

Analyse der Verbindung von Topologie und Dynamik in abstrakten Graphen und biologischen Netzwerken

Vom Fachbereich Biologie der Technischen Universität Darmstadt
zur
Erlangung des akademischen Grades
eines Doctor rerum naturalium
genehmigte
Dissertation von

Dipl.-Biol. Mark Müller-Linow
aus Darmstadt

Berichterstatter: Prof. Dr. Marc-Thorsten Hütt
Mitberichterstatter: Prof. Dr. Ralf A. W. Galuske

Tag der Einreichung: 5. März 2008
Tag der mündlichen Prüfung: 24. April 2008
Darmstadt 2008 D17

Inhaltsverzeichnis

1	Einleitung	3
2	Eigenschaften von Graphen	11
2.1	Wichtige globale und lokale Eigenschaften: Knotenzahl, Kantenzahl, Vernetzungsgrad, Grad und Clusterkoeffizient	11
2.2	Gradverteilung und <i>Hubs</i>	13
2.3	Gradkorrelationen	13
2.4	Modularität	14
2.5	Längenmaße	15
2.6	Zentralitätsmaße	16
2.7	Hierarchische Strukturen	17
3	Dynamische Modelle	19
3.1	Das diskret erregbare (DE) dynamische Modell	20
3.2	Varianten des DE-Modells	20
4	Graphen und Netzwerke	23
4.1	Abstrakte Graphen	23
4.1.1	Zufällige Graphen - das Erdős-Rényi (ER) Modell	23
4.1.2	<i>Small-World</i> Graphen - das Watts-Strogatz (WS) Modell	24
4.1.3	Skalenfreie Graphen - das Barabási-Albert (BA) Modell	25
4.1.4	Modulare Graphen - das <i>Community</i> Modell	25
4.1.5	Hierarchische Graphen	26
4.2	Reale Systeme	27
4.2.1	Neuronale Netzwerke	27
4.2.2	Metabolische Reaktions-Netzwerke	28
4.2.3	Metabolische Korrelations-Netzwerke	29
4.3	Randomisierte Graphenstrukturen	31
5	Ergebnisse Teil I - Verteilungsmuster von Anregungen auf abstrakten Graphen	33
5.1	Einführung	33
5.2	Einfluss der Vernetzungsdichte	35

5.3	Einfluss der Gradverteilung	37
5.4	Einfluss der Spontanaktivität	40
5.5	Einfluss von Gradkorrelationen	41
5.6	Diskussion	43
6	Ergebnisse Teil II - Dynamische Organisation von modularen Strukturen und dominanten Knoten in abstrakten Graphen und neuronalen Systemen	45
6.1	Einführung	45
6.2	Analyse modularer Topologien	47
6.3	Analyse von <i>Hub</i> -dominierten Topologien	52
6.4	Analyse von randomisierten Graphentopologien	53
6.5	Analyse hierarchischer Graphentopologien	58
6.6	Analyse von biologischen neuronalen Topologien	62
6.7	Diskussion	65
7	Ergebnisse Teil III - Konsistenzanalysen an metabolischen Netzwerken	69
7.1	Einführung	69
7.2	Ähnlichkeitsanalyse von Netzwerken	72
7.3	Zeitliche Konsistenzen in metabolischen Korrelations-Netzwerken	73
7.4	Ähnlichkeit von Korrelations-Netzwerken und metabolischen Reaktions-Netzwerken	76
7.5	Analyse von Modul-Modul Interaktionen	78
7.6	Diskussion	82
8	Zusammenfassung und Ausblick	85
9	Übersicht über hier entwickelte Methoden und Softwaremodule	89
10	Abkürzungsverzeichnis	91
	Literaturverzeichnis	95
	Lebenslauf	103
	Publikationen	104
	Danksagung	105
	Erklärung	106

Kapitel 1

Einleitung

Am 10. Oktober 2007 erhielt der Physiker Gerhard Ertl¹ für seine "Untersuchungen von chemischen Prozessen auf festen Oberflächen" den Nobelpreis für Chemie. Eine interessante Beobachtung, die er während seiner Studien an katalytisch aktiven Platinoberflächen² machte, waren raumzeitliche Phänomene von hoher Ästhetik (Ertl, 1991). Rund 250 Jahre zuvor sprach der Mathematiker Christian Goldbach (1690-1764) mit seinem Kollegen Leonhard Euler (1707-1783) über die Sehenswürdigkeiten seines Geburtsortes Königsberg. Dabei muss wohl das Zitat "Über sieben Brücken musst Du geh'n." gefallen sein, welches Euler zu einer neuen Betrachtungsweise in der Mathematik anregte.

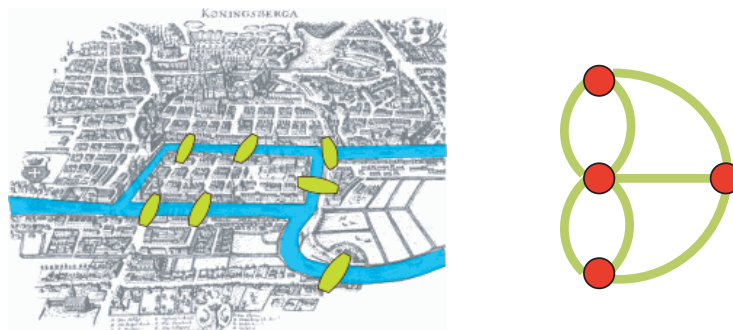


Abbildung 1.1: Die Karte (links) zeigt eine alte Darstellung von Königsberg (heute: Kaliningrad) mit seinen sieben Brücken (grün), welche die Stadtteile miteinander verbinden, die durch den Fluss Pregel getrennt werden (Quelle: aus Wikipedia - Königsberger Brückenproblem). Euler abstrahierte den Stadtplan zu einem Graphen (rechts), in dem die Knoten (rot) die Stadtteile und die Kanten (grün) die Brücken symbolisieren, und löste das Königsberger Brückenproblem graphentheoretisch, indem er nachwies, dass kein Rundweg existiert, bei dem jede Brücke nur einmal benutzt wird.

In der damaligen Aufgabenstellung ging es um den Nachweis eines Rundwegs, der jede der sieben Brücken genau einmal überquert. Die Brücken verbanden die Königsberger Stadtteile, welche durch den Fluss Pregel getrennt waren. Euler löste dieses Problem graphentheoretisch, indem er die Stadtteile zu

¹Prof. Dr. Gerhard Ertl wurde am 13.2.2008 in den Hochschulrat der TU Darmstadt berufen. Er ist Direktor am Fritz-Haber-Institut der Max-Planck-Gesellschaft in Berlin.

²Platin ermöglicht die katalytische Oxidation von Kohlenmonoxid zu Kohlendioxid

Punkten und die Brücken zu verbindenden Linien abstrahierte (siehe Abb. 1.1). Mit diesem Netzwerk (Graph) bewies er, dass es keinen solchen Rundweg gibt. Das nach ihm benannte "Eulerproblem" ist heute Bestandteil einer Reihe weiterer Problemstellungen, die die *Durchlaufbarkeit* von Knoten (den Punkten) und Kanten (den verbindenden Linien) von Graphen behandeln. Zu ihnen gehören u.a. das Hamiltonkreisproblem (vom Mathematiker Sir William Rowan Hamilton 1857 als Spiel "The Icosian Game" eingeführt) und das Briefträgerproblem (von Mei Ko Kwan 1962 untersucht). Diese mathematischen Rätsel bildeten schon in ihrer Zeit einen Zugang zum Verständnis abstrakter Strukturen. Übertragen auf große Systeme spielen sie jedoch bis heute eine wichtige Rolle bei der Lösung vielfältiger praktischer Problemstellungen, wie z.B. in Transport, Logistik und Fertigungstechnik.

Viele der ersten graphentheoretischen Denkansätze berührten geographische Probleme. Zu den denkbar einfachsten Graphen und im geographischen Sinne naheliegenden Strukturen gehören simple Gitter, welche eine Abstraktion einer Oberfläche darstellen, und bei der ein Knotenpunkt (stellvertretend für einen Ausschnitt) nur in Kontakt zu seinen direkten (z.B. vier) Nachbarn steht. Solche Vereinfachungen zu regulären Graphenstrukturen haben sich als sehr vorteilhaft zur Beschreibung verschiedenster Prozesse erwiesen, zu denen auch die oben genannten raumzeitlichen Musterbildungen auf Platinoberflächen gehören. Das Phänomen raumzeitlicher Musterbildung begründete zwar keine neue Denkweise in der Wissenschaft, steht aber stellvertretend für eine Reihe von dynamischen Beobachtungen in chemischen wie biologischen Systemen, die in den letzten Jahrzehnten unter anderem durch Methoden modelliert wurden, für die Euler einen wichtigen Grundstein mit seinem Werk "solutio problematis ad geometriam situs pertinens" aus dem Jahre 1736 gelegt hatte, und in dem das "Königsberger Brückenproblem" als strukturelles Problem eine zentrale Rolle in der neu geschaffenen Graphentheorie einnimmt.

Die Topologie (also globale Architektur oder Struktur) vieler biologischer Systeme lässt sich durch Graphen-Äquivalente darstellen. In diesen Graphen sind die einzelnen Komponenten eines biologischen Systems zu Knoten und mögliche Interaktionen zwischen den Elementen zu Kanten abstrahiert. Einfache Graphenmodelle, wie reguläre Gitter, reichen jedoch nicht zur Beschreibung aus; die Topologie biologischer Netzwerke zeichnet sich vielmehr durch eine hohe Komplexität aus, und ebenso komplex ist auch das dynamische Verhalten. Die Abstraktion realer Systeme zu Graphenmodellen eröffnet völlig neue Möglichkeiten zur Untersuchung, zu denen die graphentheoretischen Methoden, aber auch der Einsatz spezieller dynamischer Modellsysteme, gehören. Auf der anderen Seite bedeutet diese Abstraktion eine starke Vereinfachung, wodurch Untersuchungen auf systemischer Ebene erst ermöglicht werden. Einen guten Anhaltspunkt für die Qualität biologischer Netzwerke liefert der Vergleich experimenteller Daten, welche die analytische Grundlage für die Rekonstruktion biologischer Netzwerke darstellen, und Modellen, die auf der Basis bekannter Struktureigenschaften von biologischen Netzwerken erstellt worden sind. Betrachtet man sehr gut untersuchte biochemische Systeme, wie die Genregulation oder den Metabolismus, dann findet man (zumindest für einige Modellsysteme) eine Vielzahl frei verfügbarer Datensätze, die trotz gewisser Lücken eine hohe Konsistenz aufweisen. Netzwerke, die auf dieser Datengrundlage rekonstruiert werden, repräsentieren die essentiellen Eigenschaften des realen Vorbilds. Auf der anderen Seite gibt es aber auch eine Reihe sehr inkonsistenter Datensätze. Ein guter Beleg sind die Interaktionen zwischen Proteinen. Betrachtet man für bestimmte Organismen die Summe aller so rekonstruierten

Proteininteraktions-Netzwerke, so findet man nur einen geringen Anteil gemeinsamer Elemente, was darauf hindeutet, dass bislang nur verschiedene Subsysteme erfasst worden sind (Bork *et al.*, 2004). Ein anderes Problem stellt die Modellierung solcher Systeme dar. Ein reales System so zu imitieren, dass eine möglichst große Übereinstimmung hinsichtlich der topologisch charakterisierenden Größen erreicht wird, ist bislang nicht gelungen. Zum einen liegt es in der Natur komplexer Systeme, dass die Kenntnis einzelner Komponenten bzw. einzelner struktureller Eigenschaften alleine nicht ausreicht, um ein System in seiner Gesamtheit zu beschreiben, da es letztlich das komplizierte Zusammenspiel ist, das aus einfachen Bausteinen ein komplettes System hervorbringt. Zum anderen stellt sich die Frage, welche strukturellen Eigenschaften ein System dominieren und welche nur als Konsequenz von anderen Eigenschaften entstehen.

Deshalb beschränken sich die Studien verschiedener Klassen abstrakter Graphen auf ausgesuchte Eigenschaften, die auch in realen Systemen zu finden sind. Einen wichtigen Beitrag zur Erforschung von Graphenstrukturen leisteten Erdős & Rényi (1959) mit der Einführung des zufälligen Graphen (oder kurz: ER-Graph). Der ER-Graph diente lange Zeit als Studienobjekt für spezielle topologische Merkmale und ihrer Abhängigkeit von bestimmten Systemeigenschaften. Den Schritt in Richtung realer Systeme vollzogen rund 40 Jahre später Watts & Strogatz (1998), die mit Einführung des *Small-World* Modells zum ersten Mal zwei Systemeigenschaften realer Netzwerke - eine Überhöhung von Dreiecksstrukturen in der Umgebung einzelner Knoten (*Clustering*) bei gleichzeitig geringen Abständen (wenigen Kanten) zwischen beliebigen Knotenpaaren (*Weglängen*) - in ein Graphenmodell integrierten. Diese beiden Eigenschaften deuten die Grenzen struktureller Skalen an, die im Mittelpunkt topologischer Untersuchungen stehen, nämlich die Eigenschaften eines Knotens und seiner direkten Nachbarschaft (die lokale Skala) bzw. die Eigenschaften eines Graphen in seiner Gesamtheit (die globale Skala).

