoring Protocol

1: $S \leftarrow ln(d_{self})$
2: $M \leftarrow \{d_{self}, \infty, \ldots, \infty\}$
3: $C \leftarrow \{1, 0, \ldots, 0\}$
4: **loop**
5: $Sleep(delay)$
6: $r \leftarrow RandomNeighbor()$
7: {Send local vector}
8: $rqMsg.S \leftarrow S$
9: $rqMsg.M \leftarrow M$
10: $rqMsg.C \leftarrow C$
11: $Send(\{RQ, rqMsg\}, r)$
12: **end loop**

vergence that have been mentioned in sections 2.1, 2.3.3 and 4.2.1. In particular, this suggests that the gossip-based aggregation scheme can be applied in probabilistically structured overlay topologies with sufficient algebraic connectivity[10].

Gossip-based Maximum Likelihood Estimation

We now show how the gossip-based aggregation scheme can be applied to perform a distributed Maximum Likelihood estimation of the degree distribution exponent γ and the minimum degree d_min above which the fit holds. While an algorithmic description can be found in algorithms 8 - 11, we first comment on some preliminaries. First, we assume that each node v in the network initializes a local value S with the logarithm of its own degree. Furthermore, two vectors M (list of minimum degrees) and C (number of nodes with minimum degrees) of some fixed length k need to be initialized with the values ∞ and 0 respectively. The first entries of the vectors M and C are set to the own degree d_v and 1 respectively. We will later see how these vectors M and C can be used to obtain an estimation of the minimum degree d_{min} above which a power law holds with maximum likelihood.

Protocol Overview

In algorithm 8, the initialization and the main loop of the monitoring scheme is shown. Following the original epidemic aggregation scheme presented in more detail in [Jelasity *et al.*, 2005], here we assume that in periodic intervals (determined by a delay time during which nodes are assumed to sleep) nodes chose a neighbor in the network or overlay topology uniformly at ran-

[10] See for example [Boyd *et al.*, 2006] for a further investigation of the influence of spectral properties on the convergence of gossip-based averaging algorithms

Algorithm 9: Node receives $\{RQ, rqMsg\}$ from node x

1: $rspMsg.S \leftarrow S$
2: $rspMsg.M \leftarrow M$
3: $rspMsg.C \leftarrow C$
4: $Send(\{RSP, rspMsg\}, x)$
5: $Update(rqMsg)$

Algorithm 10: Node receives $\{RSP, rspMsg\}$

1: $Update(rspMsg)$

dom. A request message which contains the local values of the node is then sent to the chosen neighbor. As shown in algorithm 9, a node receiving such a request responds with its local values and updates them according to the received values and an update function. Similarly, when a response is received from a node the values of the responding node are used to update the local values (see algorithm 10).

Update Rules

When values of a neighbor are received, updates of the current local approximations are performed according to algorithm 11. Here, S is simply updated by averaging the local and the received value. In the vector M, a set of the k smallest unique node degrees is collected. In an entry $C[j]$ of vector C, values of a per-node indicator function

$$\delta_v(i) := \begin{cases} 1 & \text{if } d_v \geq M[i] \\ 0 & \text{else} \end{cases} \tag{4.13}$$

are averaged. When a vector M with a list of k smallest degrees is received from a neighbor, it is merged with the local vector, meaning that a union of both vectors is created while keeping only the k smallest unique degrees. Whenever the i-th entry of a node's local vector C is changed, the value $C[i]$ is reinitialized with the value of the indicator function shown in equation 4.13. If this procedure is applied repeatedly, due to the averaging update rule in line one of algorithm 11, a node i's local value S will converge to:

$$S \to \frac{1}{|V|} \sum_{v \in V} ln(d_v)$$

Algorithm 11: Update of local values S, M and C after receiving message msg

$S \leftarrow \frac{1}{2} \cdot (S + msg.S)$
$M' \leftarrow Merge(M, msg.M)$
for $i = 0$ to k **do**
  if $M[i] \neq M'[i]$ **then**
    $C[i] \leftarrow \delta_{self}(i)$
  else
    $C[i] \leftarrow \frac{1}{2}(C[i] + msg.C[i])$
  end if
  $i \leftarrow i + 1$
end for
$M \leftarrow M'$

Distributed Estimation of γ

Using the local approximation S and equation 4.10, a node v can then obtain a local Maximum Likelihood Estimation of the degree distribution exponent γ by solving the following equation:

$$\frac{\partial_\gamma \xi(\gamma, d_{min})}{\xi(\gamma, d_{min})} = -S \tag{4.14}$$

In cases where a numerical solution or the computation of the Hurwitz-Zeta function is too complex, a simple yet less accurate estimate can alternatively be derived from equation 4.11 as follows:

$$\gamma = 1 + \left(S - ln\left(d_{min} - \frac{1}{2} \right) \right)^{-1} \tag{4.15}$$

Bounding the possible range of d_{min}

In both equations, the lower bound d_{min} for which the power law fit holds is required. An estimation for this value can be obtained based on the above aggregation scheme and the local vectors M and C. We first observe, that the Kolmogorov-Smirnov distance from equation 4.12 cannot be directly calculated based on a node's local view since it requires a histogram of vertex degree frequencies. We can however presume that the searched degree d_{min} is among a fixed size subset of the network's smallest vertex degrees. In fact, power law distributions are well-suited for this restriction. Due to their positive skewness, a large fraction of the distribution's dynamic range is concentrated at the lower end of the degree distribution. Thus, small values for d_{min} are likely. In fact high values for d_{min} result in a large portion of data not being

considered for the fit, thus making the fit less reliable and actually uninteresting. This is the rationale behind using the size constrained vectors M and C. In M, the k smallest unique degrees in the network are collected by means of a simple minimum aggregation. In each entry $C[j]$, a per-node indicator function telling whether the local degree is greater or equal than the corresponding entry $M[j]$ is averaged. Thus, once an entry $M[j]$ has stabilized, the entries $C[j]$ will converge to:

$$C[j] \rightarrow \frac{1}{|V|} \cdot |\{v \in V : d_v \geq M[j]\}|$$

Distributed Estimation of d_{min}

We may further assume, that each node has a local estimation of the network size $n \approx |V|$. Various methods exist by which such an estimation can be obtained in a distributed fashion [Horowitz and Malkhi, 2003; Kostoulas *et al.*, 2005; Massoulié *et al.*, 2006; Merrer *et al.*, 2006], among them an application of gossip-based averaging [Jelasity *et al.*, 2005]. Multiplying, after a number of gossip iterations, the estimated network size n with an entry $C[j]$ yields:

$$n \cdot C[j] \approx |\{v \in V : d_v \geq M[j]\}|$$

With this, equation 4.12 can be applied to those vertex degrees stored in the local vectors M. Each of the nodes participating in the scheme can then estimate the lower bound d_{min} by selecting the value $y \in M$ that minimizes the following equation:

$$\bar{D}(y) = max_{x \in M : x \geq y} \left| \frac{n \cdot C[i_x]}{n \cdot C[i_y]} - P_y(X \geq x) \right| \tag{4.16}$$

where i_x and i_y are the indices of x and y respectively in M and F_y is the fitted CDF obtained from equation 4.14 or equation 4.15 when setting $d_{min} = y$.

Local Estimates	$i = 5$	$i = 10$	$i = 50$
γ-Avg	2.839	2.833	2.833
γ-Var	$1.84 \cdot 10^{-2}$	$2.22 \cdot 10^{-4}$	$4.72 \cdot 10^{-17}$
D-Avg	0.05	0.05	0.05
D-Var	$4.1 \cdot 10^{-4}$	$7.31 \cdot 10^{-6}$	$1.56 \cdot 10^{-18}$
d_{min}-Avg	5.010	5	5
d_{min}-Var	$6.850 \cdot 10^{-2}$	0	0

Figure 4.13: Average and variance of local power law parameter estimations for i iterations of the proposed gossip scheme using a vector size of $k = 10$ entries in a 10000 node Barabási/Albert network. A global Maximum Likelihood Estimation yields $\gamma = 2.833$, $d_{min} = 5$ and $D = 0.0497$.

4.3.3 Experimental Evaluation

Experimental Results

We finally evaluate the performance of the scheme by comparing the fit parameters obtained by a global Maximum Likelihood Estimation with the local values obtained in a simulation of the proposed protocol. In this simulation, we have created a power law network with 10000 nodes and roughly 50000 edges according to the Barabási/Albert preferential attachment model. For the local vectors M and C, a size constraint of $k = 10$ was used. A global MLE power law fit of the degree distribution obtained by an R implementation[11] of the method discussed in [Clauset *et al.*, 2007] yields $\gamma = 2.833$, $d_{min} = 5$ and $D = 0.0497$. For both the global and the distributed estimation of power law fit parameters, the simplified fitting methods given in equations 4.11 and 4.15 have been used. In Figure 4.13, average and variance of the node's local estimates of the degree distribution exponent γ, the minimum degree d_{min} and the Kolmogorov-Smirnov distance D are shown. The results show, that the average local estimations quickly converge towards the actual values computed globally. After only five iterations of the gossip-based protocol, the average estimated value of nodes differs by less than 0.25 % from the actual value, after ten iterations this difference falls below 0.02 %. Similar results hold for the local estimations of D and the lower bound d_{min}.

[11] In particular, the R implementation by Laurent Dubroca available at `http://tuvalu.santafe.edu/~aaronc/powerlaws/` has been used.

4.3.4 Summary and Perspectives

Summary and Contributions

In this section, we have shown that it is possible to obtain a statistically sound Maximum Likelihood Estimation of power law fit parameters in a distributed and probabilistic fashion. The proposed monitoring scheme is a direct application of recent efforts in providing reliable and automatable power law fitting methods and makes use of a gossip-based aggregation scheme which is known to work efficiently in large dynamic network topologies with good algebraic connectivity. Based on the findings summarized in section 2.3 and the arguments put forth at beginning of this section 4.3, the resulting knowledge about a network's degree distribution exponent can - at least in a number of cases - be used to make statements about network properties like diameter, average path lengths, fault and attack resilience or the performance of dynamical processes. In slowly evolving, very large network infrastructures like the Internet's router network, a measurement of such statistical properties can further be useful for the validation of network models. While a complete mapping of such networks is typically costly and laborious, the proposed protocol could be used in periods of low traffic to exchange and aggregate statistical information in a simple and distributed fashion. Referring to a vision put forth in the context of the distributed mapping project DIMES [Shavitt and Shir, 2005], at least for simple aggregate parameters, the proposed protocol can be seen as a step towards networked systems that "measure themselves".

Open Issues

We conclude this section by summarizing some open issues that have not yet been considered in more detail. First and foremost, in its current form *the proposed monitoring scheme can only be used to estimate parameters of degree distributions that are known to obey a power law*. However an important question that frequently arises in practice is whether or not a distribution obeys a power law in the first place. Unfortunately the value D of the Kolmogorov-Smirnov distance cannot be used directly to evaluate the absolute goodness of a power law fit since another probability distribution may still produce smaller distances[12]. To address this issue, in [Clauset *et al.*, 2007] a method is presented which yields a single scalar value telling whether the deviation of the observed data from the best possible power law fit is likely to result from statistical fluctuations. This method involves the creation of a large number of synthetic data sets according to the fitted power law distribution. For these synthetic data sets, the Kolmogorov-Smirnov distance is then computed and compared with that obtained for the observed data. A similar method may be applicable when using

[12] Please note that a monitoring of the *relative change* of the Kolmogorov-Smirnov distance over time still allows to assess whether the goodness of a fit becomes stronger or weaker.

several subsequent rounds of the distributed monitoring scheme. While the first of these rounds works as described above, in further rounds nodes may initialize their local values according to a synthetic degree d being randomly drawn according to the statistical parameters obtained in the first round. While a comparison of the resulting Kolmogorov-Smirnov distance for synthetically generated and actual node degrees may give clues about the goodness of the power law fit, a further study of the feasibility of such an extension remains an open issue. Another aspect that has been neglected so far is the impact of node dynamics on the quality of the fitting procedure. In [Jelasity *et al.*, 2005], the effect of failing nodes and links on the aggregation performance of the gossiping scheme underlying the protocol proposed in this section has been considered and it has been found to be rather tolerant against network dynamics while still providing reasonably good estimates. However, the impact of dynamic nodes and the resulting node degree fluctuations on the local estimation of distribution parameters has not yet been studied. We have further not yet studied the relation between network size and convergence behavior.

Further Perspectives

Clearly, the mere measurement of the network's degree distribution is interesting only in rather special cases. An important related question is therefore whether other performance-relevant, aggregate network metrics like for instance average clustering coefficient, assortativity [Newman, 2002] or degree-degree correlations can be derived in a similar fashion. In the management of probabilistically structured or unstructured overlay networks, the protocol considered here should thus rather be viewed as a particular and admittedly simplistic example for a *distributed, runtime measurement of aggregate statistical quantities*. Reconsidering the perspective that has been taken in section 3.2.2, the ability to efficiently derive such aggregate quantities and the use of results about corresponding statistical ensembles facilitates conclusions about the collective properties of large scale overlay networks. The resulting knowledge can then be used for example in the development of distributed algorithms which better adapt to the qualities of dynamically evolving network topologies. In a sense, probabilistic aggregation protocols like the one considered in this section may thus be viewed as the thermometers and barometers of thermodynamically structured overlay networks.

Chapter 5

Self-Organized Synchronization in Networked Computing Systems

> When I meet God, I am going to ask him two questions: Why relativity? And why turbulence? I really believe he will have an answer for the first.
>
> WERNER HEISENBERG

Complex Collective Phenomena

Having studied the formation of overlay networks with complex probabilistic structures, here we slightly shift our focus and examine potential applications of what is commonly called *collective phenomena*. Depending on the context, *we consider as collective phenomenon, collective dynamics or collective behavior, the formation of macroscopically coherent temporal or spatial patterns based on the interactions of a large number of dynamical elements*. Prominent examples being the seemingly coordinated, collective motion of flocks, shoals and swarms of animals, bacterial colonies and human crowds as well as the synchronous flashing of huge populations of fireflies, such collective phenomena are abundant in nature. Similar collective dynamics can also be observed in the inanimate world. Here, frequently studied occurrences are the formation of domains with aligned electron spins in ferromagnetic materials or the emergence of convection patterns in fluid dynamics. In the context of social systems, the formation of consensus, the dissemination of languages and cultural traits as well as the

dynamics of opinion formation processes represent further instances of collective behavior.

Understanding Collective Behavior in Networked Systems

The emergence of collective behavior - be it in a social, biological or physical context - is one particular facet of the complex dynamics which can occur in high-dimensional nonlinear dynamical systems. A rigorous analytical treatment of such systems is possible only under very special conditions. In recent years, methods and models originally developed in statistical physics to study the collective dynamics for example of spin models have proven to be useful also to tackle phenomena like collective motion in biology [Vicsek *et al.*, 1999] or collective behavior in society [Castellano *et al.*, 2009]. While in traditional physical models interactions are typically assumed to take place in regular lattices or fields, complex networks are now increasingly being considered as interaction topologies, thus furthering an application of the corresponding results in other domains. This development opens up interesting perspectives for an application in networked computing systems. In fact, some of the examples that have been reviewed in section 1.1 highlight the relevance of the study of collective phenomena in the context of technical systems. The inadvertent synchronization of router messages in the Internet shows that a thorough understanding of the principles underlying collective dynamics is crucial for the design of networked systems. Moreover, the example of the Millennium footbridge has shown that subtle nonlinear interactions, which can occur between pedestrians and a construction, may result in significant problems. Similarly, mutual interactions between networked computing systems and their human users can possibly give rise to collective user behavior that significantly impacts performance. With technical and social systems becoming increasingly intertwined, *the study of collective phenomena is likely to gain importance for the design of robust and manageable technical infrastructures.*

Harnessing Collective Phenomena in Distributed Computing Systems

Reconsidering collective phenomena in natural systems, their emergence usually entails some form of "benefit" for the participating elements. Collective motion in animal populations can, for example, reduce the probability of an individual falling prey to predators or optimize the energy consumption of long-distance migrations. Although not representing a "benefit" in the usual sense, in spin models for ferromagnetism collective dynamics arises as a result of a minimization of free energy. Similarly, *apart from predicting and preventing unwanted collective phenomena*, in decentralized computing systems required to operate without a central control instance, *the formation of coherent and ordered macroscopic patterns can be desirable*. Here, explicitly designing distributed protocols along suitable physical and mathematical models giving rise to benefi-

cial collective behavior is a valid approach. In the remainder of this chapter we try to implement it by considering a distributed scheme for the synchronization of periodic processes in Peer-to-Peer systems. We particularly address scenarios in which complex, probabilistic overlay topologies like for example those studied in the preceding chapter are being used. In the following section we first review some models and results about *self-synchronization phenomena in networks*. Since it will serve as the basis for the gossip-based synchronization protocol studied in section 5.2, here we particularly introduce the Kuramoto model for populations of nonlinearly coupled oscillators.

5.1 Self-Organized Synchronization in Networks

Models for Self-Organized Synchronization

An interesting type of collective dynamics frequently observed and studied in biology, chemistry, mathematics and non-equilibrium statistical physics can arise in systems being comprised of periodic processes. By means of interactions between these processes, synchronized, coherent patterns can emerge in a seemingly self-organized fashion, possibly resulting in the formation of globally coherent states or complex spatio-temporal patterns. In some respects resembling previously mentioned instances of distributed consensus, flocking and swarm behavior, self-synchronization phenomena have been studied extensively especially in the field of biology where they appear to be almost pervasive. Its occurrences range from the temporal behavior of neuronal activation patterns over the synchronized contraction of cardiac muscle cells, the circadian cycle in humans to the synchronized flashing of fireflies. Since roughly 50 years, a number of different models have been studied that can explain such examples of collective behavior. A beautiful introduction to synchronization phenomena as well as to some mechanisms underlying their emergence is given in the popular science book "Sync" [Strogatz, 2003]. While exhaustive surveys can be found for example in [Strogatz, 2000; Pikovsky *et al.*, 2003; Boccaletti *et al.*, 2002; Boccaletti *et al.*, 2006; Barrat *et al.*, 2008], in this section we summarize some aspects of synchronization models in networks insofar as they relate to the questions studied in the remainder of this chapter.

Models with Pulse Couplings

Resembling interactions taking place between the elements in a number of biological systems, a frequently studied class of synchronization models are so-called *Integrate-and-Fire models* which assume a sporadic, pulse-based coupling between dynamical elements. In each of these elements a state variable - which in practice can be interpreted for example as tension, pressure or voltage - builds up at some intrinsic phase advance rate until a threshold value

is exceeded, thus resulting in a sudden relaxation. Depending on the exact model being studied, this relaxation is assumed to generate a pulse which influences the advance of the state variables of neighboring elements. To explain the synchrony in cardiac pacemaker cells, in [Peskin, 1975] an Integrate-and-Fire model has been proposed in which a pulse advances the state variables of all other nodes by a constant amount. This model has later been generalized in [Mirollo and Strogatz, 1990], showing that any monotonic and concave down intrinsic advance of state variables eventually results in globally synchronized states except for a set of pathologic initial conditions that have a probability measure of zero.

Coupled Oscillator Models

Different from the sporadic, pulse-like coupling considered in Integrate-and-Fire models, in *coupled oscillator models* a continuous coupling between oscillatory dynamical elements is usually assumed. One of the earliest models studies a population of oscillators continuously "pulling" on each other's intrinsic frequencies [Wiener, 1958]. Considering oscillator interactions as phase-dependent influence and sensitivity functions, the generalized relaxation oscillator model introduced in [Winfree, 1967] has opened the problem to formal analysis. Building on this work, one of the most frequently studied coupled oscillator models has been introduced in [Kuramoto, 1975]. Since it constitutes the basis for the distributed synchronization scheme proposed in section 5.2, in the following we consider this *Kuramoto model* in a bit more detail and summarize some results that are relevant to the aspects addressed in the remainder of this chapter[1].

The Kuramoto Model

In its most common form, the Kuramoto model assumes n all-to-all coupled oscillators running at intrinsic frequencies ω_i which are distributed according to some arbitrary probability distribution. A constant K giving the strength of sinusoidal couplings between oscillators, the advance of an oscillator i's phase Θ_i is given in terms of the following differential equation:

$$\frac{\partial \Theta_i}{\partial t} = \omega_i + \frac{K}{N} \cdot \sum_{j=1}^{N} sin(\Theta_j - \Theta_i) \tag{5.1}$$

Hence, for $K = 0$ a node i's phase Θ_i advances with constant speed at its intrinsic frequency ω_i. Assuming $\Theta_k \in [0, 2\pi)$ and interpreting Θ_k as a node k's position on a unit circle, for $K > 0$ the advance of an oscillator's phase at any given time t is influenced by the angles between its local phase and the phases of all other nodes. Positive angles speed up, negative values slow down an

[1] More detailed surveys of analytical and experimental results on the Kuramoto model along with several dependent variations can be found for example in [Strogatz, 2000; Acebrón *et al.*, 2005]

Mean-Field Results for Kuramoto Model

oscillator's phase advance. From equation 5.1, one sees that the mathematical description of a population of n coupled Kuramoto oscillators involves a nonlinear system of n differential equations which complicates an analytical evaluation of the system's behavior. Despite this fact, in [Kuramoto, 1975] analytical results have been presented for the case $n \to \infty$. Here it was shown that for certain, sufficiently narrow distributions of intrinsic oscillator frequencies, increasingly large groups of oscillators eventually synchronize as the coupling strength K is increased above a critical coupling strength K_c that depends on the distribution of intrinsic frequencies ω_i. These analytical results have been obtained by a mean-field approach, which is facilitated by the assumption of an *all-to-all coupling* and the fact that in this case each oscillator can be viewed as being effectively influenced by the average couplings to all other oscillators.

Synchronization in Random and Complex Networks

While an all-to-all coupling possibly reduces the dimensionality of the problem in an analytical treatment, for many natural systems it is more reasonable to study *sparse networks as interaction topologies*, oscillators residing at the nodes of a network and edges representing couplings between them. The topology of the network in which couplings take place clearly influences the emergence of synchronized states as well as the dynamics of the synchronization process. For regular chain, ring and lattice topologies, it has been shown that the emergence of stable synchronized states in populations of sinusoidally coupled oscillators becomes increasingly unlikely as the size of the network increases [Ermentrout and Kopell, 1984; Ermentrout, 1985; Strogatz and Mirollo, 1988]. Contrariwise, stable synchronized states apparently emerge when using a random Erdös/Rényi as coupling topology [Satoh, 1989]. Apart from regular and completely random networks, in recent years complex coupling topologies are increasingly being considered. In [Watts, 1999], the dynamics of coupled Kuramoto oscillators in networks resulting from the random rewiring of links in an initially regular ring lattice has been studied. It was found that the random rewiring of a small fraction of links is sufficient for synchrony to emerge almost as quickly as in fully connected or completely random networks. For random networks with a given degree distribution, in [Ichinomiya, 2004] it is argued that the critical coupling required for synchrony to emerge in a variation of the Kuramoto model is proportional to the ratio $\frac{M_1}{M_2}$ of the degree distribution's first two moments. Referring to arguments presented in 2.3.1, this suggests that scale-free networks with a degree distribution in the range between two and three tend to synchronize for very

Synchronization in Scale-Free Networks

small coupling strengths, the critical coupling strength required for synchrony to emerge vanishing as the size of the network increases[2].

Spectral Network Properties and Synchronization

Reconsidering the relation between network structures and the emergence of synchronized regimes, it seems rather intuitive that the synchronization dynamics is influenced by how "well-connected" a network is and, in particular, by the network's average path lengths and the existence of small-cut node partitions. As such, the question if and how synchronization occurs in a particular network topology is closely related to the performance of dynamical processes like information diffusion, the spreading of epidemics and the convergence behavior of gossiping strategies and random walks. This immediately reminds us of chapter 4, where we have used the "good connectedness" of random scale-free networks to efficiently sample random edges and to probabilistically derive the degree distribution exponent. Moreover, in section 4.2.1 we have mentioned that a theoretical upper bound for the convergence time of a random walk can be given in terms of the second largest eigenvalue of the random walk's stochastic transition matrix. Similarly, the relation between the emergence of stable synchronized states and the spectral properties of its *Laplacian matrix* have been studied. The Laplacian matrix is tightly coupled to a network's adjacency matrix and can be seen as a discrete and finite variation of the continuous Laplacian operator[3]. For a network $G = (V, E)$ with $V = \{1, \ldots, n\}$, an entry $\left(L_{ij}\right)_{i,j=1,\ldots,n}$ of the Laplacian matrix $L(G)$ is defined as follows:

$$L_{ij} := \begin{cases} d_i & i = j \\ -1 & i \neq j \text{ and } (i,j) \in E \\ 0 & \text{else} \end{cases}$$

The spectrum of eigenvalues $0 = \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{max}$ of the Laplacian matrix $L(G)$ captures important properties of the underlying network topology and plays an important role in the stability analysis of oscillatory patterns for coupled oscillator models. In particular, the magnitude of the second smallest eigenvalue λ_2 - the so-called *algebraic connectivity* - captures how "well-connected" a network topology is. It can be related to a network's expansion and congestion properties, the existence of small cuts, the performance of gossip algorithms, the average shortest path length or its diameter [Mohar, 1991;

[2]Note that this closely resembles the vanishing epidemic threshold mentioned in section 2.3.3.

[3]The continuous Laplacian is a differential operator which is given by the sum of a function's second-order unmixed partial derivatives with respect to cartesian coordinates. It occurs for example in the description of electromagnetic and gravitational fields, diffusion dynamics and wave propagation.

Chung, 1997]. Important for the question considered in this chapter, it has been shown for example in [Barahona and Pecora, 2002; Jadbabaie *et al.*, 2004; Pecora and Barahona, 2005] that aspects of a network's tendency to synchronize, like for instance the minimally required coupling strength in the Kuramoto model or the stability of synchronized states in oscillator models with linear couplings, can be related to its Laplacian spectrum. In particular, it has been argued that networks with smaller *eigenratios*

$$\frac{\lambda_{max}}{\lambda_2}$$

generally favor the emergence of stable synchronized states. This result formally underpins the intuition that synchronization emerges in "well-connected" network topologies with a large algebraic connectivity λ_2. Similarly, networks with a large algebraic connectivity have been found to be crucial for a number of consensus, control and cooperation mechanisms that are relevant in the context of distributed computing systems [Olfati-Saber *et al.*, 2007]. In fact, with the random walk sampling and the gossip-based aggregation scheme used in chapter 4, we have considered two particular examples for distributed algorithms whose performance benefits from large algebraic connectivity.

Synchronization in Networks with Modular Structures

When considering the impact of a network's topology on synchronization, apart from studying if stable synchronized states emerge, a further interesting aspect is how *local structural features influence the dynamics of the synchronization process*. Recursively applying above arguments about the influence of algebraic connectivity to the substructures of a network, one would intuitively expect that *synchronized regimes preferentially emerge in well-connected subgraphs of a network*. Results presented in [Moreno *et al.*, 2004; Oh *et al.*, 2005; Park *et al.*, 2006] underpin this intuition. Here, it was found that nodes within the same *community* - that is a well-connected subset of nodes for which links to nodes *within* the subset are more likely than to nodes *outside* - tend to synchronize quickly. Long-range links that interconnect different (synchronized) communities have further been found to be crucial for the subsequent formation of increasingly large synchronized clusters and eventually for the emergence of globally coherent states.

Assessing Spectral and Topological Properties

Similar like the findings about the emergence of stable synchronization, the dynamics of synchronization processes in network with modular structures can again be related to the distribution of eigenvalues and eigenvectors of the Laplacian matrix that governs the geometry of oscillator couplings [Arenas *et al.*, 2006a; Arenas *et al.*, 2006b]. The resulting connections between spectral

graph properties, topological features of complex networks and the dynamics of synchronization processes open up a number of interesting perspectives. Algorithms that compute spectral properties of a network's Laplacian matrix have for example been proposed to address the problem of assessing community structures in the analysis of large scale networks [Donetti and Munoz, 2004; Capocci *et al.*, 2005]. Alternatively, it has been suggested that knowledge about a network's modular and hierarchical organization can be inferred by monitoring the progressive evolution of synchronized regimes in complex networks [Arenas and Díaz-Guilera, 2007].

5.2 Gossip-Based Synchronization in Complex Networks

Self-Organized Synchronization in Networked Computing Systems

The results about self-organized synchronization in networks summarized above have paved the way for a targeted application in engineered networked systems. Since they map well to a message-based communication paradigm, in the past Integrate-and-Fire models have primarily been considered for this purpose. Targeting the problem of time synchronization in base-station free wireless ad hoc networks, in [Hong and Scaglione, 2003] the use of the Integrate-and-Fire model presented in [Peskin, 1975] has been considered. Similarly, in [Lucarelli and Wang, 2004] the same scheme has been considered for a synchronization of small-scale, time-varying sensor network topologies based on nearest neighbor communication. Addressing the same scenario, in [Werner-Allen *et al.*, 2005] a synchronization protocol has been considered which is based on the firefly-inspired pulse coupling model studied in [Mirollo and Strogatz, 1990]. Apart from sensor networks, the use of self-synchronization models has been proposed for large scale Peer-to-Peer systems. Here, a synchronized notion of "time epochs" or "heartbeats" is for instance required by a number of turn-based distributed algorithms. For such a scenario, in [Babaoglu *et al.*, 2007] a synchronization protocol has been proposed that is based on pulse-based, nearest neighbor couplings according to the adaptive model presented in [Ermentrout, 1991]. Closest to the scheme presented in this section, in [Baldoni *et al.*, 2009] a synchronization protocol for large dynamic Peer-to-Peer systems has been proposed that is based on a model of linearly coupled oscillators.

Pulse-Couplings in Computer Networks

In the remainder of this chapter, we investigate how findings on the self-organized synchronization in complex networks can be incorporated in a distributed protocol for large scale Peer-to-Peer systems. A particular focus will be

laid upon the use of the protocol in overlays with probabilistic, complex structures as proposed in chapter 3. For this, similar like in the Integrate-and-Fire models we assume a sporadic coupling that maps well to the message-based, non-continuous communication between computing devices. However, for this *we do not make use of an Integrate-and-Fire model since we argue that the simultaneous exchange of coupling messages resulting from the synchronization of pulses threatens their applicability in large scale computing systems*. In fact, as we have seen in chapter 1.1, highly synchronized traffic peaks in the Internet can result in massive problems and countermeasures are usually taken to prevent them. To overcome this problem in situations where a synchronization of periodic processes is required without a central instance and without imposing the burden of synchronized message exchanges, in the following section we describe a synchronization protocol that makes use of gossip-like message exchanges. The scheme is inspired by the Kuramoto model and - similar like the protocol considered in [Baldoni *et al.*, 2009] - assumes that continuous couplings in the underlying coupled oscillator model are replaced by a sporadic exchange of messages. It has originally been presented in [Scholtes *et al.*, 2009] as well as - in an extended version - in [Scholtes *et al.*, 2010].

5.2.1 Coupling Protocol

A Message-based Kuramoto Model

In the following, we give a detailed description of the proposed scheme along with two different coupling models. We assume that oscillators reside at the nodes of an arbitrary undirected network $G = (V, E)$. Let each node $v \in V$ possess a randomly skewed, discrete clock t_v, a Gaussian distributed intrinsic frequency ω_v as well as an oscillating signal γ_v. Defining the period T_v of an oscillator v as $T_v = \frac{1}{\omega_v}$, a node v's phase within its period can then be given as

$$\Theta_v = \Theta_v(t_v) = 2\pi \frac{t_v \, mod \, T_v}{T_v}.$$

$\Theta_v \in [0, 2\pi)$ representing a node v's current phase in its oscillator cycle, its oscillating signal $\gamma_v(t_v)$ is then given as

$$\gamma_v = \gamma_v(t_v) = sin\left(\Theta_v\right).$$

Message-based Coupling

For the sake of making the original Kuramoto model applicable in practical distributed systems, rather than a continuous coupling to all nodes, we consider a message-based local coupling, that is we assume that nodes are coupled to their nearest neighbors via sporadic, probabilistic message exchanges. In a time-discrete manner, this sporadic coupling can for example be modeled

as a per-node vector of Bernoulli processes, that is in each time step a node v performs d_v independent Bernoulli trials X_i ($i = 1, \ldots, d_v$) with d_v being the degree of node v. Upon success of X_i, v is coupled to its i-th neighbor. Let $p \in [0, 1]$ be the uniform success probability of these Bernoulli trials. Clearly, for $p = 1$ one obtains a time discrete version of the Kuramoto model with nearest neighbor couplings. For $p = 0$ there is no coupling at all, that is the model is identical to the case of $K = 0$. Based on the first moment of the binomial distribution and the fact that each node can either choose to couple to a neighbor or be chosen as coupling partner, for the total expected number of message exchanges during n time steps in a node v one obtains

$$2 \cdot n \cdot p \cdot d_v.$$

Coupling Functions

So far, instead of defining the actual coupling we have rather considered whether a coupling occurs in a given time step. Different from the symmetric coupling constant K in the original Kuramoto model, for our simulations *we assume that the coupling strength is given as a function of the degrees of both coupling partners*. Hence we consider an asymmetric coupling strength that may differ for couplings between different nodes. If - based on the stochastic process described above - a coupling between two nodes v and w occurs, we assume that both nodes exchange their current phases Θ_v and Θ_w in their oscillator cycles. It is then assumed that both nodes v and w update their periods - and thus the speed of their phase advance - in the following way:

$$\begin{aligned} T_v = \quad & T_v + f(d_v, d_w) \cdot sin(\Theta_v - \Theta_w) \\ T_w = \quad & T_w + f(d_w, d_v) \cdot sin(\Theta_w - \Theta_v) \end{aligned} \tag{5.2}$$

With this, the cumulative effect of all couplings occurring during one time step n at a node v can be reformulated in terms of the change of the instantaneous phase advance speed $\omega_v^{(n)}$ by the following recurrence relation:

$$\omega_v^{(n+1)} = \left(\frac{1}{\omega_v^{(n)}} + \sum_{w \in V : (v,w) \in E} B_v^{(n)}[w] \cdot f(d_v, d_w) \cdot sin\left(\Theta_v^{(n)} - \Theta_w^{(n)} \right) \right)^{-1}$$

Degree-Weighted Coupling

For this we assume that in a time step n, the success of a Bernoulli trial of a node v for its neighbor w is indicated by a one-entry in a vector $B_v^{(n)}$, while a failure is indicated by a zero-entry at the corresponding position $B_v^{(n)}[w]$. In the

above definitions f denotes a coupling strength function whose value may depend on the degrees of both coupling partners. Speaking informally, the intention behind the coupling in the original Kuramoto model shown in equation 5.1 was to adjust local oscillator frequencies by K times the average angular phase difference to all other coupling partners. Applying this argument to the case of nearest neighbor couplings results in the following degree-weighted coupling function f_1:

$$f_1(d_v, d_w) := \frac{K}{d_v} \tag{5.3}$$

Here K again denotes a coupling strength constant. Own experiments (presented in [Scholtes *et al.*, 2009]) as well as arguments presented for example in [Motter *et al.*, 2005b; Li, 2008] suggest that a degree-weighted, asymmetric coupling strength as proposed in f_1 is required for stable synchronized states to emerge in networks with heterogeneous degree distributions and nearest neighbor couplings. In fact, it has been found in [Nishikawa *et al.*, 2003] that heterogeneous degree distributions can actually hinder the formation of synchronization when using constant and symmetric coupling strength. Justifying our choice of f_1, it has been shown in [Motter *et al.*, 2005b] that for a given network topology and a weighting of coupling strengths by $d_v^{-\beta}$ a maximum eigenratio and thus synchronizability is obtained for $\beta = 1$.

Harnessing Heterogeneity

With the coupling function f_1, one observes that, in terms of coupling strength, a node v does not differentiate between different neighbors w. When wanting to provide fast and stable synchronization, one is tempted to think about the *credibility* of a node's oscillator state in the sense of how "representative" it is for the macroscopic state of synchronization. While for many networks one may assume equal *credibility*, this assumption is not justified in networks with highly heterogeneous nodes like for example the case in random scale-free networks. Speaking informally, the synchronization state of highly connected hub nodes is likely to be more "reliable" since their oscillators are coupled (though weakly due to the degree weighting in f_1) to a larger number of nodes. In fact, it has been argued for example in [Moreno and Pacheco, 2004], that in in scale-free networks stable synchronized states are likely to first emerge around hub nodes and the resilience of the synchronization against perturbations is higher for nodes with large degrees. The fact that the attractiveness of hubs in real-world networks often stems from higher stability, longer uptime or better capabilities can be seen as another argument speaking in favor of a degree-based differentiation of a neighbor's *credibility*.

Target-Degree Weighted Coupling

It thus seems justified to consider a coupling model in which hubs are selectively privileged by amplifying the coupling strength to nodes with higher degrees[4]. While maintaining the degree-based weighting of f_1, one can define an alternative coupling function f_2 which incorporates this aspect as follows:

$$f_2(d_v, d_w) := \frac{K \cdot d_w \cdot \bar{d}^{-1}}{d_v} \tag{5.4}$$

In this definition $\bar{d}$ denotes the average vertex degree and K is again a constant adjusting the strength of couplings. For networks, in which all nodes have the same degree (for example in regular lattices), due to $\bar{d} = d_w$ the coupling f_2 is identical to f_1. Furthermore one observes that for fully connected network topologies, both coupling functions f_1 and f_2 are identical to the coupling in the original Kuramoto model. For other networks, the coupling strength to nodes with degrees above the global average is amplified while the coupling to nodes with degrees below the average is weakened proportionally. Moreover, the magnitude of the coupling strength fluctuations is coupled to the heterogeneity of the degree distribution.

5.2.2 Experimental Evaluation

Synchronization in Complex Network Topologies

Using the simple model presented above, we now study the emergence of synchronized oscillator states in complex network topologies. Following the work that has been presented in the previous chapter 4, we will consider the performance of the proposed scheme in scale-free network topologies. Moreover, we study the collective dynamics arising in networks generated according to the Watts/Strogatz model which has earlier been presented in 2.2. In the following section we present simulation results for both coupling functions f_1 and f_2 as well as for different coupling probabilities p. Particular emphasis is placed on whether coherent oscillator states emerge as well as on the minimally required message exchange frequency. In further experiments we will investigate the resilience of synchrony against perturbations. In order to assess the achieved level of oscillator coherence and the dynamics of the synchronization process, we will use the common time-dependent order parameter r, which for

[4]Similar ideas have been considered in [Hwang *et al.*, 2005; Motter *et al.*, 2005a].

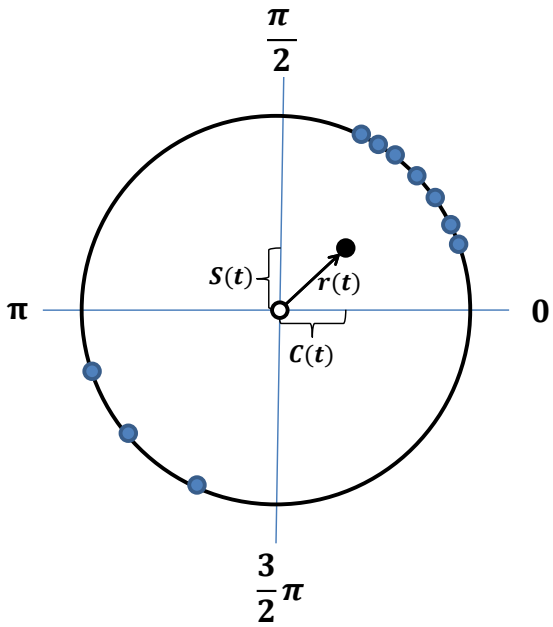


Figure 5.1: Example of order parameters $S(t)$, $C(t)$ and $r(t)$ for snapshot of 10 oscillator states

$$S(t) = \frac{\sum_{v \in V} sin(\Theta_v)}{|V|}, C(t) = \frac{\sum_{v \in V} cos(\Theta_v)}{|V|}$$

is defined as

$$r(t) = \sqrt{S(t)^2 + C(t)^2}.$$

Interpretation of Order Parameter

Here, $S(t)$ can be viewed as a macroscopic "sound" generated by the superposition of individual oscillatory "tones" $\gamma_v = sin(\Theta_v)$. When interpreting an oscillator's phase $\Theta_v \in [0, 2\pi)$ as position $e^{i\Theta_v}$ on the unit circle, $r(t)$ corresponds to the distance of the center of mass of all oscillator positions from the circle's center at time t. Values of $r(t)$ close to 0 represent unsynchronized states in which oscillator positions are equally distributed across the circle. In states of coherent oscillations, $r(t)$ is close to 1 since all oscillators are concentrated at a particular position and the center of mass approaches the unit circle. An example for a snapshot of oscillator states can be seen in Figure 5.1.

Power Spectra of Oscillator Populations

Another way of analyzing both the time-variant global "sound" S as well as the individual node's local signals γ_v, is to study the dynamics and power spectra of these signals. In a perfectly coherent state, S and γ_v are pure-sinusoidal signals with a peak-to-peak amplitude of two and the power spec-

trum is a delta function with a single peak at the (global) oscillator frequency. For S, the power spectrum can reveal intermediate states in which certain clusters oscillate at different frequencies. For γ_v, the power spectrum reveals the level of noise in the signal that is due to signal perturbations resulting from couplings with oscillators running at different frequencies. Later in this chapter, we will use a discrete Fourier transform in order to analyze power spectra of a discrete-time signal obtained by a numerical simulation of the synchronization scheme. In order to reduce spectral leakage, the signal was multiplied with a Hamming window function [Oppenheim *et al.*, 1999] prior to performing a Fast Fourier transform. Figure 5.2 illustrates above metrics in two reference scenarios: In the first case a population of 10000 oscillators has been artificially synchronized, in the second case no synchronization was used and the intrinsic periods of individual oscillators were normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 25$. In the remainder of this article we will abbreviate such an initial normal distribution of oscillator periods as $N(\mu, \sigma)$. All simulations in the remainder of this chapter using the simulation environment TOPGEN [Scholtes *et al.*, 2008b; Botev *et al.*, 2009]. Fast Fourier Transforms have been computed using the open source statistical computing environment R.