Beginnt man die Analyse auf der kleinsten Skala, dem einzelnen Knoten, so ist es naheliegend, den *Grad* (die Zahl seiner Kanten) für eine Einschätzung seiner Rolle im System heranzuziehen. Doch was sagt diese Zahl aus? Erlangt ein Knoten an Bedeutung, weil er viele Kanten besitzt, oder kann auch (oder gerade) ein Knoten mit geringem Grad wichtig sein? Ohne die Kenntnis seiner Umgebung ist eine solche Beurteilung im Allgemeinen nicht möglich, obwohl durch die Studien biologischer Systeme (in denen die individuellen Knotengrade teilweise enorm voneinander abweichen) bekannt ist, dass bestimmte Knoten mit besonders vielen Verbindungen meist auch von größerer Bedeutung für das System sind, als Knoten mit weniger Verbindungen (Jeong *et al.*, 2001). Diese besonderen Knoten, die auch als *Hubs* bezeichnet werden, stellen entweder Verknüpfungspunkte für verschiedene Graphenregionen dar, oder sie organisieren andere Knoten in ihrer direkten Nachbarschaft (Guimerà & Amaral, 2005).

Fährt man auf der nächsthöheren Skala fort, so trifft man auf Gruppen von 3 bis 4 Knoten, welche als Subgraphen oder, falls sie häufiger als in Zufallsgraphen vorliegen, als Netzwerk motive bezeichnet werden. Diese Subgraphen lassen sich nach ihrer exakten internen Vernetzung kategorisieren, und es hat sich gezeigt, dass manche von ihnen viel häufiger in biologischen Systemen vertreten sind, als durch eine zufällige Vernetzung erklärt werden kann (Milo *et al.*, 2002, 2004; Alon, 2007). Diese Autoren argumentieren, dass bestimmte Knotenkombinationen spezifischen funktionalen Kreisläufen entsprechen und daher überrepräsentiert sind (Milo *et al.*, 2004). Ob ein ausgesuchtes Motiv bzw. ein bestimmter Motivtyp

eine besondere Rolle in einem Netzwerk einnimmt, verlangt erneut eine Beurteilung im Zusammenhang mit seiner Umgebung.

Verlässt man diese Ebene, so betritt man die Mesoskala des Graphen. Hier erreicht das Zusammenspiel von Subgraphen bzw. einzelner Knoten bereits so komplexe Züge, dass zur topologischen Charakterisierung andere Größen herangezogen werden müssen. Ein bekanntes Maß, welches auf lokalen Unterschieden der Vernetzungsdichte basiert, ist die Modularität (Girvan & Newman, 2002; Rives & Galitski, 2003; Tanay *et al.*, 2004). Viele biologische Systeme, wie z.B. metabolische Netzwerke (Ravasz & Barabási, 2003; Guimerà & Amaral, 2005) und neuronale Netzwerke (Hilgetag *et al.*, 2000), zeichnen sich dadurch aus, dass die interne Vernetzung ihrer Subsysteme größer ist als die Vernetzung mit den anderen Subsystemen bzw. dem Rest des Graphen.

Auf der größten Skala betrachtet man schließlich das Netzwerk in seiner Gesamtheit, und eine Reihe globaler Kenngrößen stehen hier für die Charakterisierung zur Verfügung, allen voran die *Gradverteilung*, welche die Wahrscheinlichkeit beschreibt, dass ein bestimmter Grad in einem Graphen vorkommt (hier wurde eine lokale Größe zu einer globalen Größe extrapoliert). Mit der Einführung des skalenfreien (BA) Graphen durch Barabási & Albert (1999) wurde ein neuer Graphentyp definiert, der sich durch seine *Gradverteilung* auszeichnet und der in Struktur und in der Art der Erzeugung fundamentale Eigenschaften vieler biologischer Systeme berücksichtigt. Skalenfreie Graphen besitzen keinen typischen Knotengrad, sondern, ganz im Gegenteil, ein sehr breites Spektrum an Knotengraden. Besonders charakteristisch sind einige wenige Knoten mit extrem vielen Verbindungen und eine Vielzahl von Knoten mit geringem Knotengrad (siehe hierzu Kap. 2.2 und Kap.4.1.3). Weitere wichtige Größen sind der *Vernetzungsgrad* (die Zahl der Kanten bezogen auf die Systemgröße) und die *Gradkorrelationen* (die Systematik in den Graden benachbarter Knoten).

Zur Einordnung der Bedeutung struktureller Eigenschaften (mit den entsprechenden Größen zur Charakterisierung) in einem biologischen System müssen die dynamischen Anforderungen an die Topologie berücksichtigt werden. Die folgenden vier Sichtweisen sollen diesen Zusammenhang am Beispiel des Metabolismus demonstrieren:

- (1) Die Systemstabilität verlangt, dass Ausfälle einzelner Enzyme abgefangen werden können. Das skalenfreie Graphenmodell bietet hierfür gute Voraussetzungen, wie diverse Studien der Empfindlichkeit gegenüber Knotenläsionen gezeigt haben (Jeong *et al.*, 2001).
- (2) Die Zahl der Reaktionen, die letztlich benötigt werden, um eine Substanz in eine andere umzusetzen, sollte in Zeit- und Energieaufwand möglichst gering sein. Eine skalenfreie Architektur löst auch dieses Problem, da die Weglängen zwischen beliebigen Knotenpunkten im Vergleich zu anderen Architekturen drastisch reduziert sind.
- (3) Mit der rein räumlichen Trennung einzelner Reaktionen (z.B. durch Zellkompartimentierung mittels Zellorganellen) geht oft auch eine funktionale Trennung einzelner Stoffwechselwege einher. Stoffumwandlungen innerhalb einer biochemischen Klasse erfolgen aus graphentheoretischer Sicht oft in Modulen, mit einer Reihe von Reaktionen innerhalb eines Moduls, aber nur wenigen Reaktionen, welche die Modulgrenzen überschreiten und die einen Stoffwechselweg mit dem gesamten Metabolismus über ausgesuchte Substanzen verbinden. Solche Eigenschaften werden zunächst auf der Mesoskala des Systems beschrie-

ben. Gleichzeitig wird aber auch die besondere Rolle einzelner Knoten, die hier als zentrale Elemente innerhalb eines Moduls fungieren, deutlich.

(4) Darüber hinaus existieren zentrale Schaltstellen, welche als Bindeglied zwischen verschiedenen Stoffwechselwegen (z.B. den Modulen) fungieren. Bekannte Vertreter sind die Glykolyse und der Zitronensäurezyklus, deren Zwischenprodukte von großer Bedeutung im Katabolismus (dem Abbau von komplexen zu einfachen Substanzen) und dem Anabolismus (der Synthese komplexer Substanzen) sind. Auch hier besitzen einzelne Metaboliten (wie z.B. Pyruvat oder Acetyl-CoA) eine besondere Stellung, da sie Ausgangspunkt verschiedenster Synthesewege sind.

Die hier exemplarisch dargestellten Ansprüche erfordern die Zusammenfügung unterschiedlicher topologischer Eigenschaften, die sich auf den ersten Blick teilweise auszuschließen scheinen, wie z.B. modulare Strukturen, welche die Weglängen beträchtlich erhöhen. Mit der Einführung des Hierarchiekonzeptes durch Barabási *et al.* (2001); Ravasz & Barabási (2003) konnte dieser Widerspruch jedoch entkräftet werden (siehe auch Ravasz & Barabási (2003)). Netzwerke mit einer hierarchischen Struktur besitzen Module unterschiedlicher Größe auf verschiedenen Skalen, d.h. die modularen Elemente sind ihrerseits aus Modulen zusammengesetzt usw., und gleichzeitig sind die Weglängen vergleichsweise gering. Verschiedene Graphen dieses Typs stehen zum gegenwärtigen Zeitpunkt im Mittelpunkt der Diskussion, da sie mehrere essentielle Eigenschaften realer biologischer Systeme vereinen. Typische Varianten basieren auf der Sichtweise verschachtelter Module (Ravasz & Barabási, 2003; Kaiser *et al.*, 2007a), andere Varianten stellen die Koexistenz von Modulen und zentralen Knoten in den Mittelpunkt (*Hubs*) (Guimerà & Amaral, 2005) oder beinhalten zusätzlich eine Hierarchie der *Hubs* (Ravasz & Barabási, 2003).

Der Exkurs in den Metabolismus machte deutlich, dass Topologie und Dynamik in biologischen Systemen nicht getrennt betrachtet werden können. Die endliche Größe biologischer Systeme grenzt den Spielraum für dynamische Prozesse ein; funktionale Netzwerke beruhen nicht auf einzelnen Interaktionen, sondern auf funktionalen Zusammenhängen, wie z.B. korreliertem Verhalten einzelner Elemente (hier formt der dynamische Prozess die Netzwerktopologie). Es ist daher nicht nur interessant zu sehen, welche strukturellen Eigenschaften (auf den entsprechenden Größenskalen) ein System prägen, sondern auch wichtig zu verstehen, auf welcher Größenskala ein System funktioniert bzw. wie sich Dynamik auf verschiedenen Größenskalen abbildet. Doch welche ist die richtige Dynamik, um ein System zu beschreiben? Will man einen realen Prozess möglichst genau imitieren, oder will man bestimmte Eigenschaften eines Netzwerks dynamisch sondieren? Der exakte Nachbau eines Prozesses beinhaltet eine detaillierte Beschreibung einzelner elementarer Wechselwirkungen unter Verwendung einer Vielzahl von Parametern. Die mit der Systemgröße rapide anwachsende Menge von Parameterabhängigkeiten macht solche Vorgehensweisen nur interessant für kleinere Subsysteme. Auf Systemebene hingegen muss die Zahl der beeinflussenden Parameter entweder drastisch auf die Menge der relevanten Größen reduziert werden oder man bedient sich sogenannter dynamischer Sonden, die nicht unbedingt durch die Eigenschaften des realen Prozesses motiviert sein müssen, aber dennoch in der Lage sind, charakteristische Wechselwirkungen in einem System aufzudecken. Die Wahl eines geeigneten dynamischen Prozesses hängt letztendlich von der Problemstellung, aber auch von der Frage ab, ob eine Dynamik eher durch die Systemstruktur oder

die dynamischen Parameter beeinflusst wird.

Anhand verschiedener Beispiele sollen völlig unterschiedliche Anforderungen an das dynamische Modell verdeutlicht werden.

(1) Hierzu soll zunächst das zu Beginn angesprochene Phänomen der Musterbildung auf Platinoberflächen genauer betrachtet werden. Voraussetzung für die Entstehung der Muster ist ein zyklischer Wechsel zwischen verschiedenen Zuständen, die von der Katalysatoroberfläche eingenommen werden können und die zu unterschiedlichen Affinitäten für die Reaktanden und schließlich zu deren Umsetzung führt. Bei geschickter Kombination der äußeren Parameter (z.B. Temperatur bzw. Partialdruck der Reaktanden) werden in diesem System räumliche Oszillationen in Form von Wellen, Kreiswellen und Spiralwellen erzeugt. Für die Modellierung dieser erregbaren Medien stehen verschiedene Methoden der Mathematik zur Verfügung. Einen möglichen Ansatz zur Beschreibung dieser unterschiedlichen Typen der Musterbildung bieten bereits einfache dynamische Regelwerke, wie die zellulären Automaten. In diesen Modellen ist der Zustand eines Knotens durch die Zustände der direkten Nachbarn (4 Nachbarn in einem regulären zweidimensionalen Gitter), den eigenen Zustand und eine bestimmte Regel definiert. Ergänzt man dieses deterministische Regelwerk durch weitere Komponenten (die zufällige Auswahl von Knoten, die als Initiationspunkte von Oszillationen dienen, bzw. die zufällige Dauer eines bestimmten Zustandes, den ein Knoten einnehmen kann), dann ist man bereits in der Lage, das Auftreten dieser Oszillationen und einige statistische Eigenschaften zu beschreiben.

(2) Völlig andere Anforderungen an das dynamische Modell stellt hingegen die detaillierte Betrachtung eines einzelnen Prozesses. Diese verlangt im ersten Beispiel eine Beschreibung des Reaktionsmechanismus auf atomarer Ebene unter Einbeziehung der Konformationsänderungen in der Katalysatoroberfläche. Eine ähnliche Vorgehensweise kann bei der Modellierung von biologischen Subsystemen eine Rolle spielen. Hierzu soll die kurze Folge von Reaktionen der Glykolyse betrachtet werden. Diese repräsentiert eine der zentralen Schlüsselstellen des Metabolismus mit einem klar definierten Signaleingang (Glukose) und einem Endprodukt (Pyruvat), welches Ausgangsstoff für viele weitere Reaktionen ist. Wenn man im Detail verstehen möchte, wie ein solches System beispielsweise auf eine kontrollierte Zufuhr von Glukose reagiert, so erfordert diese Fragestellung die Analyse mit einem komplexeren dynamischen Modell, welches eine Vielzahl kinetischer Parameter berücksichtigt.

Für das Themenfeld genregulatorischer Modelle bietet der kurze Übersichtsartikel von Bornholdt (2005) einen guten Überblick über die entsprechenden Herangehensweisen.

Im Vergleich zu den Studien, in denen Netzwerke verschiedenster Systeme aus den unterschiedlichsten Organismen rekonstruiert worden sind, ist die Menge experimenteller Daten, welche dynamisches Verhalten auf systemischer Ebene repräsentieren, vergleichsweise gering (Weckwerth & Fiehn, 2002; Emmerling *et al.*, 2002). Im Mittelpunkt dieser Untersuchungen stehen meist funktionale Zusammenhänge zwischen einzelnen Komponenten (z.B. die Koregulation von Genen oder die Korrelation zwischen einzelnen Metaboliten), deren Ursache in einer entsprechenden Netzwerkrepräsentation des Prozesses zu suchen ist. Hier bietet sich unter Umständen der Umweg über sogenannte Korrelations-Netzwerke an, die auf der Basis von Koexpressions- oder Korrelations-Daten erstellt werden (Stuart *et al.*, 2003; Kose *et al.*, 2001; Weckwerth & Fiehn, 2002).

Diese Arbeit wurde im Rahmen des Graduiertenkollegs 340 (Kommunikation in biologischen Systemen: Vom Molekül zum Organismus in seiner Umwelt) erstellt. Kommunikation findet in biologischen Systemen auf den verschiedensten Größen- und Zeitskalen statt. Im Graduiertenkolleg 340 wurden diverse Projekte behandelt, die von der Interaktion zwischen Organismen bzw. zwischen Organismus und Umwelt bis hin zur Kommunikation auf molekularer Ebene (z.B. im Sinne von wechselseitigen Beziehungen bzw. Einflüssen, Signaltransduktion oder Musterbildung) reichen. Neben dem biologisch-experimentellen Feld wurden auch auf theoretischer Ebene mathematische Modelle zur Beschreibung biologischer Prozesse untersucht. Zum Beispiel kann die mathematische Beschreibung von Mustern, die aus der Interaktion vieler einzelner Elemente resultieren, Aufschlüsse über den systemischen Zustand eines Organismus geben. Der Fokus dieser Arbeit liegt auf der Beziehung zwischen der Struktur und den dynamischen Eigenschaften von biologischen Systemen und von abstrakten Graphen, welche unter Ausnutzung aller topologischen bzw. dynamischen Skalen (welche die jeweilige Graphenabstraktion erlaubt) analysiert werden soll. Zur Sondierung dieser Eigenschaften wurde ein etabliertes erregbares dynamisches Modellsystem verwendet, in dem Anregungen zufällig erzeugt werden und nach einfachen Regeln durch das System propagieren. An verschiedenen typischen abstrakten Graphenarchitekturen wurde zunächst untersucht, in welcher Weise und auf welchen Skalen sich Dynamik abbildet, welche Rolle die dynamischen Parameter spielen und welchen Einfluss verschiedene Aspekte einer Netzwerktopologie haben. Auf dieser Basis wurden verschiedene biologische Systeme unter besonderer Berücksichtigung ihrer funktionalen Ansprüche charakterisiert. Schließlich erfolgte eine Analyse experimenteller Daten, die (als Abbild von Dynamik) bei der Beantwortung der Frage helfen soll, inwieweit die isolierte Betrachtung einer Netzwerkarchitektur zur Klärung experimenteller Befunde beitragen kann.

In dieser Arbeit soll zunächst ein Überblick über die relevanten topologischen Eigenschaften von Graphen (Kap. 2) und das spezielle dynamische Modellsystem (Kap. 3), welches den Ergebnissen der ersten beiden zentralen Untersuchungen zugrunde liegt, gegeben werden. Eine Aufstellung der prinzipiellen Graphentypen bzw. der realen Netzwerke erfolgt in einem separaten Kapitel im Anschluss (Kap. 4). Die zentralen Ergebnisse dieser Arbeit sind in 3 Teile untergliedert, die wie folgt aufgebaut sind:

Im ersten Ergebnisteil (Kap. 5) wurden die dynamischen Eigenschaften von drei verschiedenen basalen Graphentypen auf der größten Skala des Systems, der Graphengröße, untersucht. Neben dem Einfluss der prinzipiell unterschiedlichen Architekturen sollen insbesondere die Auswirkungen unterschiedlicher Vernetzungsdichten und Gradkorrelationen anhand der Fähigkeit des Systems, Anregungen zu organisieren, numerisch wie analytisch untersucht werden. Diese Ergebnisse wurden auf der Grundlage dynamischer Musterbildung analysiert. In einem weiteren Schritt sollen die regulierenden Eigenschaften der Kontrollparameter des dynamischen Modells, insbesondere die Abhängigkeit der Ergebnisse von intrinsischer Spontanaktivität, untersucht werden.

Der zweite Teil (Kap. 6) führt diese Untersuchung auf einer anderen Ebene weiter. Indem hier der topologische und dynamische Fokus auf die Mesoskala bzw. die kleinste Skala gelegt wurde, soll eine essentielle Eigenschaft vieler hierarchischer Graphen, die Koexistenz von modularen Elementen und zentralen einzelnen Knoten, analysiert werden. Im Mittelpunkt stehen hier die unterschiedlichen Ausprägungen kooperativen bzw. synchronisierten Verhaltens in Abhängigkeit von spontaner Aktivität in verschiedenen

Graphen, welche diese beiden Eigenschaften in unterschiedlichem Maße repräsentieren. Anhand von zwei neuronalen Netzwerken soll die dynamische Organisation bei unterschiedlichen topologischen Voraussetzungen zudem illustriert werden.

Der dritte Teil (Kap. 7) befasst sich mit der Analyse experimenteller Daten, die aus metabolomischen Studien stammen. Diese Untersuchung erfolgte auf der Ebene von Korrelations-Netzwerken (einer funktionalen Abstraktion dynamischer Prozesse), welche aus den Kovarianzen von Metabolitenkonzentrationen verschiedener Replikate gewonnen wurden. Zunächst soll die Konsistenz der Daten mit graphentheoretischen Methoden geprüft werden. Hierfür kommt ein Verfahren in Frage, bei dem die Ähnlichkeit von Netzwerken auf der Basis der Paardistanzen analysiert wird. Durch Vergleich mit der realen Prozessstruktur, dem metabolischen Reaktions-Netzwerk, soll untersucht werden, ob die beobachteten Korrelationen zwischen den Metaboliten durch die Struktur einer zugrunde liegenden Stoffwechselkarte erklärt werden können, oder ob hierfür andere regulatorische Prinzipien in Frage kommen.

Kapitel 2

Eigenschaften von Graphen

Ein Graph besteht aus Knoten und aus Kanten (der Begriff Netzwerk wird in dieser Arbeit für reale Systeme verwendet). Die Kanten stellen die Verbindungen zwischen den Knoten her und weisen auf eine Beziehung bzw. potentielle Interaktion zwischen den beteiligten Knoten hin. Auf diese Weise entsteht eine formale Sprache, mit der sich eine Vielzahl realer Systeme beschreiben, analysieren und vergleichen lässt. Ein Graph ist vollständig beschrieben, wenn die Nachbarn jedes Knotens und die Richtung der Kanten bekannt sind. Die Verknüpfung von Knoten kann in einem Graphen entweder gerichtet oder ungerichtet erfolgen. Abb. 2.1a zeigt einen gerichteten Graphen mit 10 Knoten und 14 Kanten mit unterschiedlichen Interaktionsrichtungen (angedeutet durch die Pfeile; Knoten 7 wird hier über eine Kante von Knoten 3 beeinflusst, aber nicht umgekehrt; Knoten 4 und 5 beeinflussen sich gegenseitig über 2 Kanten). Des Weiteren sieht man, dass dieser Graph nicht verbunden ist, sondern aus zwei Fragmenten (farblich unterschieden) besteht. Die größte verbundene Komponente (Fragment1: Knoten 1-8) wird auch als *Giant Component* bezeichnet. Einige der im Folgenden vorgestellten Maße zur topologischen Charakterisierung lassen sich nur auf diesen Knotencluster anwenden, andere auf den gesamten Graphen. Eine gute Übersicht zu den hier vorgestellten topologischen Eigenschaften von Graphen geben die Veröffentlichungen von Strogatz (2001), Albert & Barabási (2002), Newman (2003a), Newman (2003b) und Barabási & Oltvai (2004).

2.1 Wichtige globale und lokale Eigenschaften: Knotenzahl, Kantenanzahl, Vernetzungsgrad, Grad und Clusterkoeffizient

Sind die Eigenschaften zur Beschreibung eines Graphen bekannt, dann lassen sich einige globale Parameter (Kenngrößen, die den gesamten Graphen betreffen) schnell ableiten, wie z.B. die Zahl der Knoten n (die der Größe eines Graphen entspricht) und die Zahl der Kanten m . Ein Maß für die Dichte der Vernetzung ist der Vernetzungsgrad z eines Graphen (im Folgenden werden synonym auch Vernetzungsdichte und Konnektivität verwandt), welcher sich aus der Knoten- und Kantenanzahl berechnen lässt: $z = 2m/(n(n - 1))$ für ungerichtete Graphen bzw. $z = m/(n(n - 1))$ für gerichtete Graphen. Das Graphenbeispiel in Abb. 2.1a besitzt $n = 10$ Knoten und $m = 14$ Kanten und hat demnach einen Ver-

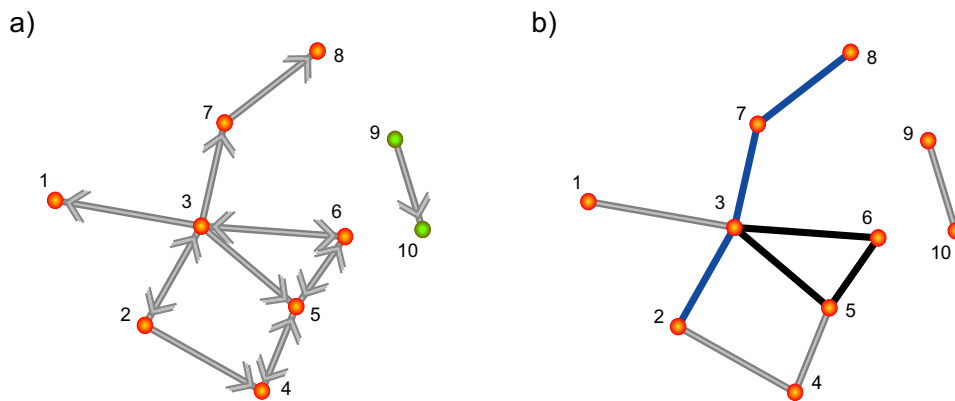


Abbildung 2.1: Basale Grapheneigenschaften. (a) Der unverbundene gerichtete Graph (aus zwei Fragmenten) besteht aus 10 Knoten (rot) und 14 gerichteten (grau mit Pfeilen) Kanten. (b) Diese ähnliche ungerichtete Fassung des Graphen aus a) mit 10 Kanten zeigt zwei wichtige Eigenschaften. Die kürzesten Pfade charakterisieren den kleinsten Abstand zwischen zwei Knoten (hier blau dargestellt für das Knotenpaar 2 und 8), während die Kantendreiecke (schwarz) für die Berechnung des Clusterkoeffizienten herangezogen werden. Hier besitzen nur die Knoten 3, 5 und 6 einen Clusterkoeffizienten $c > 0$.

netzungsgrad von $z = 14/90 = 0.16$.

Der Grad k_i eines Knotens i ist definiert als die Summe der Kanten m_i , die diesen Knoten mit dem Rest des Graphen verbinden. In gerichteten Graphen wird zwischen ein- und ausgehenden Kanten unterschieden; diese tragen zum Grad eines Knotens gleichermaßen bei wie eine ungerichtete Kante (Knoten 4 in Abb. 2.1a und b hat beispielsweise einen Grad von $k_4 = 2$). In dieser Arbeit wird als Größe häufiger das Kanten/Knoten-Verhältnis m/n eines Graphen (als Alternative zum Vernetzungsgrad z) verwendet; dieser Parameter skaliert mit dem durchschnittlichen Grad $\langle k \rangle$ eines Graphen. Der Graph in Abb. 2.1b hat ein Kanten/Knoten-Verhältnis von $m/n = 1$ und einen durchschnittlichen Grad von $\langle k \rangle = 2$.

Der Clusterkoeffizient c ist eine knotenspezifische Größe, die die lokale Vernetzung um einen Knoten herum beschreibt. Zur Berechnung werden die Nachbarn eines Knotens gesucht, die auch untereinander vernetzt sind und mit diesem Knoten ein Kantendreieck bilden. Die Zahl der Kanten zwischen den Nachbarn eines Knotens wird auf die Zahl der Knotenkombinationen normiert, die prinzipiell ein solches Dreieck an diesem Knoten erzeugen können. Knoten 5 in Abb. 2.1b ist (zusammen mit Knoten 3 und 6) Teil eines solchen Kantendreiecks (schwarz hervorgehoben). Knoten 5 hat einen Grad von $k_5 = 3$ und könnte demnach maximal drei dieser Dreiecke erzeugen. Demnach beträgt der Clusterkoeffizient dieses Knotens $c_5 = 1/3$. Der durchschnittliche Clusterkoeffizient $\langle c \rangle$, ein weiteres globales von der Graphengröße abhängiges Maß, ist zusammen mit der Funktion $c(k)$ ein wichtiger Parameter zur Charakterisierung verschiedener Graphenarchitekturen (Barabási & Oltvai, 2004).