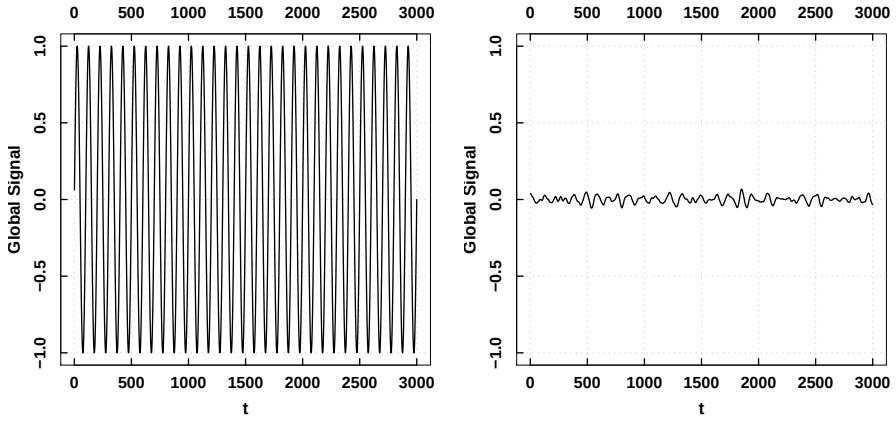
Coupling Functions

Preliminaries

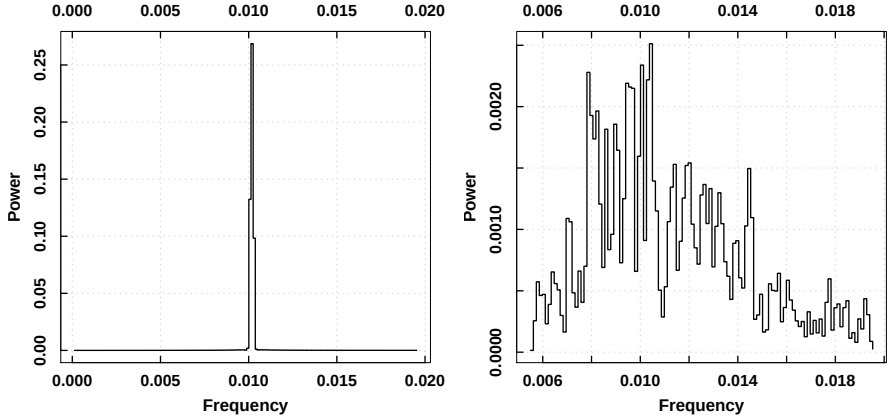
In the following we present simulation results for different complex network topologies and coupling functions f_1 and f_2. Unless explicitly stated otherwise, all results are mean values of at last five independent simulation runs on connected, random network topologies with 10000 nodes and roughly 50000 edges. For the experiments in this section, the intrinsic periods of oscillating nodes were $N(100, 25)$-distributed. Initially, phase skews of individual nodes were uniformly distributed within a node's intrinsic period. For the results in this section, no message latencies or perturbations were considered and a coupling probability $p = 1$ was used. Furthermore, results are shown for a coupling strength constant $K = 0.1$ which was checked (experimentally) to lie above the critical value K_c required for synchrony to emerge.

Degree-Weighted Coupling in Watts/Strogatz Networks

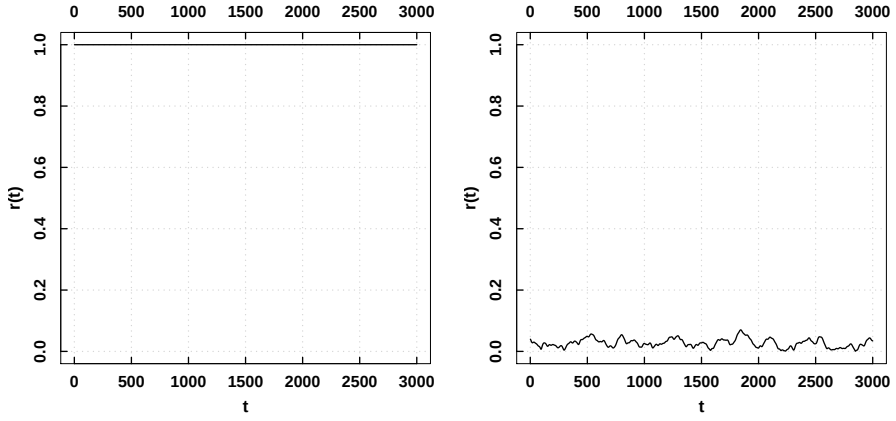
We begin our evaluation by considering networks generated according to the Watts/Strogatz model [Watts and Strogatz, 1998]. The main advantage of this model is that it allows to study the interrelation between a network's small-world characteristics and the emergence of self-synchronization. For this, two model parameters k (the number of nearest neighbors in an initial ring lattice)



(a) $S(t)$ (time domain)



(b) $S(t)$ (frequency domain)



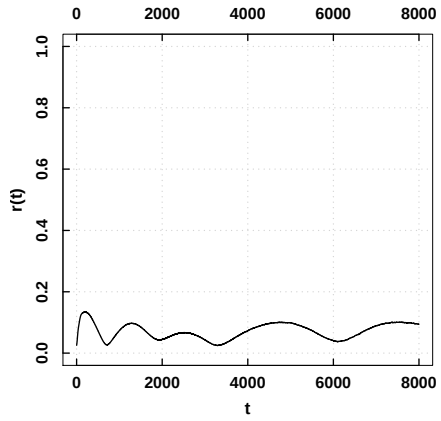
(c) Order parameter $r(t)$

Figure 5.2: Metric exemplification in a 10000 node reference graph of artificially synchronized (left) and unsynchronized (right) oscillators

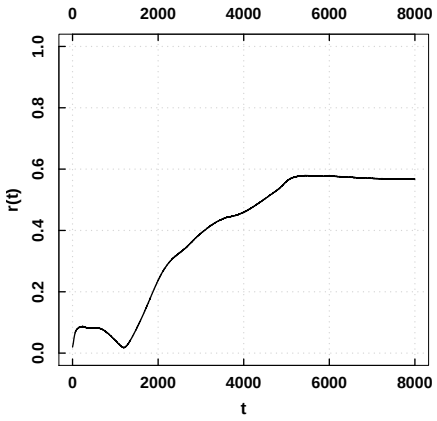
and p_r (the probability to reconnect one endpoint of an edge to a chosen uniformly at random) can be adjusted to produce graphs with tunable average path lengths, randomness and clustering. The resulting graph can range between a regular ring lattice ($p_r = 0$) and a completely random Erdös/Rényi graph ($p_r = 1$). It has been argued for example in [Barrat and Weigt, 2000] that Watts/Strogatz networks with N nodes exhibit small average path lengths if $p_r \gg \frac{1}{N}$. Throughout this chapter we denote Watts/Strogatz networks with reconnection probability p_r and k connections to nearest neighbors in the initial lattice as $WS(p_r, k)$. Because a $WS(p_r, k)$ network with N nodes has $\frac{N \cdot k}{2}$ edges and we intended to perform all simulations on equally-sized graphs with 10000 nodes and roughly 50000 edges, a parameter $k = 10$ was used in all subsequent experiments. Figure 5.3 shows the evolution of the order parameter $r(t)$ for different Watts/Strogatz networks when using the coupling function f_1. Since an averaging over several simulation runs would impair the significance of $r(t)$ as an illustration of the evolution of synchrony, results for a single representative simulation run are shown instead in Figure 5.3. These results suggest that for the degree-weighted coupling, synchronizability of a $WS(p_r, 10)$ network increases with higher reconnection probabilities p_r. For completely random $WS(1, 10)$ networks consisting of 10000 nodes and 50000 edges, synchrony emerges quickly within roughly 500 simulated time steps. This finding has been substantiated by additional experiments in which some more points in the Watts/Strogatz model's parameter space were explored. For this, the order parameter $r(t)$ eventually reached during 10000 simulated time steps has been recorded. The averaged results for different reconnection probabilities p_r are shown in Figure 5.4 on a log-scale x-axis. These results for our Kuramoto-inspired synchronization scheme with asymmetric coupling function f_1 are consistent with the findings presented in [Watts, 1999] and mentioned in [Watts and Strogatz, 1998].

Degree-Weighted Coupling in Scale-Free Networks

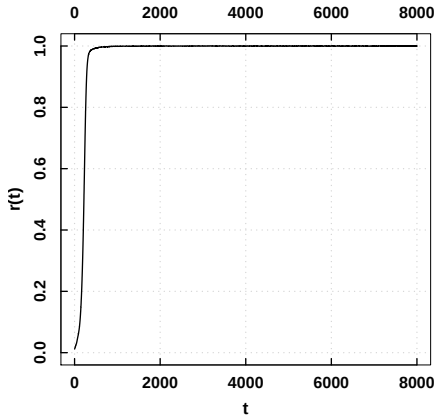
Based on the results presented in chapter 4, one can imagine situations in which one wants to use the synchronization scheme in Peer-to-Peer systems with scale-free overlay topologies. In the following, we thus present simulation results for the synchronization behavior in such networks. Being a simple and convenient non-equilibrium model for the emergence of scale-free networks, in the following we have used network topologies generated by the Barabási/Albert preferential attachment model [Barabási and Albert, 1999] which has already been mentioned in 2.3.5. For the following simulations, the number of links l that each newly added node established to preferentially



(a) $WS(0.0001, 10)$



(b) $WS(0.01, 10)$



(c) $WS(1, 10)$

Figure 5.3: Order parameter $r(t)$ in Watts/Strogatz topology with coupling f_1

chosen existing nodes has been set to $l = 5$. This resulted in network topologies with roughly the same number of nodes and edges as the Watts/Strogatz topologies considered above and allows to factor out differences in synchronizability that are due to different network sizes. The fitted degree distribution exponent of the generated 10000 node scale-free networks was slightly smaller than three. Figure 5.5 shows the evolution of the order parameter $r(t)$ for a representative simulation run in a 10000 node Barabási/Albert network using coupling strength function f_1. Similar to the results for 10000 node $WS(p_r, 10)$ graphs with $p_r \gg \frac{1}{N}$, one observes fast convergence towards a synchronized state within roughly 500 simulated time steps.

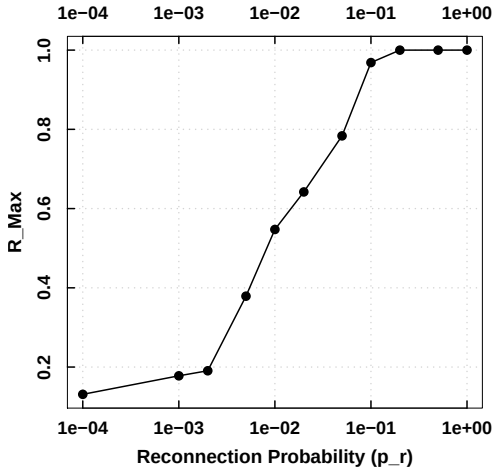


Figure 5.4: Average maximum order in $WS(p_r, 10)$ graph for coupling f_1 (Lines are drawn to guide the eye)

Target-Degree-Weighted Coupling in Watts/Strogatz Networks

We now study the emergence of synchrony when using the coupling function f_2, that is the coupling strength to high degree nodes is amplified while the coupling strength to small degree nodes is decreased proportionally. Since the node degree distribution of Watts/Strogatz topologies is rather homogeneous and all node degrees are likely to be close to the global average, one expects no significant deviation from the behavior using the f_1 coupling. The evolution of the order parameter $r(t)$ during the first 1500 iterations of two representative simulation runs with coupling functions f_1 and f_2 can be seen in Figures 5.6(a) and 5.6(b). The dynamics of the order parameter is nearly identical in both cases. The situation is different when looking at Barabási/Albert networks with a highly heterogeneous node degree distribution. Figures 5.6(c) and 5.6(d) show the order parameter $r(t)$ during the first 500 iterations of two representative simulation runs using coupling functions f_2 and f_1. Here one observes a roughly three times faster evolution of oscillator coherence. These rather illustrative results have been substantiated by further experiments in which the average number of iterations required to achieve an order parameter $r(t) > 0.95$ were recorded. The table in Figure 5.7 suggests that the number of couplings required to achieve synchrony in scale-free networks is smaller than in small-world Watts/Strogatz networks when using degree-weighted couplings. Moreover, *selectively amplifying coupling strength for high degree nodes further reduces the number of required couplings*. As such, for an application of the

Target-Degree-Weighted Coupling in Scale-Free Networks

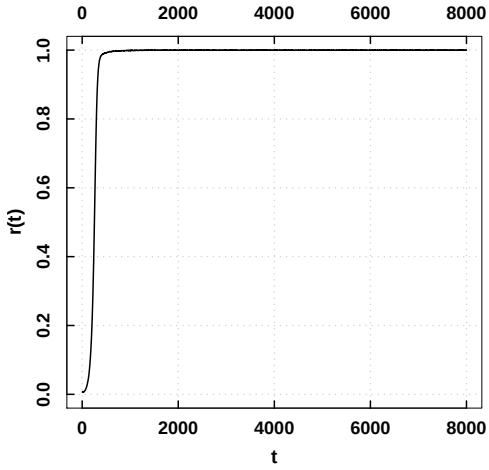


Figure 5.5: Order parameter $r(t)$ in Barabási/Albert topology with coupling f_1

proposed synchronization scheme in scale-free network topologies, the use of the coupling function f_2 seems to be desirable.

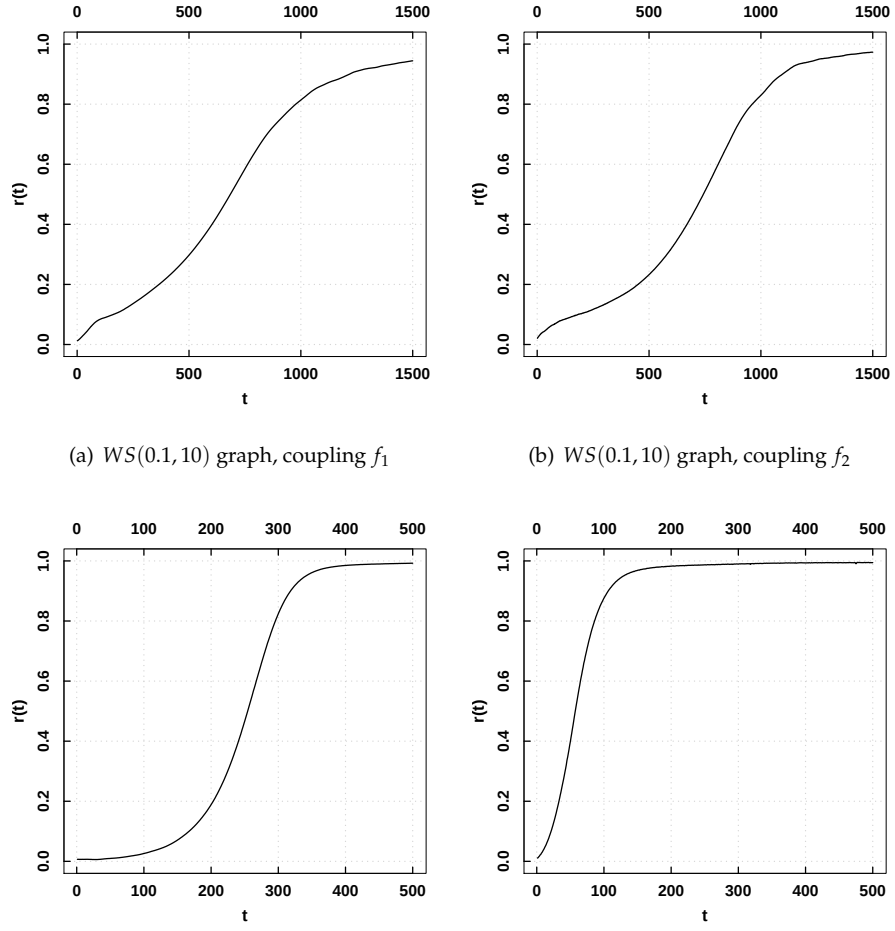
Coupling Probability

Sporadic Coupling

So far, results have been presented only for the case of a coupling probability $p = 1$, that is nodes were coupled to all nearest neighbors in each time step of the simulation. While this corresponds to a time-discrete version of the Kuramoto model, for a synchronization protocol usable in practical systems this is clearly not feasible. Hence, it is interesting to study how a decrease in coupling probability affects the onset of synchronization. In practical terms, assuming a low probability per-node coupling has obvious advantages: The number of messages exchanged per time unit and thus the bandwidth requirements of the synchronization protocol are reduced. Furthermore - at least up to a certain extent - sporadic communication errors are hidden in the underlying stochastic process and do not require special treatment.

Synchronization Time

When reducing the probability and thus the number of couplings per time unit, one may assume that the level of synchronization eventually reached remains unchanged while the time required for synchronization increases proportionately. Intuitively, one expects that halving the coupling probability and doubling the simulation duration results in the same level of synchrony since the overall number of couplings remains unchanged. The results presented in



(a) $WS(0.1, 10)$ graph, coupling f_1

(b) $WS(0.1, 10)$ graph, coupling f_2

(c) BA graph, coupling f_1

(d) BA graph, coupling f_2

Figure 5.6: Evolution of order in Watts/Strogatz and Barabási/Albert networks with couplings f_1 and f_2

Coupling	$WS(0.1, 10)$	Barabási/Albert
f_1	1175	279
f_2	1142.4	122.6

Figure 5.7: Average iterations to 95 % coherence in 10000 node networks

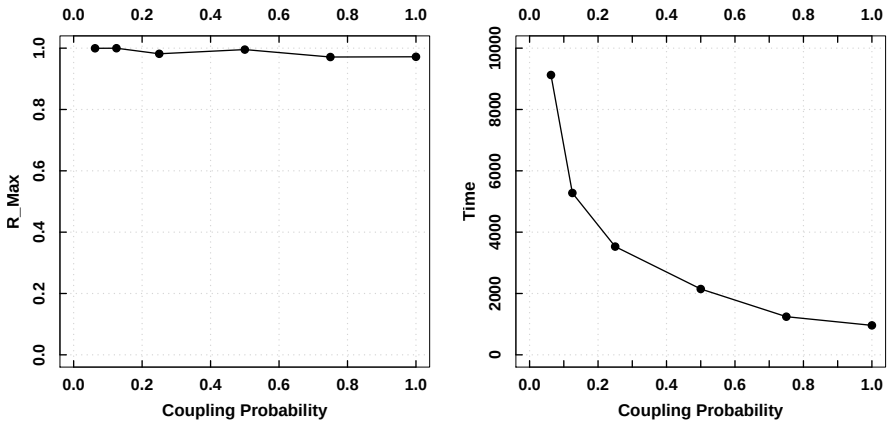


Figure 5.8: Average maximum $r(t)$ (left) and time required to reach 95 % coherence (right) for different coupling probabilities in a $WS(0.1, 10)$ network using coupling function f_1 (Lines are drawn to guide the eye)

Figure 5.8 show that this intuition seems to be correct. In these experiments, the average maximum value of the order parameter $r(t)$ has been obtained for several simulations lasting for 20000 discrete time steps. In order to compare the time at which synchronization emerges, the number of steps required on average to reach $r(t) > 0.95$ has again been recorded for $WS(0.1, 10)$ graphs with 1000 nodes and 5000 edges.

Bandwidth Demand vs. Synchronization Time

From a practical point of view, the result that halving coupling probability (and thus the frequency of coupling message exchanges) results in a doubling of the time required for oscillators to become coherent is not very interesting. This is because the overall number of messages that need to be exchanged to achieve a fully synchronized state remains roughly the same for any coupling probability. As such, the decrease of bandwidth for synchronization comes at the prize of requiring significantly more time. For an application in practice, it is an important question whether schemes with more sporadic couplings and thus smaller bandwidth demand can be used without exhibiting this drawback. One idea that comes into one's mind is whether more sporadic couplings can be compensated by a proportionate increase in coupling strength.