2.2 Gradverteilung und *Hubs*

Eine weitere wichtige Kenngröße für einen Graphen ist die Gradverteilung, mit der verschiedene basale Graphenarchitekturen unterschieden werden. Die Gradverteilung $p(k)$ ist durch die Wahrscheinlichkeit p , dass ein bestimmter Grad k im Graphen vorhanden ist, definiert.

Man unterscheidet beispielsweise die *Poisson*-Verteilung (mit einem schmalen Wertebereich für k), die δ -Verteilung (in der nur ein Grad vorkommt) und die skalenfreie Verteilung (Barabási & Albert, 1999). Letztere spielt in den meisten realen Netzwerken (darunter vielen biologischen) eine besondere Rolle und charakterisiert auch einige abstrakte Graphen, die in dieser Arbeit vorgestellt werden. In diesem Graphentyp sinkt die Wahrscheinlichkeit $p(k)$ der Interaktion eines Knotens mit k anderen Knoten einem Potenzgesetz $p(k) \propto k^{-\gamma}$ entsprechend ab. Besonders für $\gamma < 3$ treten als zentrale Eigenschaft solcher Netzwerke so genannte *Power-Law Tails* auf, das sind ein oder mehrere Knoten (*Hubs*), mit einem großen Anteil an der gesamten Vernetzung. Die Bedeutung des Parameters γ auf die Vernetzung der *Hubs* und auf die Gradverteilung wird u.a. detailliert in Barabási und Oltvai (2004) beschrieben. Interessant ist in diesem Zusammenhang die Entwicklung skalenfreier Architekturen in biochemischen Systemen, wie z.B. Protein-Interaktions-Netzwerken. Es wird vermutet, dass der Prozess der Gen-Verdopplung einen wichtigen Anteil an der Ausbildung der Skalenfreiheit hat. Infolge dieser Entwicklung werden auch die Genprodukte (d.h. die Proteine) vervielfältigt, die wiederum (als neue Knoten im System) bevorzugt Kontakte zu den Interaktionspartnern des Proteinoriginals herstellen. Dieser Wachstumsprozess, der auch als *Preferential Attachment* bezeichnet wird, führt bei ausreichender Systemgröße (genügend hohem n) zu skalenfreien Netzwerken (Barabási & Albert, 1999; Barabási & Oltvai, 2004).

Manche Graphen sind durch Knoten mit extrem hohem Grad, den *Hubs*, gekennzeichnet (Barabási & Oltvai, 2004). Solche *Hubs* spielen in realen Netzwerken eine entscheidende Rolle und bestimmen das Verhalten eines Netzwerkes fundamental (Barabási & Oltvai, 2004) (beispielsweise wurde in Protein-Interaktions-Netzwerken im Vergleich zu den restlichen Knoten eine dreifach höhere Relevanz für die *Hubs* bestimmt (Jeong *et al.*, 2001)). In Studien von Guimerà & Amaral (2005) wurden 3 unterschiedliche *Hub*-Typen charakterisiert. (1) *Hubs* innerhalb von Modulen (siehe z.B. *Community* Modell [Kap. 4.1.4]); (2) *Hubs* zwischen den Modulen (z.B. hierarchischer skalenfreier Graph [Kap. 4.1.5]); (3) *Hubs*, die mit allen Teilen eines Graphen verbunden sind.

Für diese Studie wurde die Definition des *Hub* geringfügig erweitert. Hier werden auch solche Knoten als *Hubs* bezeichnet, die im Sinne der unten aufgeführten Zentralitätsmaße eine besondere Rolle spielen.

2.3 Gradkorrelationen

Gradkorrelationen werden durch lokale Verknüpfungsschemata erzeugt. Sind in einem Graphen bevorzugt Knoten gleichen Grades miteinander verknüpft, so wird dieser als *assortativ* (positiv korreliert) bezeichnet, bei bevorzugter Verknüpfung unterschiedlicher Grade liegt ein *disassortativer* (antikorrelierter) Graph vor (Newman, 2002). Trusina *et al.* (2004) benutzten diese Charakterisierung, um hierarchische und anti-hierarchische Graphen zu unterscheiden (siehe hierzu auch Kap. 5.5). In skalenfreien Netzwerken macht

sich diese Eigenschaft insbesondere in einer Zusammenballung bzw. Trennung der *Hubs* bemerkbar (siehe Abb. 2.2). Barabási & Oltvai (2004) vermuten, dass viele biologische Netzwerke *disassortativ* aufgebaut sind, dass also *Hubs* nicht mit *Hubs* verbunden sind.

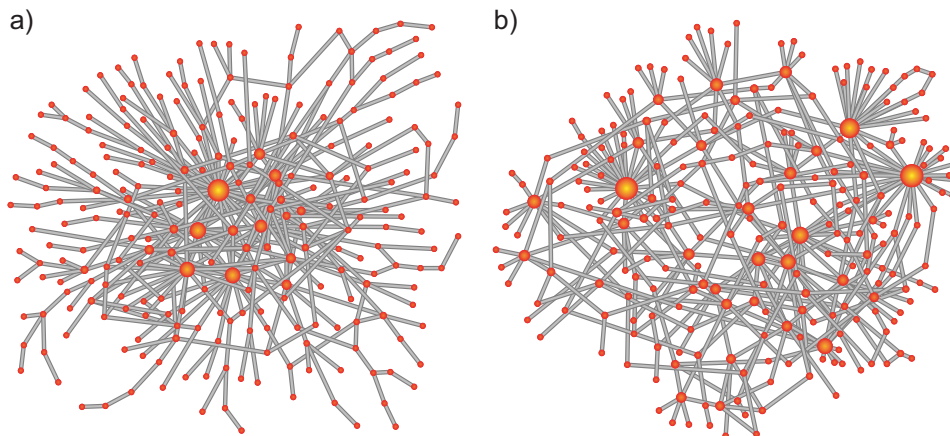


Abbildung 2.2: Unterschiedliche Gradkorrelationen in skalenfreien Graphen, die aus einem BA-Graphen (mit $n = 250$ Knoten und $m = 350$ Kanten) durch entsprechende Randomisierungen (Kap. 4.3; Prozess 3c und 3d) erzeugt wurden. Der Grad eines Knotens ist hier durch seine Größe gekennzeichnet. (a) Im hierarchisierten (positiv korrelierten) Graph liegen die *Hubs* (die größten Knoten) vernetzt im Zentrum des Graphen vor. (b) Im antihierarchisierten (negativ korrelierten) Graph sind die *Hubs* durch Knoten niedrigen Grades von einander getrennt.

2.4 Modularität

Die Modularität bezeichnet die strukturellen Auswirkungen, die eine heterogene Verteilung der Kanten innerhalb von Graphen haben kann. Die hierdurch erzeugten Subgraphen (Module) besitzen im Vergleich zur Umgebung eine deutlich höhere interne Vernetzung (Girvan & Newman, 2002; Newman & Girvan, 2004). Abb. 2.3 zeigt zwei Graphen, die sich vor allem in ihren modularen Eigenschaften deutlich voneinander unterscheiden. Beide Graphen bestehen aus $\mu = 3$ Modulen, welche durch eine ähnliche interne Vernetzungsdichte charakterisiert sind. Die Modularität des Graphen in Abb. 2.3a ist bezüglich dieser drei Module um ein vielfaches höher (gekennzeichnet durch die geringe Kantenzahl zwischen den Modulen) als die des Graphen in Abb. 2.3b, bei dem die Modulgrenzen bereits zu verschwimmen beginnen. Die Identifikation der Module bzw. die Quantifizierung der Modularität stellt eine große Hürde dar und wurde bereits in einer Vielzahl von Arbeiten ausführlich diskutiert (Clauset *et al.*, 2004; Donetti & nos, 2004; Radicchi *et al.*, 2004; Reichardt & Bornholdt, 2004; Newman & Girvan, 2004; Guimerà & Amaral, 2005; Bagrow & Bollt, 2005; Muff *et al.*, 2005). Für diese Arbeit spielt die exakte Identifikation der Module eine untergeordnete Rolle (siehe hierzu auch die Erläuterungen in Kap. 6.1). Zur Berechnung der Modularität M wird das Verhältnis von Kanten innerhalb eines Moduls zu den Kanten, welche die Knoten innerhalb dieses Moduls mit dem gesamten Graphen verbinden, herangezogen. Voraussetzung für diese Methode ist die Kenntnis der einzelnen Module (Details zur exakten Berechnung der Modularität M sind z.B.

Newman & Girvan (2004) zu entnehmen.). Die Beispielgraphen in Abb. 2.3 besitzen eine Modularität von $M = 0.92$ für (a) bzw. 0.6 für (b) bei $\mu = 3$ vorgegebenen Modulen (Werte für M liegen zwischen 0 (keine Modularität) und 1 (stärkste Modularität)). Die Bedeutung modularer Strukturen in biologischen Systemen wird z.B. in Proteininteraktions-Netzwerken deutlich, wo bestimmte Funktionalitäten durch den gleichzeitigen Kontakt vieler Proteine erreicht werden, oder in metabolischen Netzwerken, in denen einzelne Stoffwechselwege deutlich voneinander getrennt vorliegen (siehe auch die Arbeiten von Rives & Galitski (2003); Tanay *et al.* (2004); Han *et al.* (2004); Yook *et al.* (2004); Guimerà *et al.* (2004)). Synonym für die modularen Eigenschaften eines Graphen wird in dieser Arbeit auch der Begriff des Clustering verwendet (nicht zu verwechseln mit dem Clusterkoeffizienten, welcher lokale Knoteneigenschaften beschreibt).

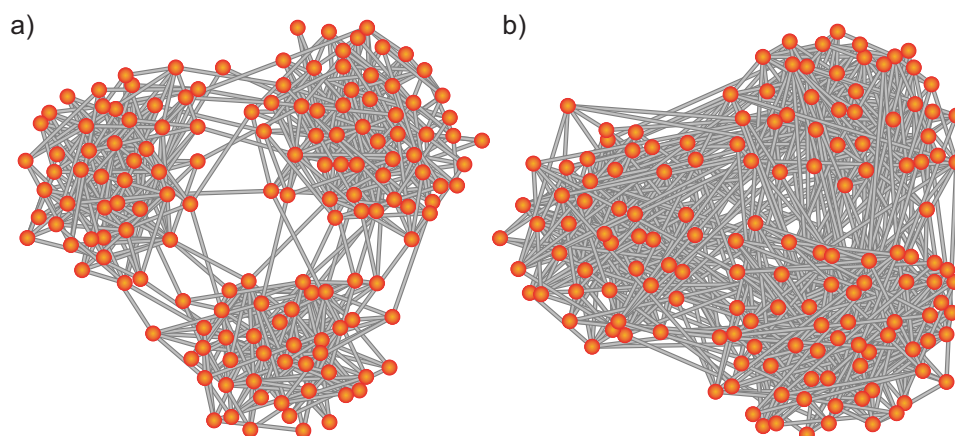


Abbildung 2.3: Modulare Eigenschaften von Graphen. Beide Graphen unterscheiden sich nur hinsichtlich der Zahl der Kanten, welche die drei Module miteinander verbinden. (a) Die deutlich geringere Zahl intermodularer Kanten führt zu einer stärkeren Trennung der ($\mu = 3$) Module und einer hohen Modularität $M^{(\mu=3)} = 0.92$ (Erläuterungen zur Berechnung von M finden sich im Text). (b) In diesem Graphen wurde die Zahl intermodularer Kanten verfünffacht, wodurch sich die einzelnen Module schwerer unterscheiden lassen und gleichzeitig die Modularität auf $M^{(\mu=3)} = 0.62$ abfällt.

2.5 Längenmaße

Die Weglänge (oder kürzester Pfad) L_{ij} zwischen zwei Knoten i und j bezeichnet die minimale Zahl der Kanten, die zwei Knoten verbindet, wobei es keine Rolle spielt, ob für eine Knotenkombination mehrere kürzeste Pfade existieren. Für den ungerichteten Graphen in Abb. 2.1b gibt es exakt einen kürzesten Pfad zwischen den Knoten 2 und 8 (blau markiert). Die Weglänge für dieses Knotenpaar beträgt 3. Für die Knotenkombination 2 und 9 ist hier keine Weglänge definiert. In gerichteten Graphen kann es für ein Knotenpaar unterschiedliche Weglängen für Hin- und Rückrichtung geben. In Abb. 2.1a ist z.B. die Weglänge für die Knotenkombination 2 und 8 nur für eine Richtung definiert. Eine weitere globale Größe, die mittlere Weglänge, errechnet sich entsprechend aus kürzesten Pfaden aller Knotenpaare. Sie skaliert ebenfalls mit der Systemgröße (Barabási & Oltvai, 2004).

Der Durchmesser eines Graphen ist der längste aller kürzesten Pfade in einer Architektur. In dem Beispiel aus Abb. 2.1b lässt sich der Durchmesser nur für die *Giant Component* bestimmen. Der Durchmesser von 4 ist hier durch den Abstand zwischen Knoten 4 und Knoten 8 gegeben. Viele reale Netzwerke besitzen trotz ihrer Größe einen geringen Durchmesser. Der Durchmesser in diesen (fast immer) skalenfreien Netzwerken hängt im Wesentlichen von der Verteilung der *Hubs* ab. Für bestimmte Netzwerke konnte gezeigt werden, dass sich der Durchmesser bereits verdoppelt, wenn nur 5% der Knoten (mit den höchsten Graden) entfernt werden (Albert *et al.*, 2000)

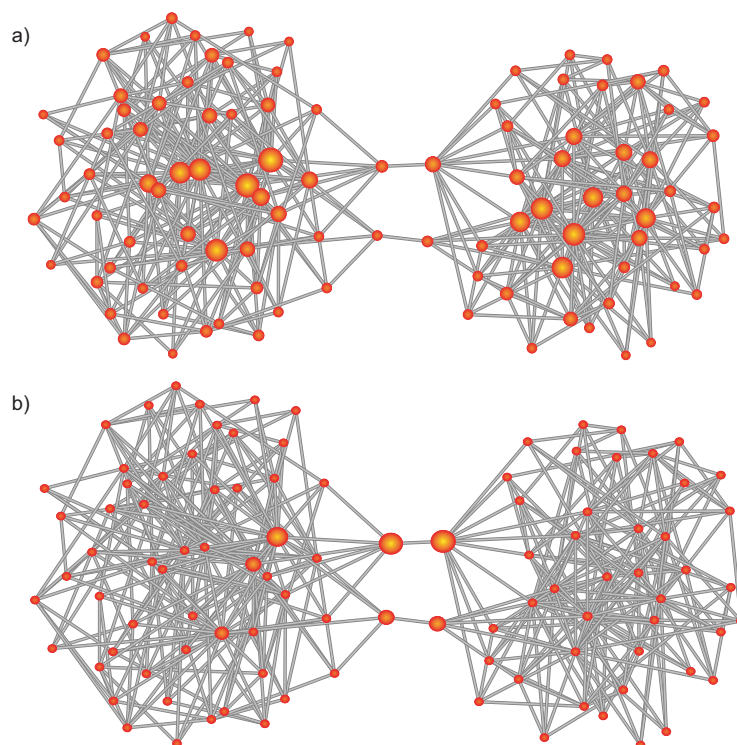


Abbildung 2.4: Zentralitätsmaße in modularen Graphen (die Größe der Knoten codiert die jeweilige Zentralitätsstärke). Durch die Grad-Zentralität (a) werden die zentralen Knoten innerhalb der Module identifiziert, während die *Betweenness*-Zentralität (b) globale Grapheneigenschaften berücksichtigt und die Verknüpfungspunkte (das sind hier Knoten mit deutlich niedrigerem Grad) zwischen den Modulen als zentrale Elemente herausstellt.

2.6 Zentralitätsmaße

Für die Bestimmung der Zentralität (Wichtigkeit bzw. zentrale Rolle) eines Knotens stehen verschiedene Maße zur Verfügung (Newman, 2003a). Die einfachste Methode, die als Maß den Grad k eines Knotens verwendet (siehe z.B. die *Hubs* im skalenfreien Graphen), beruht auf topologischen Grapheneigenschaften unterster Ordnung. Deshalb wurden andere Zentralitäts-Konzepte entwickelt, die den Knoten als Teil des gesamten Graphen betrachten. Die *Closeness*-Zentralität B_C (Wagner & Fell, 2001; Ma & Zeng,

2003) eines Knotens ist definiert als die mittlere Länge aller kürzesten Pfade, die diesen Knoten mit allen anderen Knoten verbindet. Kleine Werte deuten hier eine zentralere Lage dieses Knotens im Netzwerk an. Die *Betweenness*-Zentralität B geht hier noch einen Schritt weiter (Freeman, 1997; Freeman 1979) Freeman (1977, 1979). Sie spiegelt die Eigenschaft von manchen Knoten in einem Netzwerk wider, einen besonderen Einfluss auf dynamische Prozesse zu haben. Hier wird im Prinzip die Zahl der kürzesten Pfade (in einem Graphen) gemessen, die über einen bestimmten Knoten laufen. In dieser Arbeit wurde für die Bestimmung der *Betweenness*-Zentralität ein Algorithmus von Brandes (2001) verwandt, der für jedes Knotenpaar in effizienter Weise den kürzesten Pfad und die darin enthaltenen Knoten bestimmt.

Mögliche Unterschiede in der Identifikation zentraler Knoten werden anhand zweier dieser Zentralitätskonzepte und einem modularen Graphen in Abb. 2.4 erläutert. In dieser Darstellung sind die Stärken der jeweiligen Zentralitäten durch die Größe der Knoten codiert. In Abb. 2.4a erfolgt die Identifikation wichtiger Knoten durch lokale topologische Eigenschaften, den Knotengraden, und daher auf Subgraphenebene; die zentralen Knoten bilden das Zentrum der jeweiligen Module. Im Gegensatz dazu benutzt die *Betweenness*-Zentralität in Abb. 2.4b globale topologische Merkmale, wodurch die Knoten in den Schnittstellen zwischen den Modulen als zentrale Elemente identifiziert werden, obwohl sich diese eher durch niedrige Grade auszeichnen. Da viele biologische Systeme durch modulare Strukturen gekennzeichnet und Knoten mit niedrigem Grad prinzipiell als Schaltstellen zwischen Modulen vorstellbar sind, sollten strukturelle Analysen vorzugsweise mit der *Betweenness*-Zentralität durchgeführt werden.

2.7 Hierarchische Strukturen

Hierarchische Strukturen, die bei der Organisation biologischer Systeme eine wichtige Rolle spielen (Ravasz *et al.*, 2002; Sales-Pardo *et al.*, 2007), für die es aber bislang keine formale Definition gibt, können auf verschiedenen Ebenen in einer Graphenabstraktion existieren. Eine typische Sichtweise stellt die Organisation modularer Strukturen in den Mittelpunkt. Hier wird Hierarchie durch die Verschachtelung von Modulen unterschiedlichster Größe erzeugt (Ravasz & Barabási, 2003; Kaiser *et al.*, 2007a). Solche Graphenarchitekturen besitzen eine charakteristische gradabhängige Verteilung der Clusterkoeffizienten, welche, ähnlich zur skalenfreien Gradverteilung, einem Potenzgesetz mit $c(k) \propto k^{-\beta}$ folgt (Ravasz & Barabási, 2003). Diese Autoren fokussieren auf einen Graphentyp (der hierarchische skalenfreie Graph (Barabási *et al.*, 2001)), der zusätzlich durch Hubs gekennzeichnet ist, die ebenfalls in den Modulen bzw. den Hierarchieebenen hierarchisch verteilt vorliegen. In einem anderen Hierarchiekonzept stehen die Eigenschaften der kürzesten Pfade im Mittelpunkt. Die Systematik der Verteilung von Knotengraden in einem kürzesten Pfad führt zu einer Unterscheidung in hierarchische und antihierarchische Netzwerke. In diesen hierarchischen Netzwerken kommt es zu einer Häufung von hierarchischen Pfaden; das sind kürzeste Pfade, bei denen die Knotengrade, mit einem der Pfadenden beginnend, kontinuierlich ansteigen und zum anderen Pfadende hin abfallen, wobei ein abfallender Bereich nicht zwingend vorhanden sein muss.

Kapitel 3

Dynamische Modelle

Prozesse, wie die in der Einleitung vorgestellte durch Platin vermittelte Katalyse von Kohlenmonoxid zu Kohlendioxid, können zu einem oszillatorischen Verhalten und zu verschiedenen Typen raumzeitlicher Muster führen. Diese Organisationsprinzipien dynamischer Musterbildung sind keine Besonderheit, verursacht durch einen komplexen Reaktionsmechanismus, sie sind ganz im Gegenteil auch in vielen anderen chemischen Abläufen, wie z.B. der Belusov-Zhabotinski-Reaktion, zu beobachten, und sie charakterisieren ebenso diverse biologische Prozesse. Bekannte Beispiele sind die während der Selbstorganisation von Schleimpilzen (*Dictyostelium discoideum*) auftretenden Signalwellen, die Kalziumwellen im Zytoplasma der Oozyten des Krallenfroschs (*Xenopus sp.*) und die elektrischen Wellen am Herzmuskel.

Betrachtet man die CO-Katalyse, dann fällt auf, dass die Reaktion in drei Schritten erfolgt, die sukzessive durchlaufen werden und bei der jeder Schritt den nächsten bedingt. Diese für erregbare Medien typische Abfolge von Zuständen lässt sich als generelles Prinzip auch für die Sondierung komplexer Graphenarchitekturen verwenden. Eine einfache Variante eines solchen dynamischen Prinzips wurde unter dem Namen *Forest-Fire* Modell von Bak *et al.* (1990) eingeführt. Dieses Minimalmodell einer erregbaren Dynamik spielt in der Geschichte biologischer Modellierung von Prozessen der Signalpropagation und Musterbildung eine wichtige Rolle. Neben den dynamischen Regeln, die prinzipiell nur deterministisches Verhalten zulassen, beinhaltet dieses Modell einen stochastischen Parameter π , der die Verweildauer für einen der drei Zustände reguliert. Das Modell wurde durch Drossel & Schwabl (1992) um einen weiteren Parameter, die Rate spontaner Anregungen f , erweitert, welche im ursprünglichen Modell als *Blitzwahrscheinlichkeit* bezeichnet wurde. In dieser Form wurde es zunächst in Studien selbstorganisierter Kritizität (engl. self-organised criticality, SOC) verwandt, die an regulären gitterartigen Graphen durchgeführt wurden. Auf diesen Strukturen wird das dynamische Verhalten nur durch die beiden Modell-Parameter f und π bestimmt, deren Festlegung im Grenzbereich zu SOC- bzw. Perkulationsphänomenen führt. Dieses dynamische Verhalten machte es für die Untersuchung bestimmter Eigenschaften von Waldbränden (daher der Name) sowie für die Erforschung der Krankheiten- und Epidemieausbreitung (Jensen, 1998) interessant. Die Parallele zu anderen Epidemiemodellen, z.B. SIRS (Kuperman & Abramson, 2001) (eine zyklische Variante des SIR-Modells (Bailey, 1975; Anderson & May, 1991; Hethcote, 2000; Moreno *et al.*, 2002), ist offensichtlich. Neben den beiden bereits erwähnten dynamischen Eigenschaften wurden in einem breiten Parameterregime aber auch propagierende Wellen verschiedenen Typs beobachtet. In

diesem Bereich ist das Modell, bedingt durch die einfachen Nachbarschaftsregeln, welche die Weitergabe einer Anregung regulieren, auch für die Untersuchung von Netzwerken beliebiger Architektur interessant. In diesem Zusammenhang konnten Graham & Matthai (2003) beispielsweise zeigen, dass das Einfügen von *Shortcuts* (das sind Kanten, die deutlich getrennte Knoten und damit Graphenbereiche miteinander verbinden) in einer regulären Architektur dem dynamischen Effekt spontaner Anregung (also dem Parameter f) entspricht.

Die dynamische Untersuchung der in dieser Arbeit vorgestellten Graphenarchitekturen (Kap. 5 und Kap. 6) basiert auf diesem Modell, welches, mathematisch betrachtet, ein zellulärer Automat mit stochastischen Elementen ist. Es wird im Folgenden als diskret erregbares (DE) Modell bezeichnet. Für einige Graphen mit hoher Vernetzungsdichte wurde eine Variante mit einem zusätzlichen Parameter ξ eingeführt (siehe Kap. 6.5 und Kap. 6.6).

3.1 Das diskret erregbare (DE) dynamische Modell

Die dynamischen Regeln des DE-Modells sind denkbar einfach. Jeder Knoten kann drei diskrete Zustände (erregbar E , angeregt A und refraktär R) annehmen, welche zyklisch durchlaufen und synchron und in diskreten Zeitschritten nach folgenden Regeln aktualisiert werden (*Update*):

- (1) Ein erregbarer Knoten E wird angeregt, wenn sich mindestens ein angeregter Knoten A in direkter Nachbarschaft befindet, d.h. über eine Kante mit diesem verbunden ist. Ist dies nicht der Fall, so erfolgt spontane Anregung mit der Wahrscheinlichkeit f .
- (2) Ein angeregter Knoten A geht nach einem Zeitschritt automatisch in den refraktären Zustand R über.
- (3) Ein Knoten R regeneriert sich ($R \rightarrow E$) mit der Wahrscheinlichkeit π . Abb. 3.1 zeigt eine mögliche Propagation der Anregungen (A , rot), wenn ein Knoten (2. Bild, Knoten 7) zufällig angeregt wurde. Die Anregungen werden an die erregbaren Nachbarknoten (E , grün) weitergegeben, während die angeregten Knoten nach einem Zeitschritt in den refraktären Zustand (R , schwarz) übergehen.