Compensating Coupling Frequency by Coupling Strength

A simple way to incorporate such a compensation of sporadic couplings is to use the following coupling function

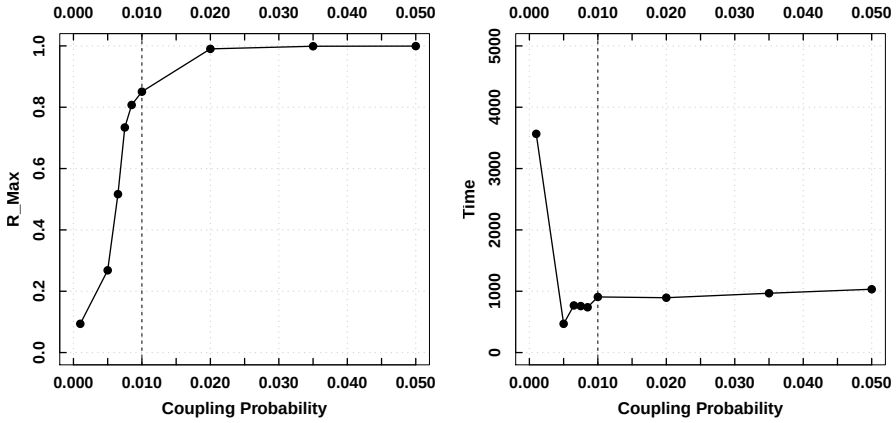
$$f_3(d_v, d_w) := \frac{K}{d_v \cdot p} \tag{5.5}$$

with p being the nodes' uniform coupling probability and K being again a coupling strength constant. For $p = 1$ we obtain the coupling function f_1 which means that simulation results for the coupling function f_1 can actually be reinterpreted in terms of the coupling f_3 with $p = 1$. More sporadic couplings $p < 1$ result in a proportional increase of coupling strength. In the limit of $p \to 0$ we obtain a model of infinitely rare and infinitely strong couplings. In Figure 5.9, the maximum order parameter $r(t)$ reached on average as well as the average number of time steps after which this maximum was reached is shown for different coupling probabilities and the coupling function f_3 in 1000 node Barabási/Albert and Watts/Strogatz $WS(0.1, 10)$ networks with roughly 5000 edges. As constant coupling strength factor, a value of $K = 0.1$ has again been used. The results suggest that the coupling probability p can in fact be largely reduced (by a factor of at least 50) without having significant effects on either the maximum coherence or the synchronization time, if the strength of the coupling is increased proportionally. Interestingly, a reduction of coupling probability below $p = 0.01$ results in a sudden drop of maximum coherence as well as a sudden increase of the time required to achieve this lower level of coherence.

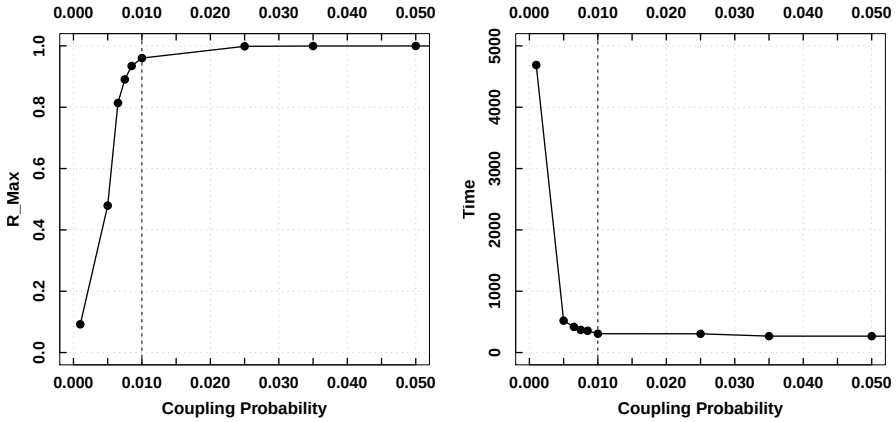
Limits of Compensating Coupling Frequency

Clearly, more sporadic couplings can only be compensated up to a certain extent as for $p \to 0$ the number of couplings approaches zero. In the following we examine how the minimally required coupling frequency in our synchronization model is influenced by other parameters. Experiments have been performed for further values of the average oscillator period μ and 1000 node Barabási/Albert and Watts/Strogatz networks. However, results are presented only for Barabási/Albert topologies since those for $WS(p_r, k)$ graphs with p_r above the small-world threshold have shown to be identical. The plots in Figure 5.10 suggest that the minimally required coupling probability depends on the initial distribution of oscillator frequencies. In a plot for an experiment with a mean oscillator frequency $f = \frac{1}{\mu}$, a dashed line indicates f on the axis corresponding to the coupling probability. When using coupling probabilities smaller than the average frequency, a node will initiate on average less than one coupling per mean oscillator period and neighbor. In our experiments, for coupling probabilities $p \ll f$, oscillator coherence did not emerge even after

a very long time. For practical applications, this suggests that the bandwidth demand of Kuramoto-based synchronization decreases proportionally with a reduction of the mean frequency of the oscillator population if more sporadic couplings are compensated for by a proportional increase in coupling strength. Thus, at least in Watts/Strogatz and Barabási/Albert network topologies, *a minimum coupling intensity of (on average) one coupling per oscillator period and neighbor seems to be sufficient for synchronized states to quickly emerge*.

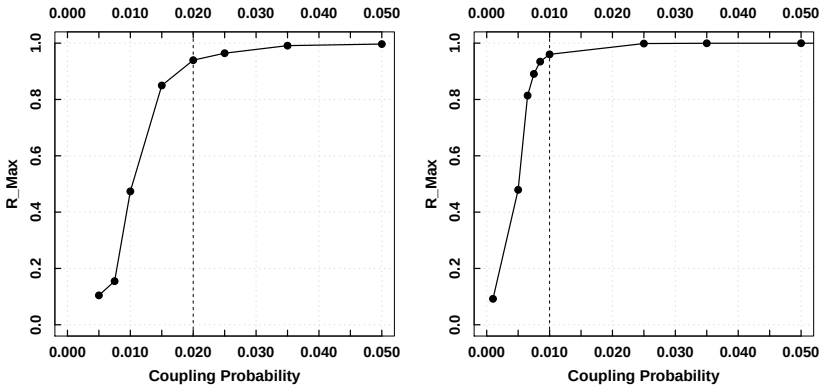


(a) Watts/Strogatz networks with $p_r = 0.1$



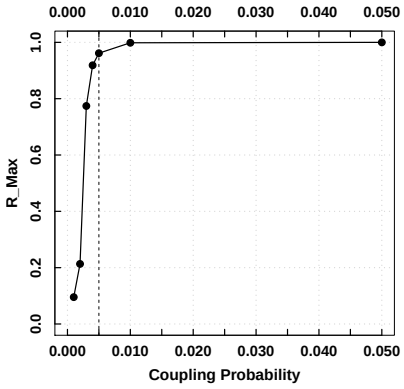
(b) Barabási/Albert networks

Figure 5.9: Average maximum $r(t)$ and sync time with strength-compensated sporadic coupling f_3 and $N(100, 25)$-distributed intrinsic oscillator periods (Lines are drawn to guide the eye)

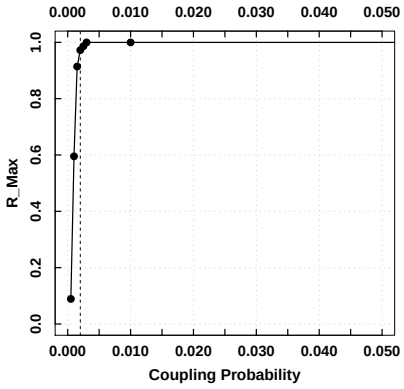


(a) $\mu = 50, \sigma = 12.5$

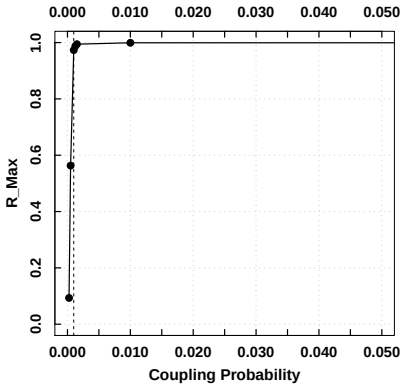
(b) $\mu = 100, \sigma = 25$



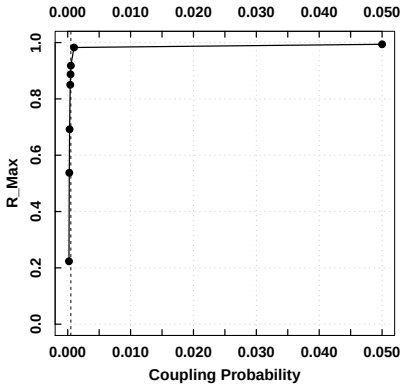
(c) $\mu = 200, \sigma = 50$



(d) $\mu = 500, \sigma = 125$



(e) $\mu = 1000, \sigma = 250$



(f) $\mu = 2000, \sigma = 500$

Figure 5.10: Average maximum order $r(t)$ with strength-compensated sporadic coupling f_3 in Barabási/Albert networks using different intrinsic oscillator period distributions (Lines are drawn to guide the eye)

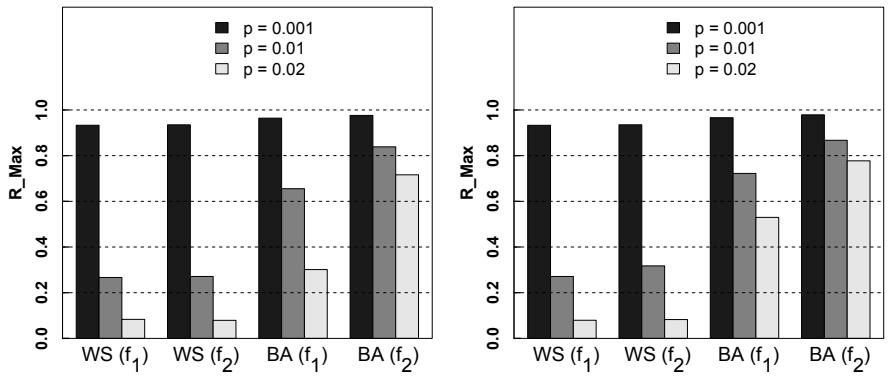
Synchronization Resilience

Fluctuations in Real-World Systems

A major hurdle that complicates the application of distributed algorithms in real-world settings like for example large scale Peer-to-Peer systems is the impact of the dynamics induced by leaving or joining participants as well as failing devices and network connections. An important task when designing distributed algorithms for such systems is to assess their susceptibility to this kind of perturbations. In the following, the effect of random node replacements on oscillator coherence is investigated. In order to measure the effect of node dynamics rather than that of the associated topology changes, it is assumed that failing nodes are immediately replaced in a way that does not change the vertex degree distribution or other network characteristics. In the following simulations, this has been modeled by reinitializing oscillator states for a certain fraction of randomly chosen nodes to frequencies and phases drawn from the initial distribution.

Random Failures

In a first series of experiments, the effect of oscillators being reinitialized uniformly at random has been investigated, that is in each simulated time step all nodes had a uniform probability of being replaced. Three per-node replacement probabilities $p = 0.02$, $p = 0.01$ and $p = 0.001$ have been simulated for Barabási/Albert and Watts/Strogatz network topologies consisting of 10000 nodes. The resulting average maximum order parameter $r(t)$ is shown in Figure 5.11(a). While a uniform replacement probability of $p = 0.001$ per node and iteration did result in only a moderate decrease of coherence for both, Watts/Strogatz and Barabási/Albert networks, a significant difference can be seen for replacement probabilities $p = 0.01$ and $p = 0.02$. For both coupling functions f_1 and f_2, the average maximum order parameter $r(t)$ is significantly higher in Barabási/Albert topologies: For f_1 and a failure rate of $p = 0.01$ the retained coherence is increased by approximately 150 % compared to Watts/Strogatz networks. For f_2 and a failure rate of $p = 0.02$, oscillator coherence in Barabási/Albert networks is roughly nine times higher than in Watts/Strogatz topologies. This can be attributed to the stabilizing effect exerted by the small number of highly connected hubs. The use of the coupling function f_2, which selectively increases the coupling strength to stable hubs, seems to be suitable to further this effect. When comparing the average maximum order in Barabási/Albert topologies using couplings f_1 and f_2, for f_2 one observes a roughly 28% increase in coherence for a failure probability $p = 0.01$ and a 137% increase for $p = 0.02$. In contrast, no significant increase can be observed for Watts/Strogatz networks.



(a) Random Churn (b) Degree-Coupled Churn

Figure 5.11: Average maximum order $r(t)$ under different types of perturbation

Degree-Based Failures

As stated in the considerations leading to the introduction of the coupling strength function f_2, for many real-world network topologies with highly diverse vertex degrees it is justified to relate a node's degree to its reliability. Highly connected Internet routers are probably - or at least hopefully - better secured against failures and attacks, websites with many incoming links would probably not have become popular if they were not well-maintained, and hubs in Peer-to-Peer overlay topologies often emerge because they have been reliable in the past. It thus seems reasonable to consider a model in which a node's probability to fail is reciprocally proportional to its degree. Figure 5.11(b) shows results of a series of simulations based on such a degree-based perturbation model. The total number of nodes being replaced in each iteration was the same as in the previous experiments with uniform replacement probabilities, that is in each simulated time step a total of 0.1%, 1% respectively 2% of all oscillators were reinitialized with a random intrinsic period and phase. Results of these experiments are shown in Figure 5.11(b). The level of coherence maintained in Barabási/Albert networks is again higher than that in Watts/Strogatz topologies. For a failure rate of $p = 0.1$, the results for the degree-coupled perturbations are identical to those for the uniform model for a failure rate of $p = 0.001$. For higher levels of perturbation, one observes that the retained level of coherence in Barabási/Albert topologies is higher for degree-coupled failures. For a failure rate of $p = 0.01$, a roughly 10% increase can be observed for the coupling f_1 and a 4% increase for the coupling f_2. For

a failure rate of $p = 0.02$, the degree-based model results in a 73% (f_1) respectively 8% (f_2) increase over the uniform random model. In summary, the results suggest that *in scale-free networks the impact of perturbations on the level of oscillator coherence is rather mild if the failure rate is reciprocally proportional to node degrees and if coupling strengths to high degree nodes are selectively amplified.*

5.2.3 Monitoring Networks by Synchronization

Message-based Kuramoto Synchronization as Monitoring Protocol

The simulation results presented above as well as the findings summarized in section 5.1 indicate that the proposed synchronization scheme can be applied practically in network topologies with a sufficiently large algebraic connectivity. While the resulting coherence of per-node oscillating signals *per se* can be useful in certain scenarios, in this section we comment on a possibly more interesting, further application of the proposed scheme. For this, we recall that both the emergence as well as the dynamics of synchronization is crucially influenced by spectral properties of a network's Laplacian matrix. Based on relations between spectral properties and a network's structural organization, it has been proposed for example in [Arenas *et al.*, 2006a] that the global dynamics of the order parameter can actually be used to detect modular structures and argue about the hierarchical organization of large scale networks. In this section we address a similar question, however we focus on the issue whether the dynamics of local, microscopic oscillatory signals and coupling exchanges allows to infer knowledge about the macroscopic and mesoscopic organization of the network. Hinting at a related problem studied in general spectral theory [Kac, 1966], one is tempted to reformulate this in terms of the question whether a node can "hear" the topology of the network into which it is embedded.

Assessing Algebraic Connectivity

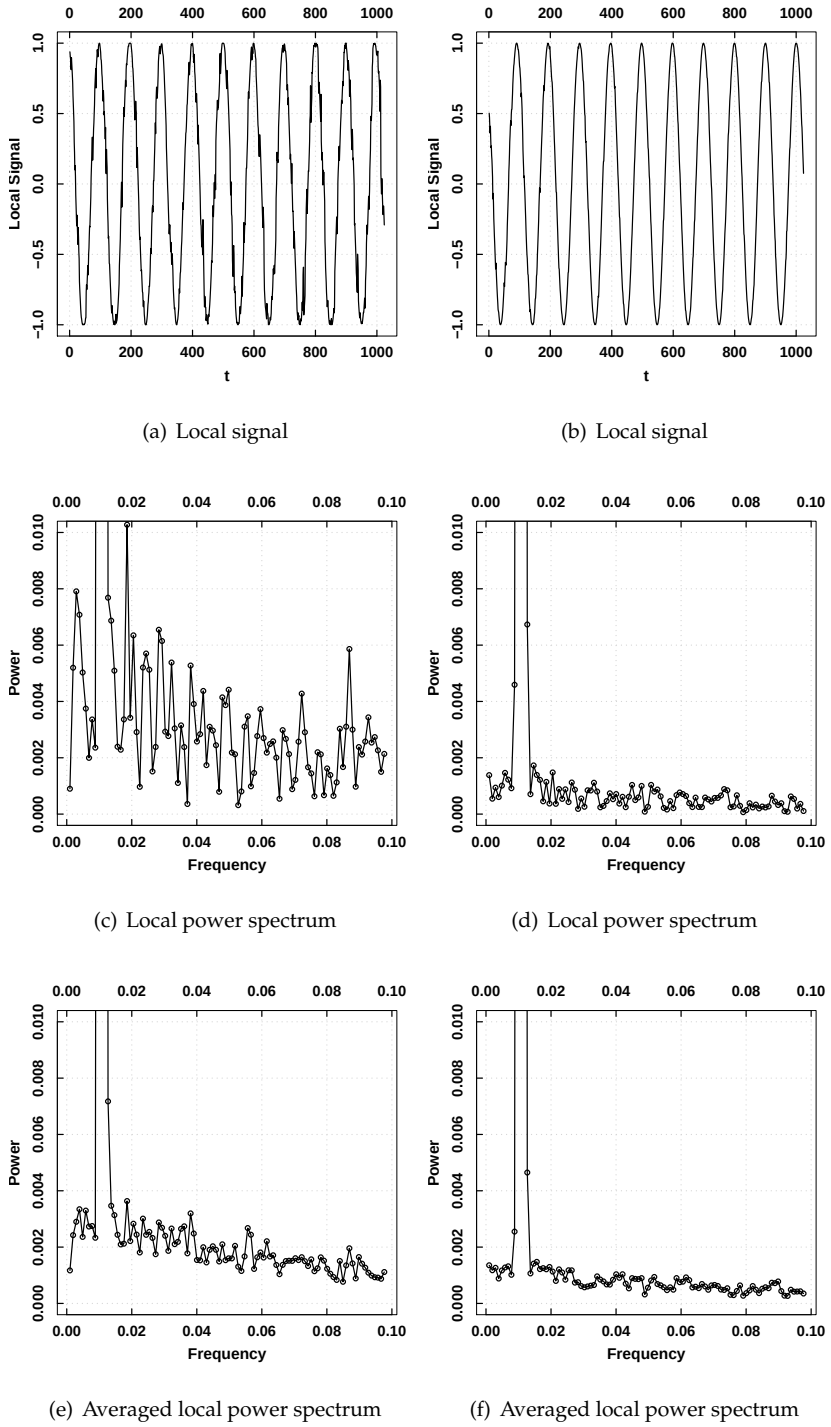
Assessing Algebraic Connectivity

We have seen in section 5.1 that whether a network topology facilitates self-organized synchronization depends on its algebraic connectivity. It has further been mentioned that this property has massive implications on the performance of other dynamical processes including information diffusion, network flows, distributed consensus and control schemes or the convergence behavior of random walks and gossip-based algorithms. The performance of such dynamical processes can be of great importance for the design of networked computing systems. Apart from monitoring the evolution of global order by means of the order parameter $r(t)$ as done in section 5.2.2, here we argue that the proposed synchronization protocol provides a very simple method to as-

sess the algebraic connectivity of the network *locally*. For this, a node merely needs to study its local oscillatory pattern as it is being influenced by couplings to neighboring nodes. In network topologies with large algebraic connectivity, oscillators will quickly synchronize and the angle between neighboring oscillators quickly ceases. As a result, deviations of the local oscillatory dynamics from a pure sinusoidal signal that are due to the period adjustments given in equations 5.2 will disappear. Contrariwise, the oscillatory signal of nodes will continue to be perturbed if no stable synchronized state emerges.

Experimental Results

In Figure 5.12, some experimental evidence for this intuition is presented. For this, simulations have been performed using 1000 node Watts/Strogatz network topologies with rewiring probabilities below and above the critical value required to obtain synchronizable networks. The left column shows the local oscillator signal $\gamma_v(t)$ of a node v chosen uniformly at random from a Watts/Strogatz network without random rewiring ($p_r = 0$) as well as a close-up view of the power spectrum of $\gamma_v(t)$ which has been obtained using a 1024 sample Fast Fourier Transform. The right column of Figure 5.12 shows the corresponding values for a randomly chosen node v in a Watts/Strogatz topology with rewiring probability $p_r = 0.1$ in the same time frame. Oscillator periods were $N(100, 25)$-distributed. A coupling with probability $p = 0.02$ and compensated coupling strengths according to equation 5.5 were used. When comparing the local signals in $WS(0, 10)$ and $WS(0.1, 10)$ topologies, one observes a significantly higher degree of noise in the local oscillations of nodes in the regular ring lattice $WS(0, 10)$. This noise is due to period adjustments driven by the prolonged incoherence of oscillators. A clear difference can also be seen in the level of noise in the nodes' power spectra shown in Figures 5.12(c) and 5.12(d). For illustration purposes, close-ups on frequencies near the center frequency are shown. While these are results of randomly chosen, individual nodes in a single simulation run, in Figures 5.12(e) and 5.12(f) local power spectra have been averaged over several simulation runs for different randomly chosen nodes. These results suggest that perturbations of a node's local oscillator can be used to reason about the synchronizability and thus the algebraic connectivity of a network topology. In the frequency domain, candidate single-valued metrics that allow a local discrimination of network topologies with large algebraic connectivity from others are average or variance of spectral power. Alternatively, a node can monitor the distribution of period offsets that are due to incoming coupling messages. Arguing that one can infer additional information from this dynamics, in the following we will study the evolution of couplings to a node's individual neighbors.



(a) Local signal

(b) Local signal

(c) Local power spectrum

(d) Local power spectrum

(e) Averaged local power spectrum

(f) Averaged local power spectrum

Figure 5.12: Local signal evolution and noise power spectra for $WS(0, 10)$ (left column) and $WS(0.1, 10)$ (right column) network topologies (Lines are drawn to guide the eye)

Assessing Community Substructures

Assessing Community Substructures

Apart from algebraic connectivity, we have already mentioned that the emergence of synchrony in networks of coupled oscillators is influenced by the network's modular organization. Here, it has been shown that well-connected subgraphs, or communities, tend to synchronize quickly while connections between these communities result in increasingly large synchronized clusters and possibly global coherence. While it has already been argued that a global picture of synchronization dynamics can reveal modular and hierarchical structures, in the context of the proposed synchronization scheme an interesting related question is whether nodes can argue about such structural features by measuring quantities that are locally available. To address this issue we first need to be able to generate networks with pronounced modular substructures. Here we consider a maximally simple random graph model which can actually be seen as a two-stage version of the classical Erdös/Rényi model. In a first step, k communities $C_1, \ldots, C_k$ are created, each community C_i being a $G(s, m)$ Erdös/Rényi network with s nodes and m edges created uniformly at random. In a second step, m' additional edges are assigned to node pairs chosen uniformly at random irrespective of community memberships. Eventually, a network consisting of $k \cdot s$ nodes and $k \cdot m + m'$ edges is created. An example for a network generated by this model is shown in Figure 5.13. Here $k = 20$ communities were created with $m = 100$ links being assigned to each community consisting of $s = 20$ nodes and $m' = 100$ additional links being added randomly to the global network. This process eventually yielded a network consisting of 20 interconnected clusters, 400 nodes and 2100 edges. Due to the choice of parameters m and m', nodes within a community C_i are much more likely to be connected than nodes in different communities.

Modular Graph Model

Global Synchronization Dynamics in the Cluster Model

In the following, we study the dynamics of the proposed synchronization protocol in the particular network depicted in 5.13. In this experiment, $N(100, 25)$ distributed oscillator frequencies, a coupling probability of $p = 0.05$ and a degree-weighted coupling strength function f_1 as defined in equation 5.5 has been used. The evolution of the global order parameter $r(t)$ in a representative simulation run in the network depicted in Figure 5.13 is shown in Figure 5.14. Here, different phases can easily be identified. A first level of coherence with $r(t) \approx 0.2$ is quickly reached after roughly 250 time units. This can be attributed to communities C_i being quickly synchronized individually. In a subsequent phase, the exchange of coupling messages across different communities results in the formation of increasingly large synchronized clusters until, after roughly 3000 time steps, a globally coherent state emerges. This global

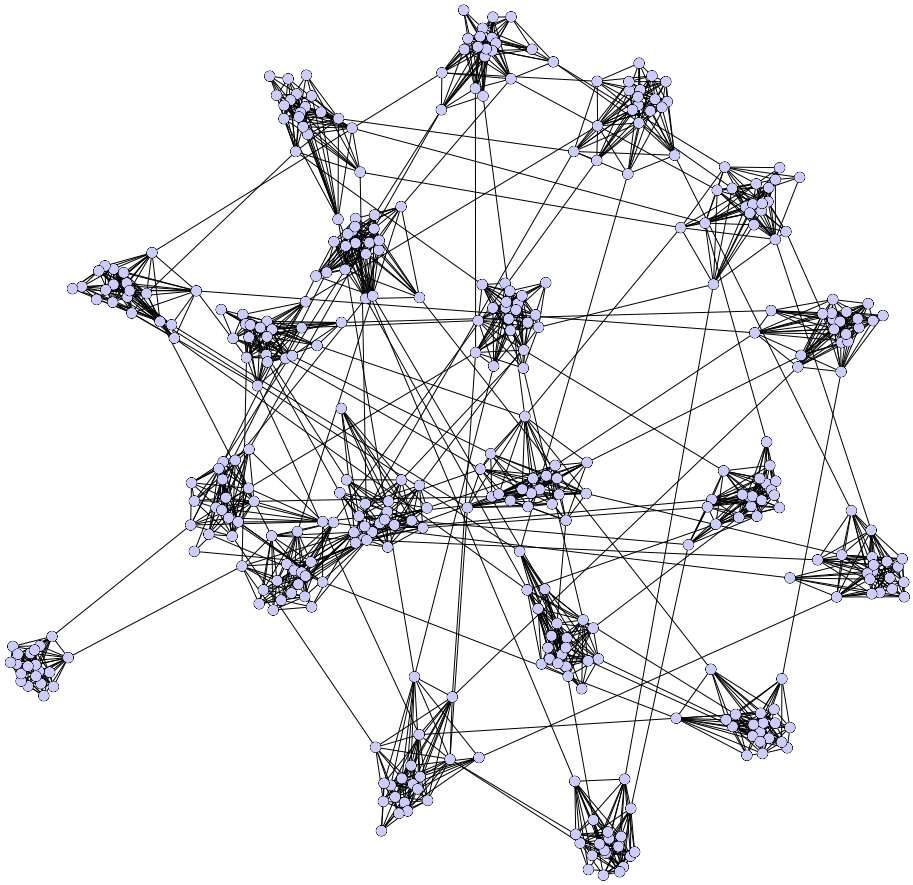


Figure 5.13: Example network with 20 communities, 400 nodes and 2100 edges

perspective is in accordance with earlier findings on the evolution of coherent states in modular networks of coupled oscillators.

Local Synchronization Dynamics

We now turn to a local picture of synchronization dynamics, taking the perspective of a node whose neighbors are members of different communities. For this, we will study the evolution of the angle

$$\delta(v, w) := sin(\Theta_v - \Theta_w)$$

between two oscillators residing in nodes v and w. Here we assume that both nodes v and w compute $\delta(v, w)$ whenever a coupling between them occurs based on the used coupling probability. Since synchronized regimes are likely to appear first in a network's most densely connected subgraphs, $\delta(v, w)$

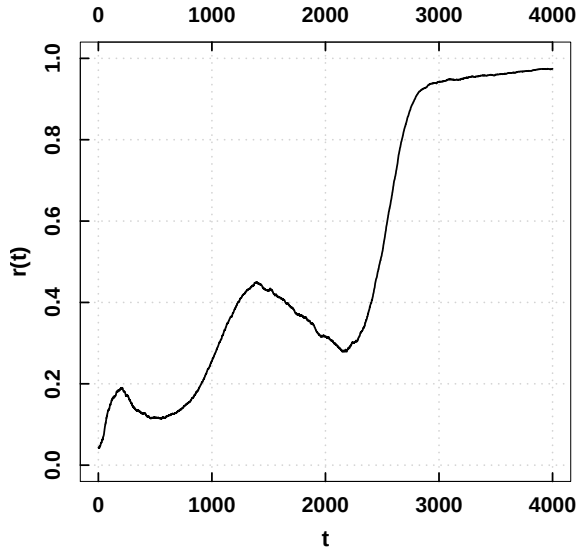
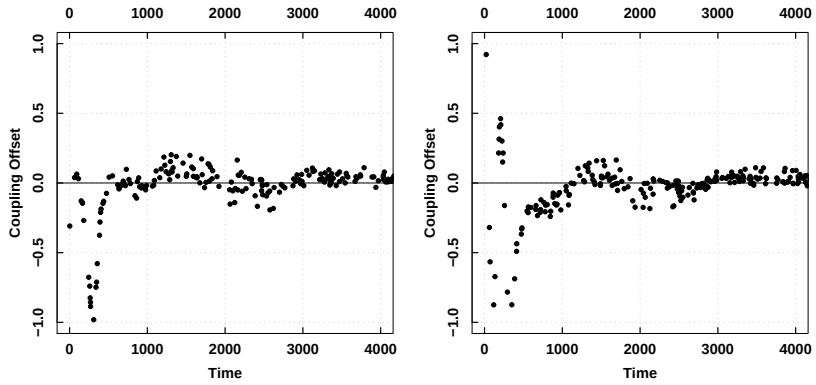


Figure 5.14: Evolution of order parameter in example network depicted in 5.13

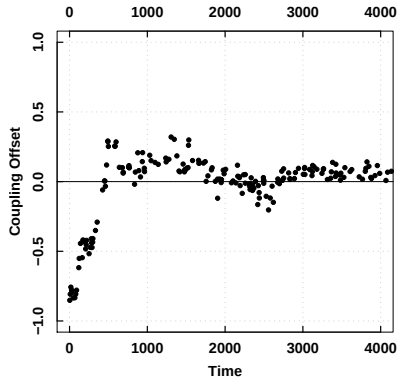
Evolution of Oscillator Angles

is expected to decrease quickly if the nodes v and w are members of the same, quickly synchronizing community. This suggests that - by monitoring the dynamics of incoming coupling offsets - a node can reason about the community memberships of its neighbors. To underpin this intuition, a particular node x has been chosen from the network depicted in Figure 5.13. In this network, x is member of a community C_7, maintaining links to nodes in the same community as well as to nodes that are members of three other communities C_9, C_{10} and C_{15}. Using the same coupling probability and coupling strength function f_1 as in the previously described experiment, $\delta(x, w)$ has then been recorded for each coupling event occurring in a simulation of the proposed synchronization scheme lasting for 4000 time steps. In Figure 5.15, the evolution of $\delta(x, w_i)$ is shown for node x and six neighbor nodes $w_1, \ldots, w_6$. A clear difference can be seen in the dynamics of couplings to different communities. In Figures 5.15(a) - 5.15(c), it can be seen that for nodes w_1, w_2 and w_3, which reside in the same community C_7 as node x, $\delta(x, w_i)$ quickly ceases, eventually fluctuating around zero. For nodes in other communities, the evolution of $\delta(x, w_i)$ shown in Figures 5.15(d) - 5.15(f) is significantly different, ceasing only after 3000 simulated time steps when global coherence is approached.

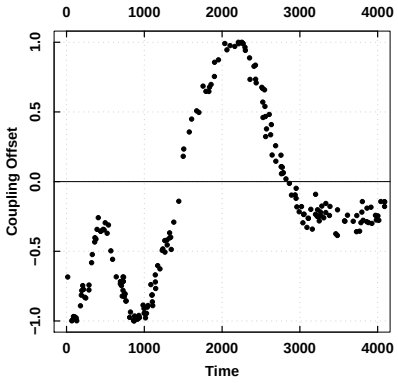


(a) w_1 in C_7

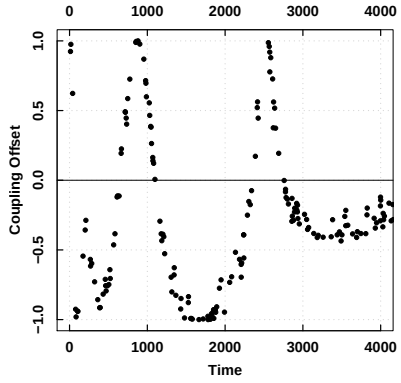
(b) w_2 in C_7



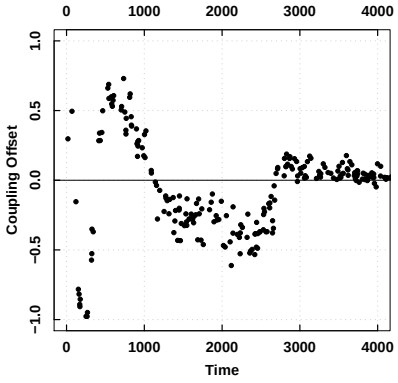
(c) w_3 in C_7



(d) w_4 in C_9



(e) w_5 in C_{10}

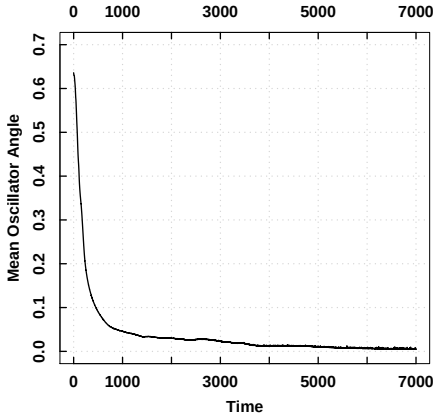


(f) w_6 in C_{15}

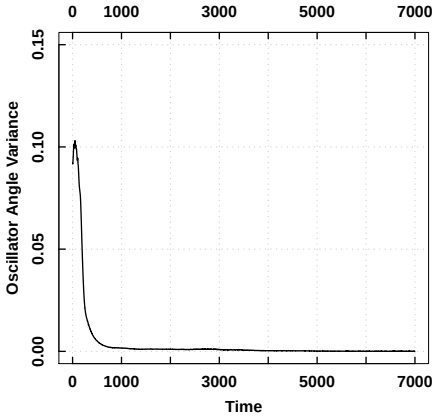
Figure 5.15: Evolution of exchanged coupling offsets $\delta(x, w_i)$ for six neighbors w_i of node x in community C_7

Evolution of Mean and Variance of Oscillator Angles

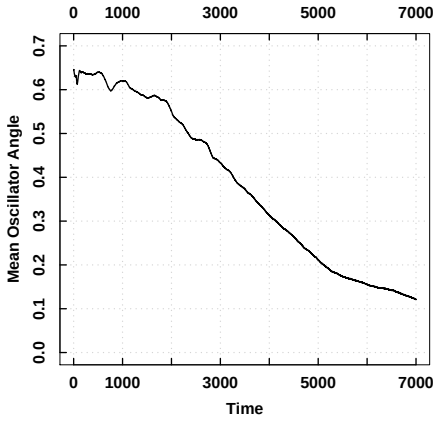
The results shown in Figure 5.15 have been obtained for one particular node chosen from the particular network realization depicted in Figure 5.13. Further experiments have thus been performed using ten different 1000 node networks generated according to the model presented above. In the network generation process, the parameters $k = 50$, $s = 20$, $m = 100$ and $m' = 100$ have been used. Based on the community memberships assigned to nodes during the construction of these networks, all edges were classified depending on whether they connect nodes in different communities or within the same community. For each edge (v, w), the evolution of $\delta(v, w)$ was recorded for all couplings taking place during a simulation of the synchronization scheme with coupling f_1 lasting for 7000 time steps. In Figure 5.16, the evolution of mean and variance of the absolute oscillator angle $|\delta(v, w)|$ is shown for edges connecting nodes in the same and in different communities. In Figure 5.16(a) and 5.16(b) one observes that for intra-community edges, both mean and variance of $|\delta(v, w)|$ quickly decays. Contrariwise, in Figures 5.16(c) and 5.16(d) mean and variance of $|\delta(v, w)|$ decrease slowly for edges connecting nodes in different communities. These results suggest, that nodes in a network running the proposed synchronization scheme can locally classify links by monitoring the evolution of angles to neighboring oscillators.



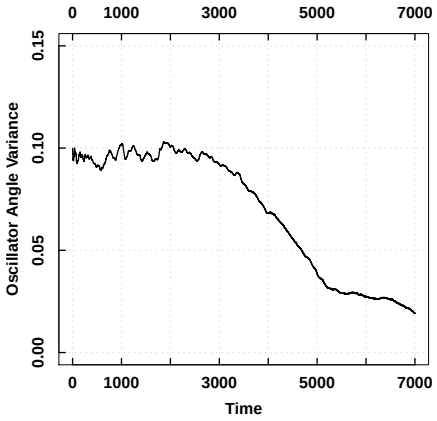
(a) Mean $|\delta(v,w)|$ for intra-community links



(b) Variance of $|\delta(v,w)|$ for intra-community links



(c) Mean $|\delta(v,w)|$ for inter-community links



(d) Variance of $|\delta(v,w)|$ for inter-community links

Figure 5.16: Evolution of mean and variance of absolute oscillator angles $|\delta(v,w)|$ in 1000 node networks with modular structure

5.2.4 Summary and Perspectives

Self-Organized Synchronization in Distributed Systems

In this chapter, we have studied a scheme for the self-organized synchronization of periodic processes that is inspired by the Kuramoto model for non-linearly coupled oscillators. Assuming a sporadic, probabilistic exchange of the oscillator's phase between nearest neighbors in networks, it can easily be implemented in terms of a distributed protocol. Compared to synchronization schemes that are based on pulse-coupled models, the randomized timing of the gossip-like coupling message exchanges can mitigate problems resulting from highly synchronous traffic. We have considered the impact of different coupling functions on the emergence of synchrony in networks with complex probabilistic structures. When using a degree-weighted coupling strength, simulations suggest that coherent oscillator states quickly emerge in networks with sufficiently large algebraic connectivity. The finding that a degree-weighted coupling strength quickly gives rise to coherent oscillator states in scale-free networks is in accordance with earlier studies of the Kuramoto model for networks with highly heterogeneous degree distributions. Considering the minimally required message exchange frequency, we observe a fast emergence of coherent states for coupling frequencies above the mean intrinsic oscillator frequency if the coupling strength is amplified proportionally. Below this coupling frequency, global coherence did not emerge even after very long times. We have further argued that the resilience of synchronized states against perturbations is particularly large in scale-free networks. Moreover, in scale-free networks both the synchronization time and the resilience against perturbations can be improved if the strength of couplings to high degree nodes are selectively amplified.

Open Issues and Threats to Validity

While these results suggest that the proposed synchronization scheme can be useful in a number of practical settings, there exist a number of open issues and threats to validity that have not yet been considered. An important issue that has not been addressed so far is how the functioning and performance of the scheme is influenced by inevitable communication latencies. Intuitively, one would expect the impact of latencies to be interrelated with the frequency of individual oscillators. In fact, analytical results presented for example in [Izhikevich, 1998; Papachristodoulou and Jadbabaie, 2005] suggest that transmission delays in the exchange of coupling message are likely to not affect the emergence of synchronization if the magnitude of delays is smaller than oscillator periods. However, further studies need to be done in order to evaluate the effect of realistic latency distributions on the proposed scheme. Apart from

disregarding message latencies, in this chapter a rather limited choice of network topologies, network sizes and intrinsic oscillator frequency distributions has been considered. In particular, a more substantial study of the interrelation between between a network's spectral properties and the dynamics and stability of stable synchronized states is needed before applying the proposed scheme in practical networked computing systems. Such a study is likely to benefit from the rapidly growing literature on the interplay between collective phenomena and complex network structures.

Monitoring Networks by Synchronization

Apart from being useful in the analytical study of synchronization processes, we have also argued that findings about the relation between the dynamics and stability of synchronization, spectral properties of Laplacian matrices and structural features of complex networks foreshadow promising further application scenarios. We have seen for example that *a node's local oscillatory signal can give clues about a network's algebraic connectivity*. Crucially influencing the functioning and performance of various dynamical processes and distributed algorithms, the ability to assess a network's algebraic connectivity in a distributed fashion can be valuable in a number of practical settings. It thus appears to be an interesting potential application of the scheme considered in this chapter. Apart from studying the aggregate effect of couplings on local oscillatory signals, we have further provided - admittedly preliminary and rough - experimental evidence that further *spectral properties of a network's Laplacian matrix leave their marks in the evolution of oscillator angles between a node and its neighbors*. Here, we have argued that the monitoring of local coupling dynamics can be used by nodes to extract valuable information about the community structure in its neighborhood. In particular, simulation results suggest that the *time evolution of angles to neighboring oscillators can be used to classify links into intra- and inter-community connections*.

Open Issues and Perspectives

A number of open issues and perspectives are associated with the proposition of using synchronization-based, distributed monitoring schemes. In this chapter, we have merely touched on the question to what extent the monitoring of a node's local oscillatory signal and the coupling dynamics to neighbors facilitates a distributed reasoning about a topology's spectral properties. In particular, results about the observed correlation between algebraic connectivity and noise in an oscillator's local spectrum are rather illustrative and more exhaustive empirical and/or analytical studies are required to substantiate them. Regarding the local detection of modular structures, it must be mentioned that the communities in the modular networks used in this chapter were rather pronounced, presumably resulting in easily distinguishable coupling dynam-

ics for intra- and inter-community links. A study of how well the scheme works in networks with less pronounced or overlapping communities is thus another open issue. It is also not yet clear, to what extent the time evolution of oscillator angles can serve as a fingerprint for particular communities, which would facilitate not only the discrimination of intra- and inter-community links but also a reasoning about community memberships of individual neighbors. A related question that has not yet been addressed is whether knowledge about the topology of network modules or their hierarchical organization can be derived from a node's local perspective on synchronization dynamics. Here, it would be interesting to investigate whether locally computable per-link metrics, like for instance the variance in the time evolution of oscillator angles or the time at which these angles start to cease, can be related to characteristics of the topology of inter-community connections. One may imagine a number of scenarios where the knowledge resulting from such a distributed synchronization-based monitoring could be useful. Referring to the importance of spectral properties for gossiping algorithms, information diffusion, distributed consensus and collective decision-making schemes, the performance of such schemes could possibly be improved if nodes knew about the community structure in their immediate neighborhood. Resembling phenomena apparently arising naturally in social systems [Granovetter, 1973], possible strategies may for example involve an adjustment of gossip-frequencies, coupling strengths or a suitable weighting of received information depending on the community to which a connection links. A further possible scenario where the scheme could be used is in the construction or adaptation of overlay networks. A reasonable rewiring of connections based on the synchronization dynamics could for instance be used to adapt a network's mixing properties or its modular substructures. A related idea of evolving network topologies based on their global synchronization dynamics has recently been considered in [Gorochowski *et al.*, 2010].