3.2 Varianten des DE-Modells

Prinzipiell muss die Frage gestellt werden, ob das *Update*-Verfahren im DE-Modell Einfluss auf die dynamischen Eigenschaften eines Graphen hat, bedenkt man beispielsweise die Musterbildung der Anregungen bei niedrigen Werten der Spontanaktivität f (siehe Kap. 5.4). Die Simulation der in dieser Arbeit vorgestellten Graphenarchitekturen ist auch mit einem modifizierten DE-Modell mit einem sequentiellen *Update* vorstellbar, bei dem ein Knoten in jedem Zeitschritt aus einer Liste verfügbarer Knoten zufällig ausgewählt wird (Unter verfügbaren Knoten versteht man solche, mit denen noch kein *Update* durchgeführt wurde. Sind die verfügbaren Knoten aufgebraucht, so können wieder alle Knoten verwendet werden.). Die Untersuchung einiger basaler Graphenarchitekturen (z.B. ER-Graph und BA-Graph) hat

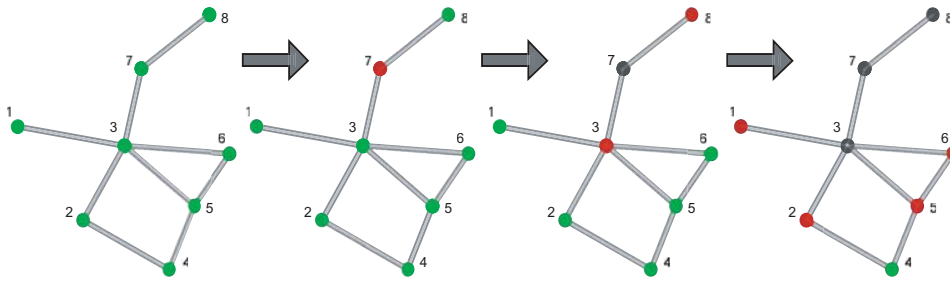


Abbildung 3.1: Ausschnitt aus einer Simulation mit dem DE-Modell. Die Zustände der Knoten sind farblich gekennzeichnet (grün: E ; rot: A ; schwarz: R ; von links nach rechts sind direkt aufeinanderfolgende *Updates* dargestellt). In einem vollständig regenerierten Graph (Bild 1) wird Knoten 7 zufällig angeregt (Bild 2). Diese Anregung wird auf die direkten Nachbarn im Zustand E übertragen (Bild 3 und 4), während nach einem Zeitschritt der angeregte in den refraktären Zustand wechselt.

gezeigt, dass bei genügend hoher Simulationsdauer keine nennenswerten Unterschiede hinsichtlich der dynamischen Musterbildung zu beobachten sind. Im Gegensatz dazu weist die Monte-Carlo-Methode - ein sequentielles *Update*-Verfahren, bei dem ein Knoten (ohne Nebenbedingung) zufällig ausgewählt wird - deutliche Unterschiede zu dem originalen Verfahren auf. Dies zeigt sich vor allem in dem Ausbleiben von wellenförmigen Propagationen, die das klassische *Forest-Fire* Modell charakterisieren.

Ein weiteres Problem stellt die Untersuchung von sehr dicht vernetzten Graphenarchitekturen dar. Bei einem hohen mittleren Grad führen die *Update*-Regeln auch bei starker Variation von f fast immer zu einer direkten (nicht-stochastischen) Anregung der Knoten im Zustand E . Da für einige der Untersuchungen der Einfluss des Parameters f auf die Dynamik eine entscheidende Rolle spielt, muss man die Bedingung für die Weitergabe von Anregungen entsprechend modifizieren. Dieses Problem wurde durch die Einführung des Parameters λ in das erregbare System gelöst. Dieser Parameter reguliert die Schwelle $\xi_i = k_i \lambda$, welche individuell für jeden Knoten i und in Abhängigkeit von seinem Grad k_i die Zahl der ankommenden Anregungen festlegt, die nötig sind, um ihn von dem erregbaren Zustand E in den angeregten Zustand A zu überführen (mit einem Minimalwert von $\xi = 1$).

Abb. 3.2 zeigt die dynamische Reaktion eines Graphen mit hoher Vernetzungsdichte ($z = 0.08$) auf die Variation des Parameters λ für verschiedene Werte von f anhand der durchschnittlichen Anregungsstärke, nämlich der Anregungsdichte ρ_A . Hierzu wurde ein BA-Graph mit $n = 250$ Knoten und $m = 2500$ Kanten ($m_A = 10$) erzeugt und ohne Schwelle ξ (dies entspricht $\lambda = 0$) bzw. mit Werten von $0 < \lambda \leq 0.15$ simuliert (je 10 Simulationen mit einer Simulationsdauer von $t_{max} = 25000$ und $\pi = 0.1$); die Spontanaktivität f wurde in einem Bereich von $10^{-4} \leq f \leq 0.1$ variiert. Die Anregungsdichte, welche der durchschnittlichen Zahl angeregter Zustände A (normiert auf die Netzwerkgröße) entspricht, ist ein einfaches dynamisches Maß zur Beurteilung der dynamischen Eigenschaften des Graphen auf Systemebene. Die Ergebnisse für das basale DE-Modell ($\lambda = 0$, rote Kurve in Abb. 3.2) zeigen deutlich, dass die Spontanaktivität f keinen Einfluss auf die Anregungsdichte ρ_A hat. Mit zunehmenden Werten von λ reagiert das System immer stärker auf Änderungen im Parameter f . Die grüne Kurve in Abb. 3.2 zeigt für den hier behandelten Graph einen idealen Wert von λ , bei dem erstens die deutlichsten Unterschiede in ρ_A zu

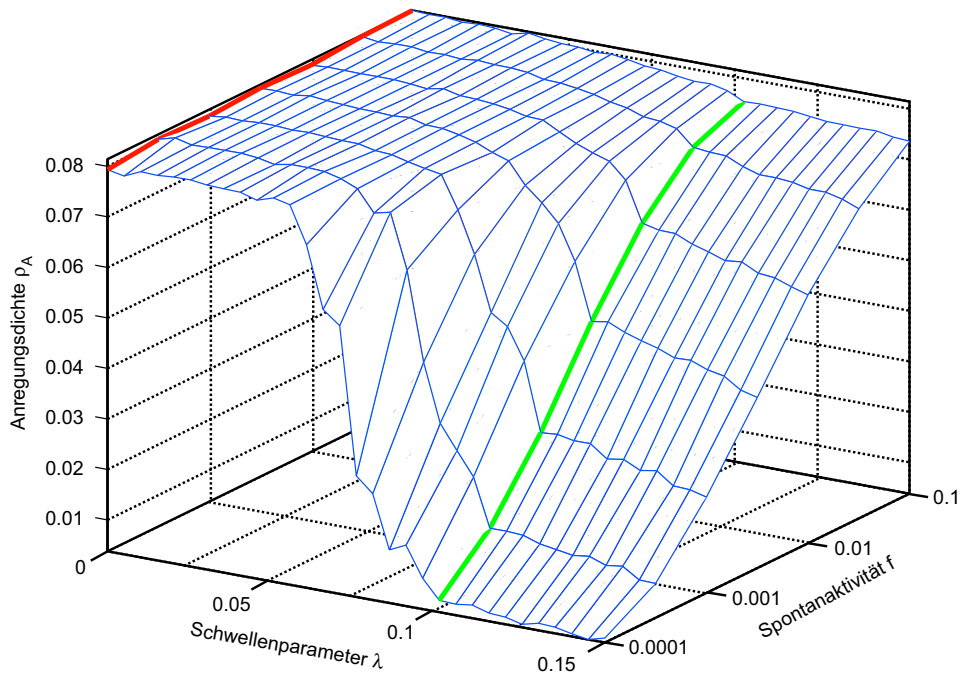


Abbildung 3.2: Dynamisches Verhalten eines dicht vernetzten BA-Graphen mit und ohne Schwellenparameter λ . Die rote Kurve steht für die Simulationsergebnisse des originalen DE-Modells ($\lambda = 0$). Es ist deutlich zu erkennen, dass die Anregungsdichte ρ_A selbst bei geringen Werten der spontanen Anregungsrate f unverändert hoch bleibt. Erhöht man den Schwellenparameter λ , so ändern sich die dynamischen Reaktionen, d.h. der Graph zeigt eine zunehmende Sensitivität gegenüber f . Die grüne Kurve bei $\lambda = 0.1$ markiert den Anfangsbereich von λ , ab dem ρ_A am stärksten durch die Spontanaktivität f beeinflusst wird.

sehen sind und zweitens das ursprüngliche DE-Modell einer nicht allzu starken Abwandlung unterzogen wurde. Dichter vernetzte Graphen wurden in dieser Arbeit mit Werten von $0.1 \leq \lambda \leq 0.15$ untersucht.

Kapitel 4

Graphen und Netzwerke

4.1 Abstrakte Graphen

In dieser Arbeit wird zwischen Graphen und Netzwerken unterschieden. Der Begriff "Graph", der im Prinzip die mathematische Bezeichnung für ein Netzwerk ist, wird hier ausschließlich für abstrakte Systeme verwendet, während der Begriff "Netzwerk" in dieser Arbeit für reale Systeme reserviert ist. Kap. 5 befasst sich ausschließlich mit einfachen abstrakten Graphen. In Kap. 6 werden komplexere abstrakte Graphenarchitekturen hinzugenommen und auch zwei neuronale Systeme vorgestellt, während sich Kap. 7 ausschließlich mit metabolischen Netzwerken befasst. Fast alle vorgestellten Analysen basieren auf verbundenen Graphen, in denen weder Selbst- noch Doppelkanten vorkommen. Daher wurden alle Graphen so erzeugt, dass sie die oben genannten Bedingungen erfüllen. Ausnahmen bilden die ER-Graphen und die metabolischen Korrelations-Netzwerke, die bei niedrigen Vernetzungsdichten fragmentiert sind.

4.1.1 Zufällige Graphen - das Erdős-Rényi (ER) Modell

Das Modell des zufälligen Graphen, welches von Erdős & Rényi (1959) eingeführt wurde, stellt eines der am besten untersuchten klassischen Graphenmodelle dar (siehe Abb. 4.1a). Entsprechend werden solche Graphen auch als Erdős-Rényi (ER)-Graphen bezeichnet. Im Mittelpunkt der Forschung der letzten Jahre standen insbesondere die Abhängigkeiten bestimmter Grapheneigenschaften, wie z.B. des Durchmessers, der mittleren Weglänge und der Verbundenheit, vom Vernetzungsgrad bzw. der Graphengröße.

Die Konstruktion eines ER-Graphen erfolgt evolutiv: Man beginnt mit n unverbundenen Knoten und generiert Kanten zwischen jedem Knotenpaar mit einer Wahrscheinlichkeit p . Für $p \ll 1$ sind ER-Graphen durch eine Poisson'sche Gradverteilung gekennzeichnet, welche (bei großer Knotenzahl) relativ schmal ist und dazu führt, dass nur geringe Abweichungen in den Knotengraden entstehen. Durch Variation von p können ER-Graphen unterschiedlichen Vernetzungsgrades erstellt werden, wobei niedrige p -Werte zu fragmentierten Graphen und hohe p -Werte zu fast vollständiger Vernetzung führen. Abb. 4.1a zeigt als weitere Eigenschaft von ER-Graphen das Phänomen, in einem bestimmten Bereich des Vernetzungsgrades z zu fragmentieren und gleichzeitig einen großen zusammenhängenden Cluster (*Giant Component*) zu bilden. Dieser Cluster wird von sehr kleinen Fragmenten bzw. vereinzelt Knoten begleitet.

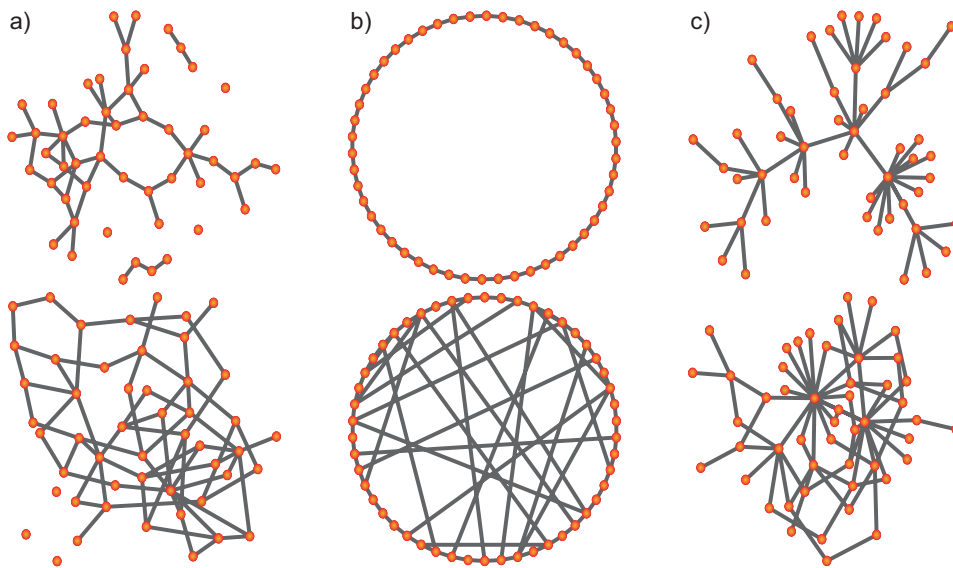


Abbildung 4.1: Drei basale Graphenmodelle in je zwei Ausführungen mit unterschiedlichem Vernetzungsgrad (obere Reihe: $z = 0.04$; untere Reihe: $z = 0.06$) (a). Der zufällige (ER)-Graph tendiert bei niedrigen Vernetzungsgraden zur Fragmentierung. Bei einem durchschnittlichen Grad von $\langle k \rangle = 2$ (oben) besitzt dieser Graphentyp bereits eine große *Giant Component* mit zyklisch vernetzten Elementen. (b) Das *Small-World* (WS)-Modell in der Variante mit Kantenaddition besteht bei $\langle k \rangle = 2$ (oben) aus einem einzigen Ring. Durch die Addition weniger Kanten (unten) werden der Durchmesser und die mittlere Weglänge drastisch herabgesetzt. (c) Der skalenfreie (BA)-Graph zeichnet sich durch eine geringe Zahl stark vernetzter und eine hohe Zahl schwach vernetzter Knoten aus und ist bei $\langle k \rangle = 2$ (oben) durch eine baumartige Topologie charakterisiert.