Applications of Synchronization-based Monitoring

Further Perspectives

Having considered some aspects of using a model for self-synchronization in the context of networked computing systems, we conclude by stressing that self-synchronization is merely one, particularly well-studied example for the kind of collective dynamics that can occur in systems of interconnected dynamical elements. Consequently, the aspects discussed here should be considered as a demonstration that foreshadows how recent findings about the interplay between the emergence of collective behavior and the topologies of complex networks can be applied in a computing context. Incorporating these findings in the development of network topologies and protocols is likely to facilitate the design of robust, adaptive and scalable distributed systems.

Chapter 6

Conclusion

> I don't know where my expertise is; my expertise is no disciplines. I would recommend to drop disciplinarity wherever one can.
>
> HEINZ VON FOERSTER

The Complexity Crisis

There is no doubt about the fact that we are living in a connected world, surrounded by increasingly large, dynamic and interwoven technical systems that pervade more and more aspects of our daily lives. But what can we say about the large scale properties of the Internet, global scale Peer-to-Peer systems or other large networked computing systems that are more and more becoming critical - in some cases even vital - technical infrastructures? How can we obtain sound and reliable statements about the structure and behavior of technical systems that are not controlled by a central organization, government or institution and that evolve in a distributed fashion, embedded into political, economic and social realities? How do favorable macroscopic network properties emerge and how do these relate to the behavior of individual machines, users and organizations? And how does a change of regulations, incentives or individual behavior affect the utility or performance of a system as a whole? Are we still in full control of our largest and most complex technical systems or are they taking on a life of their own, thus dooming us to study how they evolve and behave?

Mastering Complexity in Distributed Systems

Triggered by the availability of appropriate data sets, the study of complex structures and collective dynamics unfolding in distributed systems like the World Wide Web, Peer-to-Peer networks, power grids, social networks or the Internet has begun only recently. While the resulting findings are clearly important for the modeling and evaluation of existing systems, in the future we

will likely be required to go one step further and design systems that *systematically monitor, control and manage complex structures and collective dynamics*. In this dissertation, we have investigated different facets of the question how abstractions and results stemming from the field of complex sciences can help to reach this goal and in the following we summarize our main contributions and highlight some challenges and opportunities lying ahead.

6.1 Summary

Complexity in Engineered Systems - An Interdisciplinary Perspective

A major point of this dissertation being an *interdisciplinary perspective* on large scale distributed systems, we have reviewed findings and abstractions from complex systems science and pointed at possible applications in the context of distributed computing systems. This perspective is for the most part due to the last decade's surge of interest in the interaction topologies of complex social, biological, economic, physical and technical systems. Here, the simple yet powerful *network abstraction* has proven its value as a kind of *lingua franca* that has massively pollinated the interdisciplinary study of complex systems. However, considering the fields of *complex sciences* on the one hand and *computer science* on the other hand, differences in scientific culture as well as differences in the problems being addressed constitute a potential hurdle on the road towards a combined perspective on the engineering of networked systems. The study of complex systems - mostly from the perspective of statistical physics - typically aims at understanding the behavior of existing systems. For this, maximally simple models are usually employed that abstract away as many of a system's specifics as possible in order to investigate universal macroscopic phenomena that are independent of microscopic details. In contrast, practical computer science is a design science involved with the engineering of systems that are optimized with respect to a particular utility and that are subject to a multitude of technological, geographic and economic constraints. As has been argued in chapter 4 of this dissertation, these different perspectives have occasionally resulted in mutual misunderstanding although both are justified and necessary as our technical infrastructures grow larger. Connecting both perspectives to a complementary view on engineered networked systems is important and requires both research communities to tune in to each other more than hitherto. In particular, for the findings of the complex sciences community to be applicable to existing and highly optimized engineered systems, models need to be enriched with domain-specific information that is crucial for a system's functioning and performance. At the same time, engineers can benefit

from results about simple models for complex structures and behavior, for example by explicitly designing systems along such models.

Distributed Stochastic Management of Overlay Networks

Becoming increasingly important for the deployment of novel distributed services on top of existing infrastructures, we have argued that virtual overlay networks constitute a particularly promising domain for the application of findings from complex sciences. Here, the field of complex networks can help to establish distributed, stochastic mechanisms by which favorable, complex network structures emerge that are *neither completely random* in the sense of classical random graphs *nor completely deterministic* in the sense of highly structured overlays. Ranging between the commonly employed classification in structured and unstructured overlays, we have argued in chapter 3 that a design along abstractions stemming from the study of complex and random networks and statistical mechanics provides a number of interesting opportunities. In particular, the fact that network properties emerging at a macroscopic level can be related to local dynamical, stochastic processes like for example a certain random rewiring scheme or the behavior of users *facilitates a distributed stochastic management of overlay topologies and a bottom-up design of networked computing systems with distributed control*. Resembling the reasoning about equilibrium and non-equilibrium thermodynamic systems, a particularly appealing aspect of this approach is the possibility to derive reliable statements about large populations of dynamical, inherently stochastic elements. Rather than factoring out uncertainty by means of sophisticated corrective measures, *sensibly incorporating noise and randomness as a first class citizen is essential for the design of scalable, robust and predictable networked computing systems.*

Managing Random Scale-Free Overlays

In chapter 4, we have demonstrated some aspects of this approach in the context of random scale-free overlay networks. We have presented a distributed overlay construction and maintenance protocol that generates topologies which are drawn from a parameterizable statistical ensemble of random scale-free networks. Furthermore, here we have used the fact that - other than in systems being usually studied in natural sciences - the stochastic behavior of machines in distributed computing systems can often be influenced in a targeted fashion. In the proposed rewiring scheme, by a meaningful change of a random walk bias it is possible to actively trigger a transition between phases with bounded and unbounded node degree fluctuations. The abrupt change of macroscopic properties taking place at this critical point in the protocol's parameter space can be seen in analogy to phase transition phenomena occurring in thermodynamic systems. Finally, we have presented a scheme suitable for a distributed measurement of the degree distribution exponent in

networks with power law degree distributions. Here, we have argued that the real-time measurement of such statistical ensemble parameters by means of probabilistic aggregation schemes can help to derive reliable statements about macroscopic network properties. In general, *a shift towards simple, stochastic mechanisms that give rise to network topologies with favorable and predictable emergent properties seems to be an important cornerstone of mastering complexity in a number of sufficiently large distributed systems.*

Collective Dynamics in Distributed Computing Systems

Apart from the formation of complex network structures, a further and more general class of self-organization phenomena being studied in complex sciences and statistical physics is the formation of coherent spatio-temporal patterns from local, possibly noisy interactions between dynamical elements. In distributed computing systems, the occurrence of such coherent patterns can be both threat and opportunity. Since roughly a decade, mathematical, physical and biological models for such collective phenomena are increasingly being studied in the framework of complex networks, thus significantly furthering their relevance for distributed computing systems. We have seen that models for collective behavior in networks can serve as an inspirational framework for the design of distributed algorithms giving rise to particular types of macroscopically coherent behavior like distributed consensus, synchronization or decentralized decision-making. Complementing the interdisciplinary perspective on the stochastic, distributed management of networks with complex structures, results about the relation between the topology and the dynamics of a network are particularly interesting. A knowledge about these relations is crucial both in the network construction process and in the design of distributed schemes making use of collective phenomena.

Self-Organized Synchronization in Networks

In chapter 5 we have considered a distributed scheme for the self-organized synchronization of periodic processes in large scale networks that has been inspired by the Kuramoto model for non-linearly coupled oscillators. We have shown that a single message exchange on average per neighbor and oscillator period is seemingly sufficient to give rise to coherent oscillator states. For scale-free networks, we have provided evidence that a weighting of coupling strength proportional to the degree of the coupling partner can increase both the speed and the stability of the synchronization process. We have finally commented on the potential application of the scheme for a distributed monitoring of certain spectral properties of a network's Laplacian.

6.2 Perspectives

Broader Perspectives

We have argued that the study of complex systems bears a number of interesting aspects that can be useful in the design of resilient, predictable and efficient distributed computing systems. However, the choice of topics that have been discussed in more detail in this dissertation is necessarily rather limited. In the following, we thus highlight some broader perspectives which - at least from the author's point of view - deserve to become the subject of more in-depth future studies.

- We have seen that for a number of basic network qualities like for instance algebraic connectivity, relations to aggregate parameters of statistical ensembles of networks can be established. In systems where one wants to employ dynamical processes whose functioning or performance relies on such basic properties, the use of simple, stochastic connection rules giving rise to overlay topologies drawn from statistical ensembles with suitable parameters is a reasonable approach. In this dissertation we have commented on the use of probabilistic distributed search, random walk and synchronization schemes as well as gossip-based algorithms in probabilistically structured networks. In order to broaden the range of applicability for such a stochastic topology management, network classes and distributed schemes need to be investigated that facilitate the efficient completion of further algorithmic tasks.

- An interesting aspect that is a direct result of the statistical mechanics' perspective on complex networks is the possibility to obtain strong stochastic guarantees about a system's large scale properties from a model that incorporates randomness and uncertainty at the level of its constituents. With resilience and dependability becoming an increasingly important focus of distributed systems research, the use of statistical methods to make sound statements about the qualities of networks emerging from stochastic processes is likely to gain importance. The methods being used in the study of complex networks seem to be well-suited for this task.

- In this dissertation, models for collective behavior in networks and distributed schemes by which these networks emerge have been considered as rather separate issues. However both can actually complement each other. A combination of distributed, stochastic topology construction and adaptation protocols like the one presented in chapter 4 with models for

collective consensus or decision-making being studied in complex sciences constitutes a promising approach for the design of systems with decentralized control.

- It further seems rather natural that network structures and collective behavior mutually influence each other. In social systems, the network of contacts clearly influences an individual's views and opinions. At the same time, these views and opinions are likely to influence to whom connections are being created or maintained. Obtaining a better understanding of the co-evolution of collective behavior and complex networks in real-world systems is thus an important and current research theme. The results of this research are likely to have influence also on the design of computing systems with distributed control and variable network topologies.

- In the context of distributed systems with decentralized governance, a further interesting question is how control strategies can be derived that neither are completely egalitarian nor make use of designated leaders. In most practical networked computing systems, the reliability, trustworthiness or the degree of knowledge of participating machines or users is rather heterogeneous. At the same time, the use of designated leaders or fixed control hierarchies is usually not desirable. Here, the growing literature on the interplay between collective behavior, heterogeneous node characteristics and the structure of interaction topologies promises to facilitate the development of robust decentralized control schemes for distributed computing systems.

- A particularly promising aspect of the interdisciplinary field of complex sciences is the fact that the abstraction of technical, social, economic and biological systems in terms of complex networks opens up the possibility to study facets of all these systems within a unified theoretical framework. Considering the importance of the interplay between economic, social, political and technological aspects for large, evolving infrastructures like the Internet, this combined perspective can play an important role for the derivation of policies, regulation and incentives which eventually lead to systems with favorable properties. For this, existing simple models giving rise to networks whose statistical properties resemble those of technical infrastructures need to be adjusted and enriched with domain-specific aspects.

- A promising perspective of studying both biological and technical systems in the same conceptual framework is that it potentially furthers the design of bio-inspired computing systems. Here, models that capture the characteristics of self-organizing and adaptive biological systems can serve as inspiration for the design of distributed systems that inherit some of the favorable properties of natural systems. As such, the study of complex networks and complex systems constitutes an important cornerstone for the engineering of *organic computing systems*.

- The possibility to model the structure and behavior of social systems can further be important for the operation of technical systems whose functioning and performance critically depends on human behavior. Peer-to-Peer systems, content distribution networks or online games may for instance benefit from knowledge about the topology of collaborations and communication as well as about the dynamics of opinion formation and information spreading processes. As computing systems are becoming increasingly aware of the social networks into which their users are embedded, the use of such information in the management of distributed computing systems is likely to gain importance.

Final Note

In the light of these perspectives, the results presented in this dissertation must necessarily appear preliminary. However, our intention was to point out that augmenting the views of distributed systems engineers with those of the complex systems community bears promising prospects for the design of networked computing systems. The field of complex systems is disclosing some of the principles that govern the structure and dynamics of biological, physical and social systems. As engineers we should embrace the simplicity and elegance of these principles. Harnessing them bears the potential to transfer some of the remarkable characteristics of natural systems to our largest networked infrastructures, thus making them more manageable, resilient and adaptive.

List of Figures

List of Algorithms

List of Equations

Author's Publications

[Botev *et al.*, 2008a] Jean Botev, Markus Esch, Alexander Höhfeld, Hermann Schloss, and Ingo Scholtes. The HyperVerse - Concepts for a Federated and Torrent Based 3D Web. In *Proceedings of the First International Workshop on Massively Multiuser Virtual Environments (MMVE 2008) at IEEE Virtual Reality 2008, Reno, Nevada, USA*. IEEE, March 2008.

[Botev *et al.*, 2008b] Jean Botev, Markus Esch, Alexander Höhfeld, Hermann Schloss, Ingo Scholtes, and Peter Sturm. HyperVerse - Towards a Self-Organizing and Sustainable Global-Scale Virtual Environment. In *In Proceedings of the Fourth Microsoft Academic Days, Berlin, Germany*. Microsoft Press, November 2008.

[Botev *et al.*, 2008c] Jean Botev, Alexander Höhfeld, Hermann Schloss, Ingo Scholtes, Peter Sturm, and Markus Esch. The HyperVerse - Concepts for a Federated and Torrent Based 3D Web. *International Journal of Advanced Media and Communication*, 2(4):331–350, December 2008.

[Botev *et al.*, 2009] Jean Botev, Markus Esch, Hermann Schloss, Ingo Scholtes, and Peter Sturm. HyperVerse: Simulation and Testbed Reconciled. In *In Proceedings of the Second International Workshop on Massively Multiuser Virtual Environments (MMVE2009) at IEEE Virtual Reality 2009, LaFayette, Louisiana, USA*. IEEE, March 2009.

[Kolos and Scholtes, 2005] Serguei Kolos and Ingo Scholtes. Event monitoring design. Technical Report ATL-DQ-EN-0023 v.2, CERN, https://edms.cern.ch/document/639958/2, September 2005.

[Scholtes and Sturm, 2006] Ingo Scholtes and Peter Sturm. Web Federates - Towards A Middleare for Highly Scalable Peer-To-Peer Services. In *Proceedings*

of the Second International Conference on Web Information Systems and Technologies (WEBIST), Setúbal, Portugal. INSTICC Press, April 2006.

[Scholtes *et al.*, 2006] Ingo Scholtes, Serguei Kolos, and Pasquale Federico Zema. The ATLAS Event Monitoring Service - Peer-to-Peer Data Distribution in High Energy Physics. In *In Proceedings of the 2006 Nuclear Science Symposium, San Diego, California, USA*. IEEE, November 2006.

[Scholtes *et al.*, 2008a] Ingo Scholtes, Jean Botev, Markus Esch, Alexander Hoehfeld, and Hermann Schloss. Awareness-Driven Phase Transitions in Very Large Scale Distributed Systems. In *Proceedings of the Second IEEE International Conferences on Self-Adaptive and Self-Organizing Systems (SaSo), Venice, Italy*. IEEE, October 2008.

[Scholtes *et al.*, 2008b] Ingo Scholtes, Jean Botev, Markus Esch, Alexander Hoehfeld, Hermann Schloss, and Benjamin Zech. TopGen - Internet Router-Level Topology Generation Based on Technology Constraints. In *Proceedings of the First International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SIMUTools), Marseille, France*. IEEE, February 2008.

[Scholtes *et al.*, 2008c] Ingo Scholtes, Serguei Kolos, and Pasquale Federico Zema. The ATLAS Event Monitoring Service - Peer-to-Peer Data Distribution in High-Energy Physics. *IEEE Transactions on Nuclear Science*, 55(3):1610–1620, June 2008.

[Scholtes *et al.*, 2009] Ingo Scholtes, Jean Botev, Markus Esch, and Peter Sturm. Epidemic Self-Synchronization in Complex Networks. In *Proceedings of the First International Conference, Complex 2009, Shanghai, China, February 23-25, 2009, Revised Papers*, volume 5, pages 1794–1809. Springer Berlin/Heidelberg, February 2009.

[Scholtes *et al.*, 2010] Ingo Scholtes, Jean Botev, Markus Esch, and Peter Sturm. Epidemic Self-Synchronization in Complex Networks of Kuramoto Oscillators. *Advances in Complex Systems*, 13(1):33–58, June 2010.

[Scholtes, 2005] Ingo Scholtes. A Reimplementation Of The CORBA-based Event Monitoring System For The ATLAS LHC Experiment At CERN. Diploma Thesis, September 2005.

[Scholtes, 2010] Ingo Scholtes. Distributed Creation and Adaptation of Random Scale-Free Overlay Topologies. In *Proceedings of the Fourth IEEE Confer-*

ence on Self-Organizing and Self-Adaptive Systems (SASO), Budapest, Hungary. IEEE, October 2010.

[The community of testers and developers in ATLAS DAQ/HLT, 2005] The community of testers and developers in ATLAS DAQ/HLT. ATLAS DAQ/HLT Software Large Scale Functionality and Performance Tests July 2005. Technical Report ATL-D-TR-0003 v.1, CERN, https://edms.cern.ch/document/685256/1, November 2005.

Bibliography

[Aberer, 2001] Karl Aberer. P-grid: A self-organizing access structure for p2p information systems. *Sixth International Conference on Cooperative Information Systems (CoopIS 2001), Lecture Notes in Computer Science*, 2172:179–194, 2001.

[Acebrón *et al.*, 2005] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler. The kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77:137–185, January 2005.

[Achlioptas *et al.*, 2005] Dimitris Achlioptas, Aaron Clauset, David Kempe, and Cristopher Moore. On the bias of traceroute sampling: or, power-law degree distributions in regular graphs. In *STOC '05: Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 694–703, New York, NY, USA, 2005. ACM.

[Adamic *et al.*, 2001] Lada A. Adamic, Rajan M. Lukose, Amit R. Puniyani, and Bernardo A. Huberman. Search in power-law networks. *Physical review. E, Statistical, nonlinear, and soft matter physics .*, 64, 2001.

[Aiello *et al.*, 2000] William Aiello, Fan Chung, and Linyuan Lu. A random graph model for massive graphs. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 171–180, New York, NY, USA, 2000. ACM Press.

[Albert and Barabási, 2002] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 72(1):48–97, 2002.

[Albert *et al.*, 1999] Réka Albert, Hawoong Jeong, and Albert-László Barabási. The diameter of the world wide web. *Nature*, 401:130, 1999.

[Albert *et al.*, 2000] Réka Albert, Hawoong Jeong, and Albert-László Barabási. The internet's achilles heel: Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[Alderson and Doyle, 2010] David L. Alderson and John C. Doyle. Contrasting views of complexity and their implications for network-centric infrastructures. *Trans. Sys. Man Cyber. Part A*, 40(4):839–852, 2010.

[Alderson *et al.*, 2005] David Alderson, Lun Li, Walter Willinger, and John C. Doyle. Understanding internet topology: principles, models, and validation. *IEEE/ACM Trans. Netw.*, 13(6):1205–1218, 2005.

[Alderson, 2008] David L. Alderson. Catching the "network science" bug: Insight and opportunity for the operations researcher. *Operations Research*, 56:1047–1065, 2008.

[Amaral *et al.*, 2000] L.A.N. Amaral, A. Scala, M. Barthelemy, and Heugene E. Stanley. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97:11149–11152, 2000.

[Arenas and Díaz-Guilera, 2007] A. Arenas and A. Díaz-Guilera. Synchronization and modularity in complex networks. *The European Physical Journal - Special Topics*, 143(1):19–25, April 2007.

[Arenas *et al.*, 2006a] A. Arenas, A. Díaz-Guilera, and C. J. Pérez-Vicente. Synchronization Reveals Topological Scales in Complex Networks. *Physical Review Letters*, 96(11):114102–+, March 2006.

[Arenas *et al.*, 2006b] Alex Arenas, Albert Díaz-Guilera, and Conrad J. Pérez-Vicente. Synchronization processes in complex networks. *Physica D: Nonlinear Phenomena*, 224:27–34, December 2006.

[Azar *et al.*, 1992] Yossi Azar, Andrei Z. Broder, Anna R. Karlin, Nathan Linial, and Steven Phillips. Biased random walks. In *STOC '92: Proceedings of the twenty-fourth annual ACM symposium on Theory of computing*, pages 1–9, New York, NY, USA, 1992. ACM.