4.1.2 *Small-World* Graphen - das Watts-Strogatz (WS) Modell

Dieser Graph wurde von Watts & Strogatz (1998) zur Untersuchung der *Small-World*-Eigenschaften von Netzwerken eingeführt. Das Modell bietet die Möglichkeit, durch Variation des Parameters verschiedene p Graphenarchitekturen zu erzeugen, die zwischen den regulären Graphen (z.B. einem Ring von Knoten) und den zufälligen Graphen (z.B. ER-Graphen) liegen. In einem bestimmten Parameter-Regime sind solche Graphen durch drastisch reduzierte mittlere Weglängen bei gleichzeitig hohem mittleren Clusterkoeffizienten gekennzeichnet. Da diese *Small-World*-Eigenschaften auch in vielen realen Netzwerken gefunden wurden, ist diese Graphenklasse von hohem Interesse.

Die in Watts & Strogatz (1998) vorgestellte Variante zur Erzeugung eines WS-Graphen mit neu verknüpften Kanten beginnt mit einem regulären Graphen in Form einer kreisförmigen Kette, bei der jeder Knoten mit n_{WS} Nachbarknoten (im Uhrzeigersinn) verbunden ist. Der Parameter p gibt hier die Wahrscheinlichkeit an, dass ein Kantenende mit einem zufällig ausgewählten Knoten neu verknüpft wird. Mit zunehmender Randomisierungstiefe wird aus dem regulären ein zufälliger Graph. Eine Variation der Vernetzungsdichte ist hier nur bedingt möglich, indem man die Anfangsbedingungen (n_{WS}) ändert. Um

den Einfluss des Vernetzungsgrades genauer zu bestimmen, bietet sich eine andere Variante des WS-Modells an (Newman & Watts, 1999), bei der, ausgehend von einem regulären Graphen, Kanten mit einer Wahrscheinlichkeit p eingestreut werden. Der Parameter p reguliert hier den Vernetzungsgrad sowie die *Small-World*-Eigenschaften (Abb. 4.1b zeigt diese Variante des WS-Modells mit Kantenaddition).

4.1.3 Skalenfreie Graphen - das Barabási-Albert (BA) Modell

Untersuchungen an vielen realen Netzwerken zeigten, dass die Verteilung der Grade meist einem Potenzgesetz folgt. Barabási & Albert (1999) führten daher das Modell des skalenfreien Graphen ein (Abb. 4.1c). Solche Netzwerke zeichnen sich durch einige wenige Knoten mit extrem starker Vernetzung und einer Vielzahl an Knoten mit niedrigen Graden aus (siehe hierzu auch Kap. 2.2). Der evolutive Erzeugungsalgorithmus orientiert sich an der Entstehung realer Systeme, bei denen einzelne Komponenten nach und nach zugefügt und auf spezielle Art miteinander verbunden werden. Hierbei handelt es sich um so genanntes *Preferential Attachment*, also bevorzugtes Verknüpfen der hinzukommenden Knoten mit bereits bestehenden Knoten des Graphen, die schon viele Verbindungen aufweisen. Die Wahrscheinlichkeit einer Kante zwischen dem neuen Knoten und einem alten Knoten ist also in jedem Iterationsschritt proportional zum Grad des alten Knotens.

Die Erzeugung eines BA-Graphen beginnt in der Regel mit einem vollständig vernetzten Modul von $n_0 = 2$ oder 3 Knoten. In jedem iterativen Schritt wird ein neuer Knoten mit m_A Kanten bevorzugt an die Knoten addiert, welche den größten Grad aufweisen. Die Wahrscheinlichkeit einer Verknüpfung ist hierbei proportional zum Knotengrad. Da für einige der in dieser Arbeit vorgestellten Studien bestimmte Werte des Vernetzungsgrades z erforderlich sind, wurde der Erzeugungsalgorithmus des BA-Graphen so modifiziert, dass auch nicht-ganzzahlige Werte von m_A angenommen werden können. Die Nachkommastelle des Parameters m_A reguliert in jedem Iterationsschritt die Wahrscheinlichkeit, dass der nächst höhere (statt des nächst niedrigeren) ganzzahlige Wert von m_A angenommen wird.

4.1.4 Modulare Graphen - das *Community* Modell

Das *Community* Modell wurde von Newman & Girvan (2004) zur Untersuchung modularer Netzwerkstrukturen, so genannter *Communities*, eingeführt. In ihrer Version wird ein unverbundener Graph (aus n Knoten und $m = 0$ Kanten) vorgegeben und die Knoten auf μ Module aufgeteilt. Die Kanten werden mit zwei unterschiedlichen Wahrscheinlichkeiten, die im Prinzip das Verhältnis von inter- zu intramodularen Kanten sowie den Vernetzungsgrad regulieren, zufällig zwischen Knotenpaare eingestreut. Die in dieser Arbeit verwandten modularen Graphen basieren auf einer von Oh *et al.* (2005) eingeführten Variante des *Community* Modells, bei der Module mit skalenfreier Gradverteilung erzeugt werden.

Gemäß dem Algorithmus zur Erzeugung eines BA-Graphen werden μ vollständig vernetzte Module der Größe $n_0 = 2$ oder 3 vorgegeben. Alle weiteren Knoten werden mit m_A Kanten durch *Preferential Attachment* addiert, bis die finale Graphengröße n erreicht ist. Anschließend werden $m_E \cdot \mu(\mu - 1)/2$ Kanten zufällig zwischen den Modulen eingefügt. Ein entsprechender modularer Graph ist in Abb. 6.3 dargestellt.

4.1.5 Hierarchische Graphen

Die folgenden vier Graphenarchitekturen zeichnen sich alle durch eine hierarchische Organisation ihrer modularen Strukturen aus. Auch wenn der Hierarchiebegriff in Graphenstrukturen nach der Meinung mancher Autoren (Barabási *et al.*, 2001) nur für den hierarchischen skalenfreien Graphen reserviert ist, können hierarchische Elemente auf verschiedenen Ebenen existieren, weshalb die nun folgenden Graphen unter dieser Bezeichnung zusammengefasst sind. Darstellungen dieser Graphen sind in Abb. 6.10 und Abb. 6.11 zu sehen.

Hierarchische skalenfreie Graphen, die prominenteste Variante hierarchischer Graphen, wurden von Barabási *et al.* (2001) eingeführt, um drei wichtige Eigenschaften vieler realer Netzwerke, und zwar Modularität, hohes Clustering und eine skalenfreie Gradverteilung, in einem Modell zu vereinen. Gekennzeichnet sind solche Graphen durch eine komplexe Verschachtelung der Module auf verschiedenen Größenskalen. Die Erzeugung des Graphen basiert auf einem fraktalen Algorithmus, der durch Multiplikation und Vernetzung von bereits existierenden Graphenstrukturen ein deterministisches Netzwerk von hoher Selbstähnlichkeit produziert. Auffällig ist zum einen die starke Vernetzung der *Hubs* mit allen Knoten niedrigen Grades, wobei der Grad des zentralen *Hubs* besonders hoch ist, zum anderen die Vernetzung der *Hubs*, die sich auf derselben Hierarchieebene befinden.

Neben der geringen Variationsmöglichkeit bezüglich des Vernetzungsgrades z zeichnet sich der hierarchische skalenfreie Graph vor allem durch die extreme Dominanz seines zentralen *Hubs* aus. Da beide Eigenschaften für einen Teil der Untersuchungen ungeeignet sind, wurden im Rahmen dieser Arbeit zwei Varianten des hierarchischen skalenfreien Modells eingeführt, welche zum einen die Dominanz des zentralen *Hubs* reduzieren, zum anderen mehr Flexibilität bei der Konstruktion der Module sowie der Vernetzung erlauben.

Für **Variante 1** wurde ein weiterer Parameter, die Wahrscheinlichkeit g , eingeführt, welche insbesondere den Grad des zentralen Knotens k_h reguliert. Der Parameter g gibt die Wahrscheinlichkeit an, dass eine Kante zwischen dem zentralen *Hub* und einem anderen Knoten (gemäß Barabási *et al.* (2001); Ravasz & Barabási (2003)) gebildet wird. Während des Iterationsprozesses erhalten auch die Kopien die Eigenschaften des *Hubs* (in der existierenden Graphenstruktur), so dass im Prinzip die Dominanz des gesamten *Hub*-Systems verringert wird.

Die **Variante 2** greift bereits bei der Erzeugung der elementaren Module ein, welche in Barabási *et al.* (2001); Ravasz & Barabási (2003) aus 4 vollständig vernetzten Knoten bestehen. Dieses Modul kann durch einen BA-Graphen der Größe n_0 und der Kantenzahl $m = m_A n_0$ ersetzt werden. In jedem Iterationsschritt werden μ Kopien der existierenden Graphenstruktur erzeugt. Innerhalb einer hierarchischen Ebene wird die Vernetzung der Knoten mit den höchsten Graden durch den Parameter m_y reguliert, welcher im Prinzip die mittlere Zahl der Kanten zwischen jedem Modulpaar festlegt. Unter Modul versteht man in diesem Fall die Kopien, die bei jedem Iterationsschritt neu erzeugt werden. Die Vernetzung des zentralen Moduls, welches als Vorlage diente, mit seinen Kopien erfolgt über den zentralen *Hub* h und wird durch den Parameter m_h gesteuert, welcher die mittlere Zahl der Kanten zwischen dem zentralen Modul und jeder Kopie bestimmt. Durch Anpassung dieser Parameter in einem bestimmten Bereich

können die modularen Eigenschaften und die Dominanz der *Hubs* gesteuert werden, ohne die typische Steigung der skalenfreien Gradverteilung wesentlich zu verändern.

Hierarchische Cluster-Graphen, welche in einer Arbeit von Kaiser *et al.* (2007a) vorgestellt wurden, zeichnen sich durch hierarchisch angeordnete modulare Strukturen aus. Auf Grund der engen Gradverteilung (vergleichbar mit der Poisson-Verteilung der ER-Graphen) fehlen ihnen allerdings echte *Hubs*. Für die Konstruktion werden μ Cluster vorgegeben, die ihrerseits in μ Sub-Cluster unterteilt sind. Da jeder dieser Sub-Cluster μ Knoten beinhaltet, besitzt ein solcher Graph μ^3 Knoten. Die Verknüpfung der Knoten erfolgt für jede dieser drei hierarchischen Ebenen separat nach folgendem Schema: Innerhalb der kleinsten Ebene, also innerhalb der aus μ Knoten bestehenden Module, werden m_{H_1} Kanten zufällig eingestreut, innerhalb der mittleren Ebene m_{H_2} Kanten und schließlich m_{H_3} Kanten über den gesamten Graph. Durch entsprechende Einstellung dieser Parameter werden die hierarchischen Ebenen mehr oder weniger stark voneinander getrennt.

Fraktale modulare Graphen, welche als Modell für die Analyse des cerebralen Kortex dienen und von Sporns (2006) eingeführt wurden, sind ebenfalls durch mehrere hierarchische Ebenen charakterisiert. In der Variante als *Mapped Fractal Graph* unterscheidet er sich jedoch von dem hierarchischen Cluster-Graphen durch ein komplexes Verknüpfungsschema.

Der fraktale Erzeugungsalgorithmus verdoppelt die bereits bestehende Struktur und vernetzt beide Module so systematisch, dass die Kopien der einzelnen Knoten gemäß einer Normalverteilung am wahrscheinlichsten mit ihren *Originalen* verknüpft werden. Dadurch wird gewissermaßen eine räumliche Struktur imitiert, die nur bedingt das Überkreuzen von Verbindungen bzw. das weiträumige Überbrücken zulässt. Details zur Erzeugung sind Sporns (2006) zu entnehmen; die dort verwandten Parameter sind hier durch ein S gekennzeichnet. Der Parameter E_S reguliert den Vernetzungsgrad des Graphen, die Parameter m_S und n_S legen die Größe des ersten Subgraphen fest. Dieser Graphentyp ist durch eine äußerst enge Gradverteilung gekennzeichnet.