[Babaoglu *et al.*, 2007] Ozalp Babaoglu, Toni Binci, Mark Jelasity, and Alberto Montresor. Firefly-inspired heartbeat synchronization in overlay networks. In *SASO '07: Proceedings of the First International Conference on Self-Adaptive and Self-Organizing Systems*, pages 77–86, Washington, DC, USA, 2007. IEEE Computer Society.

[Baiesi and Manna, 2003] M. Baiesi and S. S. Manna. Scale-free networks from a hamiltonian dynamics. *Physical Review E*, 68, 2003.

[Bak *et al.*, 1987] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: An explanation of the 1/f noise. *Phys. Rev. Lett.*, 59(4):381–384, Jul 1987.

[Bakker *et al.*, 2009] Arno Bakker, Elth Ogston, and Maarten van Steen. Collaborative filtering using random neighbours in peer-to-peer networks. In *CNIKM '09: Proceeding of the 1st ACM international workshop on Complex networks meet information & knowledge management*, pages 67–75, New York, NY, USA, 2009. ACM.

[Balakrishnan *et al.*, 2003] Hari Balakrishnan, M. Frans Kaashoek, David Karger, Robert Morris, and Ion Stoica. Looking up data in p2p systems. *Commun. ACM*, 46(2):43–48, 2003.

[Baldoni *et al.*, 2009] Roberto Baldoni, Angelo Corsaro, Leonardo Querzoni, Sirio Scipioni, and Sara Tucci Piergiovanni. Coupling-based internal clock synchronization for large-scale dynamic distributed systems. *IEEE Transactions on Parallel and Distributed Systems*, 99(RapidPosts):607–619, July 2009.

[Barabási and Albert, 1999] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[Barabási *et al.*, 1999] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A Statistical Mechanics and its Applications*, 272:173–187, 1999.

[Barabási *et al.*, 2001] Albert-László Barabási, Erzsébet Ravasz, and Tamás Vicsek. Deterministic scale-free networks. *Physica A Statistical Mechanics and its Applications*, 299:559–564, October 2001.

[Barahona and Pecora, 2002] Mauricio Barahona and Louis M. Pecora. Synchronization in small-world systems. *Phys Rev Lett.*, 89(5), July 2002.

[Barrat and Weigt, 2000] Alain Barrat and M. Weigt. On the properties of small-world networks. *European Physics Journal B*, 13:547–560, 2000.

[Barrat *et al.*, 2008] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[Bauke, 2007] Heiko Bauke. Parameter estimation for power-law distributions by maximum likelihood methods. *The European Physical Journal B*, 58(2):167–173, 2007.

[Berg and Lassig, 2002] Johannes Berg and Michael Lassig. Correlated random networks. *Physical Review Letters*, 89:228701, 2002.

[Bianconi, 2009] Ginestra Bianconi. The entropy of network ensembles. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 79(3):036114, 2009.

[Blanchard *et al.*, 2002] Ph Blanchard, Ch H. Chang, and T. Krueger. Epidemic thresholds on scale-free graphs: the interplay between exponent and preferential choice. *Annales Henri Poincare*, 4:957–970, 2002.

[Boccaletti *et al.*, 2002] S. Boccaletti, J. Kurths, G. Osipov, D. L. Valladares, and C. S. Zhou. The synchronization of chaotic systems. *Physics Reports*, 366:1–101, August 2002.

[Boccaletti *et al.*, 2006] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175 – 308, 2006.

[Bollobás *et al.*, 2001] Béla Bollobás, Oliver Riordan, Joel Spencer, and Gábor Tusnády. The degree sequence of a scale-free random graph process. *Random Struct. Algorithms*, 18(3):279–290, 2001.

[Bollobás, 1998] Béla Bollobás. *Modern Graph Theory*. Springer, 1998.

[Bollobás, 2001] Béla Bollobás. *Random Graphs*. Cambridge University Press, 2001.

[Boltzmann, 1896] Ludwig Boltzmann. *Vorlesungen über Gastheorie*. Johann Ambrosius Barth, Leipzig, 1896.

[Boyd *et al.*, 2006] Stephen Boyd, Arpita Ghosh, Balaji Prabhakar, and Devavrat Shah. Randomized gossip algorithms. *IEEE/ACM Trans. Netw.*, 14(SI):2508–2530, 2006.

[Boykin *et al.*, 2004] P.O. Boykin, V.P. Roychowdhury, and N. Sarshar. Percolation search in power law networks: making unstructured peer-to-peer networks scalable. *Proceedings of the Fourth International Conference on Peer-to-Peer Computing*, 2004.

[Braden *et al.*, 1998] B. Braden, D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Partridge, L. Peterson, K. Ramakrishnan, S. Shenker, J. Wroclawski, and L. Zhang. Recommendations on queue management and congestion avoidance in the internet. RFC2309, 1998.

[Britton and Adler, 2004] Nicholas F. Britton and Fred Adler. Essential mathematical biology. *Physics Today*, 2004.

[Broder *et al.*, 2000] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. Graph structure in the web. *Comput. Netw.*, 33(1-6):309–320, 2000.

[Buldyrev *et al.*, 2010] Sergey V. Buldyrev, Roni Parshani, Gerald Paul, H. Eugene Stanley, and Shlomo Havlin. Catastrophic cascade of failures in interdependent networks. *Nature*, 464:1025–1028, 2010.

[Bustamante and Qiao, 2008] Fabián E. Bustamante and Yi Qiao. Designing less-structured p2p systems for the expected high churn. *IEEE/ACM Trans. Netw.*, 16(3):617–627, 2008.

[Caldarelli *et al.*, 2002a] G. Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz. Scale-free networks without growth or preferential attachment: Good get richer. *ArXiv Condensed Matter e-prints*, July 2002.

[Caldarelli *et al.*, 2002b] Guido Caldarelli, A. Capocci, P. De Los Rios, and M. A. Munoz. Scale-free networks from varying vertex intrinsic fitness. *Physical Review Letters*, 89(25), December 2002.

[Caldarelli, 2007] Guido Caldarelli. *Scalefree Networks*. Oxford University Press, 2007.

[Capocci *et al.*, 2005] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori. Detecting communities in large networks. *Physica A Statistical Mechanics and its Applications*, 352:669–676, July 2005.

[Carlson and Doyle, 1999] J. M. Carlson and John Doyle. Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review Letters E*, 60(2):1412–1427, August 1999.

[Carreras *et al.*, 2002] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman. Dynamics, criticality and self-organization in a model for blackouts in power transmission systems. In *Proceedings of the Hawaii International Conference on Systems Sciences*, 2002.

[Castellano *et al.*, 2009] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81(2):591–646, May 2009.

[Chawathe *et al.*, 2003] Yatin Chawathe, Sylvia Ratnasamy, Lee Breslau, Nick Lanham, and Scott Shenker. Making Gnutella-like P2P systems scalable. In *SIGCOMM '03: Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 407–418, New York, NY, USA, 2003. ACM Press.

[Chung and Lu, 2002] Fan Chung and Linyuan Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6:125–145, 2002. 10.1007/PL00012580.

[Chung, 1997] Fan R. K. Chung. *Spectral Graph Theory*. American Mathematical Society, 1997.

[Clauset *et al.*, 2007] Aaron Clauset, Cosma Rohilla Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51:661, 2007.

[Clausius, 1864] Rudolf Clausius. *Abhandlungen über die mechanische Wärmetheorie*. Friedrich Vieweg und Sohn, 1864.

[Cohen and Havlin, 2003] Reuven Cohen and Shlomo Havlin. Scale-free networks are ultrasmall. *Phys. Rev. Lett.*, 90(5):058701, Feb 2003.

[Cohen *et al.*, 2000] Reuven Cohen, Keren Erez, Daniel ben Avraham, and Shlomo Havlin. Resilience of the internet to random breakdowns. *Phys. Rev. Lett.*, 85:4626–4628, 2000.

[Cohen *et al.*, 2001] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Physical Review Letters*, 86, 2001.

[Cohen, 1988] Joel E. Cohen. Threshold phenomena in random structures. *Discrete Applied Mathematics*, 19:113–128, 1988.

[Cohen, 2003] Bram Cohen. Incentives build robustness in bittorrent. In *Proceedings of the Workshop on Economics of Peer-to-Peer Systems*, June 2003.

[Dallard *et al.*, 2001] P. Dallard, A. J. Fitzgerald, A. Flint, S. Le Bourva, A. Low, R. M. Ridsdill Smith, and M. Wilford. The london millennium footbridge. *The Structural Engineer*, 79(22):17–33, November 2001.

[Demers *et al.*, 1987] Alan Demers, Dan Greene, Carl Hauser, Wes Irish, John Larson, Scott Shenker, Howard Sturgis, Dan Swinehart, and Doug Terry. Epidemic algorithms for replicated database maintenance. In *PODC '87: Proceedings of the sixth annual ACM Symposium on Principles of distributed computing*, pages 1–12, New York, NY, USA, 1987. ACM.

[Diaconis and Stroock, 1991] P. Diaconis and D. Stroock. Geometric bounds for eigenvalues of markov chains. *Ann. Appl. Prob.*, 1:36–61, 1991.

[Donetti and Munoz, 2004] L. Donetti and M. A. Munoz. Detecting network communities: a new systematic and efficient algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 10:12, October 2004.

[Dorogovtsev and Mendes, 2000] Sergey N. Dorogovtsev and José Fernando Ferreira Mendes. Evolution of networks with aging of sites. *Phys. Rev. E*, 62(2):1842–1845, Aug 2000.

[Dorogovtsev and Mendes, 2003] Sergey N. Dorogovtsev and José Fernando Ferreira Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, 2003.

[Dorogovtsev *et al.*, 2000a] Sergey N. Dorogovtsev, José Fernando Ferreira Mendes, and Alexander N. Samukhin. Growing network with heritable connectivity of nodes. *ArXiv Condensed Matter e-prints*, November 2000.

[Dorogovtsev *et al.*, 2000b] Sergey N. Dorogovtsev, José Fernando Ferreira Mendes, and Alexander N. Samukhin. Structure of growing networks: Exact solution of the barabasi–albert's model. *Physical Review Letters*, 85:4633, 2000.

[Dorogovtsev *et al.*, 2003] Sergey N. Dorogovtsev, José Fernando Ferreira Mendes, and Alexander N. Samukhin. Principles of statistical mechanics of random networks. *Nuclear Physics B*, 666:396, 2003.

[Doyle *et al.*, 2005] John C. Doyle, David L. Alderson, Lun Li, Steven Low, Matthew Roughan, Stanislav Shalunov, Reiko Tanaka, and Walter Willinger. The "robust yet fragile" nature of the internet. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41):14497–14502, October 2005.

[Erdös and Rényi, 1959] Pál Erdös and Alfred Rényi. On random graphs. *Publ. Math. Debrecen.*, 6:290–297, 1959.

[Ermentrout and Kopell, 1984] George Bard Ermentrout and Nancy Kopell. Frequency plateaus in a chain of weakly coupled oscillators. *SIAM Journal on Mathematical Analysis*, 15(2):215–237, 1984.

[Ermentrout, 1985] Bard Ermentrout. Synchronization in a pool of mutually coupled oscillators with random frequencies. *Journal of Mathematical Biology*, 22(1):1–9, June 1985.

[Ermentrout, 1991] Bard Ermentrout. An adaptive model for synchrony in the firefly pteroptyx malaccae. *Journal of Mathematical Biology*, 29(6):571–585, June 1991.

[Fabrikant *et al.*, 2002] Alex Fabrikant, Elias Koutsoupias, and Christos H. Papadimitriou. Heuristically optimized trade-offs: A new paradigm for power laws in the internet. In *ICALP '02: Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, pages 110–122, London, UK, 2002. Springer-Verlag.

[Faloutsos *et al.*, 1999] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM '99: Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262, New York, NY, USA, 1999. ACM Press.

[Farkas *et al.*, 2004] Illés Farkas, Imre Derényi, Gergely Palla, and Tamás Vicsek. Equilibrium statistical mechanics of network structures. *LECT.NOTES PHYS.*, 650:163, 2004.

[Ferretti and D'Angelo, 2010] Stefano Ferretti and Gabriele D'Angelo. Multiplayer online games over scale-free networks: a viable solution? In *Proceedings of the First Workshop on Distributed Simulation and Online Gaming (DISIO)*. ICT/CREATE-NET, March 2010.

[Floyd and Jacobson, 1994] Sally Floyd and Van Jacobson. The synchronization of periodic routing messages. *IEEE/ACM Transactions on Networking*, 2:122–136, 1994.

[Fraigniaud *et al.*, 2005] Pierre Fraigniaud, Philippe Gauron, and Matthieu Latapy. Combining the use of clustering and scale-free nature of user exchanges into a simple and efficient p2p system. In *Proceedings of Euro-Par 2005, Parallel Processing, 11th International Euro-Par Conference, Lisbon, Portugal, August 30 - September 2, 2005*, pages 1163–1172, 2005.

[Gama-Castro *et al.*, 2008] Socorro Gama-Castro, Veronica Jimenez Jacinto, Martin Peralta-Gil, Alberto Santos-Zavaleta, Monica I. Peñaloza-Spindola, Bruno Contreras-Moreira, Juan Segura-Salazar, Luis Muñiz-Rascado, Irma Martinez-Flores, Heladia Salgado, Cesar Bonavides-Martinez, Cei Abreu-Goodger, Carlos Rodriguez-Penagos, Juan Miranda-Rios, Enrique Morett, Enrique Merino, Araceli M.Huerta, and Julio Collado-Vides. Regulondb (version 6.0): gene regulation model of escherichia coli k-12 beyond transcription, active (experimental) annotated promoters and textpresso navigation. *Nucleic Acids Research, Database issue:D120-4*, 36, 2008.

[Ganesh *et al.*, 2003] Ayalvadi J. Ganesh, Anne-Marie Kermarrec, and Laurent Massoulié. Peer-to-peer membership management for gossip-based protocols. *IEEE Trans. Comput.*, 52(2):139–149, 2003.

[Garlaschelli *et al.*, 2006] Diego Garlaschelli, Sebastian E. Ahnert, Thomas M. A. Fink, and Guido Caldarelli. Temperature in complex networks, 2006.

[Gkantsidis *et al.*, 2003] Christos Gkantsidis, Milena Mihail, and Amin Saberi. Conductance and congestion in power law graphs. *SIGMETRICS Perform. Eval. Rev.*, 31(1):148–159, 2003.

[Gkantsidis *et al.*, 2006] Christos Gkantsidis, Milena Mihail, and Amin Saberi. Random walks in peer-to-peer networks: algorithms and evaluation. *Perform. Eval.*, 63(3):241–263, 2006.

[Goh *et al.*, 2001] K. I. Goh, B. Kahng, and D. Kim. Universal behavior of load distribution in scale-free networks. *Physical Review Letters*, 87:278701, 2001.

[Gorochowski *et al.*, 2010] Thomas E. Gorochowski, Mario di Bernardo, and Claire S. Grierson. Evolving enhanced topologies for the synchronization of dynamical complex networks. *Phys. Rev. E*, 81(5):056212, May 2010.

[Govindan and Tangmunarunkit, 2000] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery. In *Proceedings of the Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. IEEE INFOCOM 2000*, pages 1371 –1380. IEEE, March 2000.

[Granovetter, 1973] Mark S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360–1380, May 1973.

[Gribble, 2001] Steven D. Gribble. Robustness in complex systems. In *HOTOS '01: Proceedings of the Eighth Workshop on Hot Topics in Operating Systems*, page 21, Washington, DC, USA, 2001. IEEE Computer Society.

[Guclu and Yuksel, 2009] Hasan Guclu and Murat Yuksel. Limited scale-free overlay topologies for unstructured peer-to-peer networks. *IEEE Trans. Parallel Distrib. Syst.*, 20(5):667–679, 2009.

[Gupta *et al.*, 2006] Indranil Gupta, Anne-Marie Kermarrec, and Ayalvadi J. Ganesh. Efficient and adaptive epidemic-style protocols for reliable and scalable multicast. *IEEE Trans. Parallel Distrib. Syst.*, 17(7):593–605, 2006.

[Gutenberg and Richter, 1949] B. Gutenberg and C. F. Richter. *Seismicity Of The Earth And Associated Phenomena*. Princeton University Press, 1949.

[Halim *et al.*, 2008] Felix Halim, Yongzheng Wu, and Roland H. C. Yap. Small world networks as (semi)-structured overlay networks. In *SASOW '08: Proceedings of the 2008 Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems Workshops*, pages 214–218, Washington, DC, USA, 2008. IEEE Computer Society.

[Hastings, 1970] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[Helbing and Vicsek, 1999] Dirk Helbing and Tamás Vicsek. Optimal self-organization. *New Journal of Physics*, 1, 1999.

[Hildrum *et al.*, 2002] Kirsten Hildrum, John D. Kubiatowicz, Satish Rao, and Ben Y. Zhao. Distributed object location in a dynamic network. In *SPAA '02: Proceedings of the fourteenth annual ACM symposium on Parallel algorithms and architectures*, pages 41–52, New York, NY, USA, 2002. ACM.

[Hong and Scaglione, 2003] Yao-Win Hong and A. Scaglione. Time synchronization and reach-back communications with pulse-coupled oscillators for uwb wireless ad hoc networks. In *Ultra Wideband Systems and Technologies, 2003 IEEE Conference on*, pages 190 – 194, November 2003.

[Hoory *et al.*, 2006] Shlomo Hoory, Nathan Linial, Avi Wigderson, and An Overview. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S*, 43:439–561, 2006.

[Horowitz and Malkhi, 2003] Keren Horowitz and Dahlia Malkhi. Estimating network size from local information. *Inf. Process. Lett.*, 88(5):237–243, 2003.

[Hwang *et al.*, 2005] D.-U. Hwang, M. Chavez, A. Amann, and S. Boccaletti. Synchronization in complex networks with age ordering. *Physical Review Letters*, 94(13):138701, 2005.

[IBM, 2001] IBM. Autonomic computing: Ibm's perspective on the state of information technology. Technical report, IBM, 2001.

[Ichinomiya, 2004] Takashi Ichinomiya. Frequency synchronization in a random oscillator network. *Phys. Rev. E*, 70(2):026116, Aug 2004.

[Izhikevich, 1998] Eugene M. Izhikevich. Phase models with explicit time delays. *Phys. Rev. E*, 58(1):905–908, Jul 1998.

[Jadbabaie *et al.*, 2004] Ali Jadbabaie, Nader Motee, and Mauricio Barahona. On the stability of the kuramoto model of coupled nonlinear oscillators. In *In Proceedings of the American Control Conference*, pages 4296–4301, 2004.

[Jelasity and Montresor, 2004] Márk Jelasity and Alberto Montresor. Epidemic-style proactive aggregationin large overlay networks. In *Proceedings of The 24th International Conference on Distributed ComputingSystems (ICDCS 2004)*, pages 102–109, Tokyo, Japan, 2004. IEEE Computer Society.

[Jelasity *et al.*, 2005] Márk Jelasity, Alberto Montresor, and Ozalp Babaoglu. Gossip-based aggregation in large dynamic networks. *ACM Trans. Comput. Syst.*, 23(3):219–252, 2005.

[Jelasity *et al.*, 2007] Márk Jelasity, Spyros Voulgaris, Rachid Guerraoui, Anne-Marie Kermarrec, and Maarten van Steen. Gossip-based peer sampling. *ACM Trans. Comput. Syst.*, 25(3):8, 2007.

[Jelasity *et al.*, 2009] Márk Jelasity, Alberto Montresor, and Ozalp Babaoglu. T-man: Gossip-based fast overlay topology construction. *Comput. Netw.*, 53(13):2321–2339, 2009.

[Jeong *et al.*, 2000] Hawoong Jeong, B. Tombor, Réka Albert, Zoltan Nagy Oltvai, and Albert-Laszló Barabási. The large-scale organization of metabolic networks. *Nature*, 407:651–654, July 2000.

[Ji *et al.*, 2008] Li Ji, Wang Bing-Hong, Wang Wen-Xu, and Zhou Tao. Network entropy based on topology configuration and its computation to random networks. *Chinese Physics Letters*, 25(11):4177, 2008.

[Kac, 1966] Mark Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73:1–23, 1966.

[Karger *et al.*, 1997] David Karger, Eric Lehman, Tom Leighton, Rina Panigrahy, Matthew Levine, and Daniel Lewin. Consistent hashing and random

trees: distributed caching protocols for relieving hot spots on the world wide web. In *STOC '97: Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, pages 654–663, New York, NY, USA, 1997. ACM.

[Kempe and McSherry, 2004] David Kempe and Frank McSherry. A decentralized algorithm for spectral analysis. In *STOC '04: Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 561–568, New York, NY, USA, 2004. ACM.

[Kermack and McKendrick, 1927] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character (1905-1934)*, 1927.

[Keyani *et al.*, 2002] Pedram Keyani, Brian Larson, and Muthukumar Senthil. Peer pressure: Distributed recovery from attacks in peer-to-peer systems. In *Revised Papers from the NETWORKING 2002 Workshops on Web Engineering and Peer-to-Peer Computing*, pages 306–320, London, UK, 2002. Springer-Verlag.

[Kirkpatrick *et al.*, 1983] Scott Kirkpatrick, C.D. Gelatt Jr., and M.P. Vecci. Optimization by simulated annealing. *Science*, 220:671, 1983.

[Kleinberg, 2007] Robert Kleinberg. Geographic routing using hyperbolic space. In *In Proceedings of the 26th IEEE International Conference on Computer Communications*, 2007.

[Kostoulas *et al.*, 2005] Dionysios Kostoulas, Dimitrios Psaltoulis, Indranil Gupta, Ken Birman, and Al Demers. Decentralized schemes for size estimation in large and dynamic groups. In *NCA '05: Proceedings of the Fourth IEEE International Symposium on Network Computing and Applications*, pages 41–48, Washington, DC, USA, 2005. IEEE Computer Society.

[Ktari *et al.*, 2009] Salma Ktari, Artur Hecker, and Houda Labiod. A construction scheme for scale free dht-based networks. In *Proceedings of the Global Communications Conference, 2009. GLOBECOM 2009, Honolulu, Hawaii, USA, 30 November - 4 December 2009*, pages 1–6. IEEE, 2009.

[Kubiatowicz, 2003] John Kubiatowicz. Extracting guarantees from chaos. *Commun. ACM*, 46(2):33–38, 2003.

[Kuramoto, 1975] Y Kuramoto. Self-entrainment of a population of coupled nonlinear oscillators. In *International symposium on mathematical problems in theoretical physics*, pages 420–422. Springer, 1975.

[Lakhina *et al.*, 2003] A. Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling biases in ip topology measurements. In *Proceedings of the 22nd Annual Joint Conference on the IEEE Computer and Communications Societies. INFOCOM 2003*, 2003.

[Lee *et al.*, 2005] D.-S. Lee, K.-I. Goh, B. Kahng, and D. Kim. Scale-free random graphs and potts model. *Pramana Journal of Physics*, 64(6):1149–1159, June 2005.

[Li *et al.*, 2005] Lun Li, David Alderson, Reiko Tanaka, John C. Doyle, and Walter Willinger. Towards a theory of scale-free graphs: Definition, properties, and implications (extended version). *Internet Mathematics*, 2(4):431–523, 2005.

[Li, 2008] Xiang Li. The role of degree-weighted couplings in the synchronous onset of kuramoto oscillator networks. *Physica A: Statistical Mechanics and its Applications*, 387(26):6624–6630, November 2008.

[Liljeros *et al.*, 2001] F. Liljeros, C.R. Edling, L.A.N. Amaral, and Heugene E. Stanleyand Y. Aberg. The web of human sexual contacts. *Nature*, 411:907–908, 2001.

[Lovász, 1993] László Lovász. Random walks on graphs: A survey. *Combinatorics, Paul Erdös is Eighty*, 2:1–46, 1993.

[Lu and Fang, 2005] Tianbo Lu and Binxing Fang. Building scale-free overlay mix networks with small-world properties. In *ICITA '05: Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05) Volume 2*, pages 529–534, Washington, DC, USA, 2005. IEEE Computer Society.

[Lua *et al.*, 2005] Keong Lua, J. Crowcroft, M. Pias, R. Sharma, and S. Lim. A survey and comparison of peer-to-peer overlay network schemes. *Communications Surveys & Tutorials, IEEE*, pages 72–93, 2005.

[Lucarelli and Wang, 2004] Dennis Lucarelli and I-Jeng Wang. Decentralized synchronization protocols with nearest neighbor communication. In *SenSys '04: Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 62–68, New York, NY, USA, 2004. ACM.

[Lv *et al.*, 2002] Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. Search and replication in unstructured peer-to-peer networks. In *Proceedings of the 16th international conference on Supercomputing (ICS 2002)*, pages 84–95, New York, NY, USA, 2002. ACM.

[Mandelbrot, 1982] Benoît B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Company, New York, 1982.

[Massoulié *et al.*, 2006] Laurent Massoulié, Erwan Le Merrer, Anne-Marie Kermarrec, and Ayalvadi Ganesh. Peer counting and sampling in overlay networks: random walk methods. In *PODC '06: Proceedings of the twenty-fifth annual ACM symposium on Principles of distributed computing*, pages 123–132, New York, NY, USA, 2006. ACM.

[Maymounkov and Mazieres, 2002] Petar Maymounkov and David Mazieres. Kademlia: A peer-to-peer information system based on the xor metric. In *Proceedings of IPTPS02, Cambridge, USA, March 2002.*, 2002.

[Merrer *et al.*, 2006] E. Le Merrer, A.-M. Kermarrec, and L. Massoulie. Peer to peer size estimation in large and dynamic networks: A comparative study. *High-Performance Distributed Computing, International Symposium on*, 0:7–17, 2006.

[Metropolis *et al.*, 1953] N. Metropolis, A. W. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[Milgram, 1967] Stanley Milgram. The small-world problem. *Psychology Today*, pages 60–67, 1967.

[Mirollo and Strogatz, 1990] Renato E. Mirollo and Steven H. Strogatz. Synchronization of pulse-coupled biological oscillators. *SIAM J. Appl. Math.*, 50(6):1645–1662, 1990.

[Mitzenmacher, 2002] Michael Mitzenmacher. A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251, 2002.

[Müller-Schloer *et al.*, 2004] Christian Müller-Schloer, Christoph von der Malsburg, and Rolf P. Würtz. Aktueller schlagwort: Organic computing. *Informatik Spektrum*, 27:2–6, August 2004.

[Mogul, 2006] Jeffrey C. Mogul. Emergent (mis)behavior vs. complex software systems. *SIGOPS Oper. Syst. Rev.*, 40(4):293–304, 2006.

[Mohar, 1991] Bojan Mohar. The Laplacian spectrum of graphs. In *Graph Theory, Combinatorics, and Applications*, volume 2, pages 871–898. Wiley, 1991.

[Molloy and Reed, 1995] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–180, 1995.

[Moreno and Pacheco, 2004] Y. Moreno and A. F. Pacheco. Synchronization of kuramoto oscillators in scale-free networks. *EPL (Europhysics Letters)*, 68(4):603–609, 2004.

[Moreno *et al.*, 2004] Y. Moreno, M. Vazquez-Prada, and A. F. Pacheco. Fitness for synchronization of network motifs. *Physica A*, 343, 2004.

[Motter *et al.*, 2005a] Adilson E. Motter, C. Zhou, and Jürgen Kurths. Weighted networks are more synchronizable: how and why. In J. Mendes, J. G. Oliveira, F. V. Abreu, A. Povolotsky, and S. N. Dorogovtsev, editors, *Science of Complex Networks: From Biology to the Internet and WWW; CNET 2004*, volume 776 of *American Institute of Physics Conference Series*, pages 201–214, June 2005.

[Motter *et al.*, 2005b] Adilson E. Motter, Changsong Zhou, and Jurgen Kurths. Network synchronization, diffusion, and the paradox of heterogeneity. *Physical Review Letters E (Statistical, Nonlinear, and Soft Matter Physics)*, 71(1):016116, 2005.

[Motter, 2002] Adilson E. Motter. Cascade-based attacks on complex networks. *Physical Review Letters E (Statistical, Nonlinear, and Soft Matter Physics)*, 66, 2002.

[Newman *et al.*, 2000] M. E. J. Newman, C. Moore, and D. J. Watts. Mean-field solution of the small-world network model. *Phys. Rev. Lett.*, 84(14):3201–3204, Apr 2000.

[Newman *et al.*, 2001] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review Letters E (Statistical, Nonlinear, and Soft Matter Physics)*, 64, 2001.

[Newman, 2001] M. E. J. Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, 98:404, 2001.

[Newman, 2002] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters E (Statistical, Nonlinear, and Soft Matter Physics)*, 89, 2002.

[Nishikawa *et al.*, 2003] T. Nishikawa, A. E. Motter, Y.-C. Lai, and F. C. Hoppensteadt. Heterogeneity in Oscillator Networks: Are Smaller Worlds Easier to Synchronize? *Physical Review Letters*, 91(1):014101–+, July 2003.

[Oh *et al.*, 2005] E. Oh, K. Rho, H. Hong, and B. Kahng. Modular synchronization in complex networks. *Physical Review Letters E (Statistical, Nonlinear, and Soft Matter Physics)*, 72(4), October 2005.

[Olfati-Saber *et al.*, 2007] R. Olfati-Saber, J. A. Fax, and R. M. Murray. Consensus and cooperation in networked multi-agent systems. *Proceedings of the IEEE*, 95(1):215–233, 2007.

[Oppenheim *et al.*, 1999] Alan V. Oppenheim, Ronald W. Schafer, and John R. Buck. *Discrete-time signal processing (2nd ed.)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1999.

[Palla *et al.*, 2004] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Statistical mechanics of topological phase transitions in networks. *Physical Review Letters E (Statistical, Nonlinear, and Soft Matter Physics)*, 69:046117, 2004.

[Papachristodoulou and Jadbabaie, 2005] A. Papachristodoulou and A. Jadbabaie. Synchronization in oscillator networks: Switching topologies and non-homogeneous delays. In *44th IEEE Conference on Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05.*, 2005.

[Papadopoulos *et al.*, 2010] Fragkiskos Papadopoulos, Dmitri Krioukov, Marian Boguna, and Amin Vahdat. Greedy forwarding in dynamic scale-free networks embedded in hyperbolic metric spaces. In *Proceedings of the IEEE Infocom Conference*, 2010.

[Park *et al.*, 2006] Kwangho Park, Ying-Cheng Lai, Saurabh Gupte, and Jong-Won Kim. Synchronization in complex networks with a modular structure. *Chaos*, 16, 2006.

[Pastor-Satorras and Vespignani, 2001] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14), 2001.

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

[Pecora and Barahona, 2005] L. Pecora and M. Barahona. Synchronization of oscillators in complex networks. *Chaos and Complexity Letters*, 1(61):61–91, 2005.

[Peskin, 1975] Charles S. Peskin. Mathematical aspects of heart physiology. Technical report, Courant Institute of Mathematical Sciences, 1975.

[Pikovsky *et al.*, 2003] Arkady Pikovsky, Michael Rosenblum, and Jürgen Kurths. *Synchronization: A Universal Concept in Nonlinear Sciences*. Cambridge University Press, 2003.

[Qi and Yu, 2008] Jing Qi and Jiguo Yu. Scale-free overlay structures for unstructured peer-to-peer networks. In *Seventh International Conference on Grid and Cooperative Computing (GCC), Shenzhen, PRC, 2008*, 2008.

[Ratnasamy *et al.*, 2001] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Schenker. A scalable content-addressable network. In *SIGCOMM '01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 161–172, New York, NY, USA, 2001. ACM.

[Ripeanu *et al.*, 2002] Matei Ripeanu, Adriana Iamnitchi, and Ian Foster. Mapping the gnutella network. *IEEE Internet Computing*, 6(1):50–57, 2002.

[Rossi *et al.*, 2009] D. Rossi, M. Mellia, and M. Meo. Evidences behind skype outage. In *Proceedings of IEEE International Conference on Communication (ICC'09)*, 2009.

[Rowstron and Druschel, 2001] Antony Rowstron and Peter Druschel. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. *Lecture Notes in Computer Science*, 2218, 2001.

[Sandberg, 2006] Oskar Sandberg. Distributed routing in small-world networks. In *Algorithm Engineering and Experiments*. SIAM, 2006.

[Saramaeki and Kaski, 2004] Jari Saramaeki and Kimmo Kaski. Scale-free networks generated by random walkers. *Physica A: Statistical Mechanics and its Applications*, 341:80–86, 2004.

[Sarshar and Roychowdhury, 2004] Nima Sarshar and Vwani Roychowdhury. Scale-free and stable structures in complex ad hoc networks. *Physical Review E*, 69:026101, 2004.

[Sarshar and Roychowdhury, 2005] Nima Sarshar and Vwani Roychowdhury. Multiple scale-free structures in complex ad-hoc networks. *PHYS.REV.E*, 72, 2005.

[Satoh, 1989] Kazuhiro Satoh. Computer experiment on the cooperative behavior of a network of interacting nonlinear oscillators. *Journal of the Physical Society of Japan*, 58:2010–2021, 1989.

[Schmeck, 2005] Hartmut Schmeck. Organic computing. *Künstliche Intelligenz*, pages 68—69, 2005.

[Servedio and Caldarelli, 2004] Vito D. P. Servedio and Guido Caldarelli. Vertex intrinsic fitness how to produce arbitrary: How to produce arbitrary scale-free networks. *Physical Review E*, 70(5), November 2004.

[Shahabi and Banaei-Kashani, 2005] Cyrus Shahabi and Farnoush Banaei-Kashani. Modelling peer-to-peer data networks under complex system theory. In Subhash Bhalla, editor, *Databases in Networked Information Systems*, volume 3433 of *Lecture Notes in Computer Science*, pages 238–243. Springer Berlin / Heidelberg, 2005.

[Shannon, 1948] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948.

[Shavitt and Shir, 2005] Yuval Shavitt and Eran Shir. Dimes: Let the internet measure itself. *ACM SIGCOMM Computer Communication Review*, 35(5):71–74, October 2005.

[Simonyi, 1995] Gábor Simonyi. Graph entropy: A survey. *Combinatorial Optimization, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 1995.

[Sinclair, 1992] Alistair Sinclair. Improved bounds for mixing rates of markov chains and multicommodity flow. *Combinatorics, Probability and Computing*, 1:351–370, 1992.

[Stoica *et al.*, 2001] Ion Stoica, Robert Morris, David Karger, M. Frans Kaashoek, and Hari Balakrishnan. Chord: A scalable peer-to-peer lookup service for internet applications. In *SIGCOMM '01: Proceedings of the 2001 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 149–160, New York, NY, USA, 2001. ACM Press.

[Strogatz and Mirollo, 1988] Steven H. Strogatz and Renato E. Mirollo. Phase-locking and critical phenomena in lattices of coupled nonlinear oscillators with random intrinsic frequencies. *Phys. D*, 31(2):143–168, 1988.

[Strogatz *et al.*, 2005] Steven Strogatz, Daniel M. Abrams, Allan McRobie, Bruno Eckhardt, and Edward Ott. Crowd synchrony on the millennium bridge. *Nature*, 438:43–44, November 2005.

[Strogatz, 2000] S H Strogatz. From kuramoto to crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D*, 143(1-4):1–20, 2000.

[Strogatz, 2003] Steven H. Strogatz. *Sync: The Emerging Science of Spontaneous Order*. Hyperion, 2003.

[Stutzbach and Rejaie, 2006] Daniel Stutzbach and Reza Rejaie. Understanding churn in peer-to-peer networks. In *IMC '06: Proceedings of the 6th ACM SIGCOMM on Internet measurement*, pages 189–202, New York, NY, USA, 2006. ACM Press.

[Stutzbach *et al.*, 2006] D. Stutzbach, R. Rejaie, N. Duffield, S. Sen, and W. Willinger. Sampling techniques for large, dynamic graphs. In *Proceedings of the 25th IEEE International Conference on Computer Communications. INFOCOM 2006*, pages 1–6, April 2006.

[Stutzbach *et al.*, 2008] Daniel Stutzbach, Reza Rejaie, and Subhabrata Sen. Characterizing unstructured overlay topologies in modern p2p file-sharing systems. *IEEE/ACM Trans. Netw.*, 16(2):267–280, 2008.

[Stutzbach *et al.*, 2009] Daniel Stutzbach, Reza Rejaie, Nick Duffield, Subhabrata Sen, and Walter Willinger. On unbiased sampling for unstructured peer-to-peer networks. *IEEE/ACM Trans. Netw.*, 17(2):377–390, 2009.

[Terpstra *et al.*, 2007] Wesley W. Terpstra, Jussi Kangasharju, Christof Leng, and Alejandro P. Buchmann. Bubblestorm: resilient, probabilistic, and exhaustive peer-to-peer search. In *SIGCOMM '07: Proceedings of the 2007 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 49–60, New York, NY, USA, 2007. ACM.

[Thommes and Coates, 2005] R. Thommes and M. Coates. Epidemiological modelling of peer-to-peer viruses and pollution. Technical report, Department of Electrial and Computer Engineering, McGill University, 2005.

[Tian *et al.*, 2005] Ruixiong Tian, Yongqiang Xiong, Qian Zhang, Bo Li, Ben Y. Zhao, and Xing Li. Hybrid overlay structure based on random walks. In *In Proc. of the 4th Intl. Workshop on Peer-toPeer Systems (IPTPS05*, 2005.

[UCTE, 2007] UCTE. Final report system disturbance on 4 november 2006, January 2007.

[van Steen, 2004] Maarten van Steen. On the complexity of simple distributed systems. *IEEE Distributed Systems Online*, 5(5):3, May 2004.

[van Steen, 2010] Maarten van Steen. *Graph Theory and Complex Networks: An Introduction*. Maarten van Steen, 2010.

[Vazquez *et al.*, 2001] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Modeling of protein interaction networks. *ArXiv Condensed Matter e-prints*, August 2001.

[Vazquez, 2000] A. Vazquez. Knowing a network by walking on it: emergence of scaling. *ArXiv Condensed Matter e-prints*, June 2000.

[Vicsek *et al.*, 1995] Tamás Vicsek, Andras Czirok, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical Review Letters*, 75:1226 – 1229, 1995.

[Vicsek *et al.*, 1999] Tamás Vicsek, András Czirók, Illés J. Farkas, and Dirk Helbing. Application of statistical mechanics to collective motion in biology. *Physica A: Statistical Mechanics and its Applications*, 274(1-2):182 – 189, 1999.

[Vicsek, 1992] Tamás Vicsek. *Fractal Growth Phenomena*. World Scientific, 1992.

[Vishnumurthy and Francis, 2006] V. Vishnumurthy and P. Francis. On heterogeneous overlay construction and random node selection in unstructured p2p networks. In *Proceedings of the 25th IEEE International Conference on Computer Communications INFOCOM 2006*, pages 1 –12, apr. 2006.

[von Foerster, 1985] Heinz von Foerster. *Sicht und Einsicht*. Vieweg, Braunschweig, 1985.

[Voulgaris *et al.*, 2005] Spyros Voulgaris, Daniela Gavidia, and Maarten van Steen. Cyclon: Inexpensive membership management for unstructured p2p overlays. *J. Network Syst. Manage.*, 13(2), 2005.

[Waclaw, 2007] Bartlomiej Waclaw. *Statistical mechanics of complex networks*. PhD thesis, Jagellonian University, Cracow, Poland, April 2007.

[Waldhorst *et al.*, 2010] Oliver P. Waldhorst, Roland Bless, and Martina Zitterbart. Overlay-netze als innovationsmotor im internet. *Informatik Spektrum*, 33(2):171–185, 2010.

[Watts and Strogatz, 1998] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature.*, 393, 1998.

[Watts, 1999] Duncan J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 1999.

[Weaver, 1948] Warren Weaver. Science and complexity. *American Scientist*, 36:536–544, 1948.

[Wehrle *et al.*, 2005] Klaus Wehrle, Stefan Goetz, Rüdiger Schollmeier, Simon Rieche, Jörg Eberspächer, Ralf Steinmetz, Oliver Heckmann, Andreas Mauthe, Vasilios Darlagiannis, and Christian Koppen. *Peer-to-Peer Systems and Applications*. Springer Berlin / Heidelberg, 2005.

[Werner-Allen *et al.*, 2005] Geoffrey Werner-Allen, Geetika Tewari, Ankit Patel, Matt Welsh, and Radhika Nagpal. Firefly-inspired sensor network synchronicity with realistic radio effects. In *Proceedings of The Third International Conference on Embedded Networked Sensor Systems*, pages 142–153, 2005.

[Wiener, 1958] Norbert Wiener. *Nonlinear Problems in Random Theory*. MIT Press, 1958.

[Willinger and Doyle, 2002] Walter Willinger and John Doyle. Robustness and the internet: Design and evolution. In *Informational [Page 21] 3426 Architectural and Policy Considerations*, 2002.

[Winfree, 1967] A. T. Winfree. Biological rhythms and the behavior of populations of coupled oscillators. *Journal of Theoretical Biology*, 16:15–42, 1967.

[Yates *et al.*, 2009] Christian A. Yates, Radek Erban, Carlos Escudero, Iain D. Couzin, Jerome Buhl, Ioannis G. Kevrekidis, Philip K. Maini, and David J. T. Sumpter. Inherent noise can facilitate coherence in collective swarm motion. *Proceedings of the National Academy of Sciences of the United States of America*, 106(14):5464–5469, 2009.

[Zhang *et al.*, 2008] Honggang Zhang, Giovanni Neglia, Donald F. Towsley, and Giuseppe Lo Presti. Stability and efficiency of unstructured file sharing networks. *IEEE Journal on Selected Areas in Communications*, 26(7):1284–1294, 2008.

[Zhong *et al.*, 2005] Ming Zhong, Kai Shen, and Joel Seiferas. Non-uniform random membership management in peer-to-peer networks. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 2, pages 1151 – 1161 vol. 2, march 2005.

[Zhong *et al.*, 2008] Ming Zhong, Kai Shen, and Joel Seiferas. The convergence-guaranteed random walk and its applications in peer-to-peer networks. *IEEE Trans. Comput.*, 57(5):619–633, 2008.