4.2 Reale Systeme

4.2.1 Neuronale Netzwerke

Für die Untersuchung neuronaler Systeme wurden die Datensätze von zwei Organismen, (1) der Katze und (2) des Fadenwurms (Nematoden) *Caenorhabditis elegans*, von Claus C. Hilgetag¹ zur Verfügung gestellt.

(1) Der erste Datensatz, welcher auf Systemebene die Verbindungen zwischen verschiedenen Bereichen des cerebralen Kortex der Katze beschreibt, basiert auf einer Zusammenstellung von 892 Verbindungen in 55 Regionen, welche aus einem Datensatz (beschrieben in Scannell *et al.* (1995)) entwickelt wurde. Für diese Abstraktion werden funktionale Bereiche des Gehirns, die eine Vielzahl von Nervenfasern (mit

¹C. C. Hilgetag, School of Engineering and Science, Jacobs University Bremen, 28759 Bremen

sehr unterschiedlicher Morphologie) enthalten können, zu Knoten und die entsprechenden Projektionen (in andere Hirnareale) zu Kanten zusammengefasst.

(2) Der zweite Datensatz, der die zelluläre neuronale Vernetzung des Nematoden *C. elegans* aus 277 Neuronen und 2105 synaptischen Verbindungen repräsentiert, wurde von Achacoso & Yamamoto (1992) übernommen. Auf dieser Abstraktionsebene wird jede Nervenfasern wie ein Knoten und die synaptische Verbindung wie eine Kante behandelt. Diese Zusammenstellung beruht im Wesentlichen auf dem Datensatz von White *et al.* (1986), für den die Verbindungen durch elektronenmikroskopische Identifikation rekonstruiert wurden. Die Vernetzungsdaten von Achacoso *et al.* (1992) wurden folgendermaßen modifiziert. Neuronen des pharyngalen Rings, für den keine interne Vernetzungsinformation vorliegt, wurden aus dem Netzwerk entfernt. Zusätzlich wurden drei Neuronen (AIBL, AIYL und SMDVL) entfernt, da für diese keine räumliche Information vorlag. Letztlich wurden 277 Neuronen in die Analyse einbezogen. Die Größe der globalen und lokalen Datensätze von *C. elegans*, die hier analysiert wurden, sind vergleichbar mit denen früherer Studien. Beispielsweise wurden für die Analyse der *Small-World* Eigenschaften (Watts & Strogatz, 1998) oder charakteristischer Motive (Reigl *et al.*, 2004) von *C. elegans* 282 und 187 Neuronen berücksichtigt. Sowohl chemische als auch elektrische Synapsen (*Gap Junctions*) wurden als Verbindungen in diese Analyse einbezogen.

4.2.2 Metabolische Reaktions-Netzwerke

Verschiedene Datenbanken stellen Informationen zu metabolischen Reaktionen zur Verfügung, die auf der Basis genomischer Analysen gewonnen wurden. Für die Rekonstruktion metabolischer Netzwerke wurde die *Kyoto Encyclopedia of Genes and Genomes* (KEGG) -Datenbank (Ogata *et al.*, 1999) bereits von vielen Forschergruppen verwendet. Die als relativ rauscharm geltende Klasse biochemischer Netzwerke stellt den Reaktionsraum für metabolische Transformationen. Bei der Rekonstruktion solcher Netzwerke sind im Wesentlichen zwei Herangehensweisen zu unterscheiden, die extrem häufige Verbindungen (engl. *currency metabolites*), wie H_2O , ATP und NADH, unterschiedlich beurteilen. In einem Reaktions-Netzwerk repräsentieren solche Metaboliten extrem wichtige *Hubs*, die die durchschnittliche Weglänge drastisch herabsetzen (siehe dazu auch die Diskussion zwischen Arita (2004) und Jeong *et al.* (2000)) und die konventionelle Ordnung der Reaktionen (wie man sie von den Stoffwechselkarten kennt) zerstören. Um diesem Effekt Rechnung zu tragen, rekonstruierten Ma & Zeng (2003) metabolische Netzwerke ohne jene Reaktionen, die einem bestimmten Typ, wie z.B. der Übertragung von Elektronen oder bestimmter funktioneller Gruppen, entsprachen. Zusätzlich korrigierten sie Fehler und Inkonsistenzen des Original-Datensatzes und überprüften zudem die Reversibilität bzw. Richtung jeder einzelnen Reaktion.

Die Rekonstruktion des metabolischen Netzwerks von *Arabidopsis thaliana*, welches dieser Untersuchung zu Grunde liegt, basiert ebenfalls auf der Datenbank von Ma & Zeng (2003).² In der *Verbindungs*-Tabelle sind 2070 verschiedene enzymkatalysierte Transformationen (unterteilt in reversible und irreversible Reaktionen) für 107 verschiedene Organismen zu finden. Eine Reaktion in dieser Auflistung repräsentiert eine potentielle Verbindung in einem metabolischen Reaktions-Netzwerk, während die Existenz dieser

²<http://genome.gbf.de/bioinformatics/index.html>

Reaktion für einen bestimmten Organismus durch einen numerischen Eintrag, der größer oder gleich 1 ist, angezeigt wird (dieser Wert steht für die Zahl der Gene, die für das entsprechende Enzym codieren). Das endgültige Reaktions-Netzwerk entsteht in drei Schritten: (1) Vernetzung aller Metaboliten in dieser Tabelle, (2) Transformation aller gerichteten zu ungerichteten Kanten und (3) Identifikation der größten verbundenen Komponente (*Giant Component* in einem fraktionierten Netzwerk). Die Umwandlung der gerichteten Netzwerkarchitektur in eine ungerichtete, welche im biologischen Sinne zu einer Reaktionskarte aus Gleichgewichtsreaktionen führt, ist hier unbedingt erforderlich, um auf der rein strukturellen Ebene einen Vergleich mit den Korrelations-Netzwerken zu ermöglichen. Auf diese Weise wurde ein Netzwerk mit 404 Knoten und 492 ungerichteten Kanten (aus ursprünglich 438 reversiblen und 332 irreversiblen Reaktionen) rekonstruiert.

4.2.3 Metabolische Korrelations-Netzwerke

Die Analyse der Korrelationsdaten und der entsprechenden Korrelations-Netzwerke basiert auf den experimentellen Studien von Weckwerth (2003); Morgenthal *et al.* (2005). Wolfram Weckwerth³ hat freundlicherweise die Konzentrationsdaten von 257 verschiedenen Metaboliten aus *Arabidopsis thaliana* (Acker-Schmalwand) zur Verfügung gestellt. Für diese wurden 10 Pflanzen zunächst für 26 Tage unter einheitlichen Bedingungen mit einem diurnalen Rhythmus von 8 h Licht und 16 h Dunkelheit kultiviert. Innerhalb eines Tages wurden zu 6 unterschiedlichen Zeitpunkten (nach 0.5, 4, 7.5 h Licht und nach 0.5, 8, 15.5 h Dunkelheit; siehe hierzu auch Abb. 7.3a) Pflanzenproben entnommen und auf ihren Metabolitengehalt hin analysiert (dies entspricht 60 Messungen).

Prinzipiell sind zwei praktikable Ansätze zur Rekonstruktion eines Korrelations-Netzwerkes aus diesen Daten denkbar. Betrachtet man Fluktuationen auf der langen Zeitskala, würde man die pflanzenspezifischen temporalen Veränderungen verwenden, die durch sechs Konzentrationswerte gegeben sind. In dieser Untersuchung soll allerdings eine statistisch verlässlichere Methode angewendet werden, welche die Fluktuationen auf einer kurzen Zeitskala (in Form der 10 Pflanzenproben) berücksichtigt. Bei der Verwendung von 6 Probenahmen (statt 10 Pflanzenproben) wird die statistisch verlässliche Region von κ (siehe unten) drastisch reduziert, so dass die untere Schwelle von κ (die für eine p -Statistik von 0.05 in etwa bei 0.82 liegen würde) in etwa mit der oberen Schwelle von κ , bei der die Korrelations-Netzwerke zu fragmentieren beginnen, übereinstimmen würde.

Korrelations-Netzwerke werden aus den gemessenen Metabolitendaten durch paarweise Kombination aller Metabolitenkonzentrationen erstellt. Eine Kante zwischen zwei Metaboliten wird unter der Bedingung erstellt, dass der Absolutwert des Korrelationskoeffizienten (der aus den Konzentrations-Vektoren beider Substanzen berechnet wird) eine bestimmte Schwelle κ übertrifft, d.h. positive Korrelationen und Antikorrelationen werden gleich behandelt. Abb. 4.2a stellt als Beispiel für diese Zusammenhänge die Vernetzung zwischen Glukose und Glukose-6-Phosphat (G6P) in einem der Korrelations-Netzwerke von

³W. Weckwerth, Max-Planck-Institut für Molekulare Pflanzenphysiologie, Wissenschaftspark Golm, 14476 Potsdam - Golm

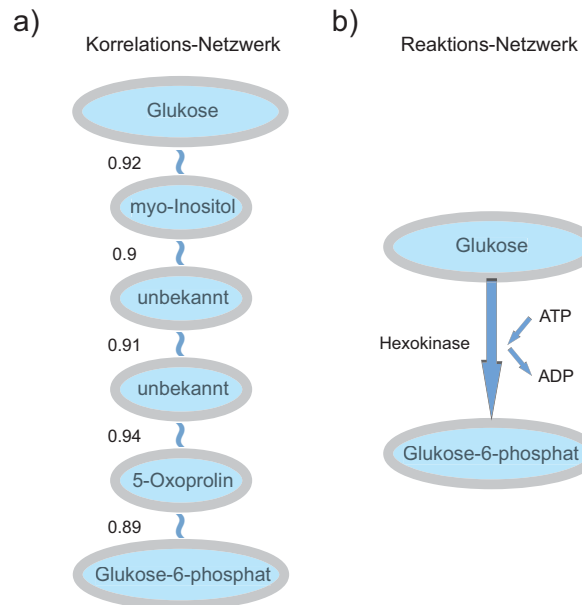


Abbildung 4.2: Beispiel für einen kürzesten Pfad in einem metabolischen Korrelations-Netzwerk und einem metabolischen Reaktions-Netzwerk von *A. thaliana*. Die Abweichung in den Pfadlängen in diesem Beispiel demonstriert die möglichen strukturellen Unterschiede zwischen diesen Netzwerkklassen. (a) In dem Beispiel eines schwach vernetzten (jedoch nicht fragmentierten) Korrelations-Netzwerkes (Parameterkonstellation: Probenahme bei 0.5 h Licht, $\omega = 8$ und $\kappa = 0.85$) hat der kürzeste Pfad zwischen diesem Metabolitenpaar eine Länge von 5 Kanten (die Zahl links neben den Kanten steht für die individuelle Korrelationsstärke der beteiligten Metabolitenpaare). (b) Die chemische Umwandlung von Glukose zu Glukose-6-Phosphat (in der Glykolyse) ist im Netzwerk der Reaktionswege durch eine Verbindung repräsentiert.

A. thaliana aus Weckwerth (2003); Morgenthal *et al.* (2005) dar. Im Reaktions-Netzwerk stellen diese beiden Metaboliten den ersten Reaktionsschritt der Glykolyse dar, der durch das Enzym Hexokinase unter Verbrauch von ATP katalysiert wird (Abb. 4.2b). Der Korrelationskoeffizient von Glukose und G6P liegt nahe bei Null, was bedeutet, dass es quasi keine Korrelation zwischen diesen Metaboliten unter den gegebenen Bedingungen gibt. Zieht man nur starke Korrelationen in Betracht, dann werden mehrere Kanten benötigt, um von einer Verbindung zur anderen zu gelangen. In dem Beispiel aus Abb. 4.2a (0.5 h Licht mit $\omega = 8$ und $\kappa = 0.85$) beträgt die Korrelationsstärke zwischen beiden Verbindungen 0.13. Zwei alternative kürzeste Pfade mit jeweils 5 Kanten existieren für diesen Fall. Einer der beiden wurde hier mit den dazu gehörigen Korrelationsstärken für jedes Metabolitenpaar dargestellt. Dieser Pfad beinhaltet einige Verbindungen, die im Experiment zwar gefunden wurden, bislang aber noch nicht identifiziert werden konnten.

Der Schwellenparameter κ reguliert den Vernetzungsgrad eines Korrelations-Netzwerkes. Für $\kappa = 0$ erhält man einen vollständig vernetzten Graphen, d.h. jeder Knoten ist mit jedem anderen Knoten verbunden. In manchen Messungen tauchen einige der Metaboliten nicht auf, wodurch Lücken in der Datenmatrix entstehen. Um diesen fehlenden Daten Rechnung zu tragen, wurde ein zweiter Parameter, die Qualitätsschwelle ω , in diese Analyse eingeführt, durch den die Menge der Metaboliten im Netzwerk auf jene

beschränkt wird, für die mindestens ω Meßergebnisse vorliegen. Auf diese Weise kann man die Verlässlichkeit der Daten, die in die Analyse eingehen, qualitativ variieren. Dieser Parameter wurde in einem Intervall von $4 \leq \omega \leq 8$ variiert, um die Robustheit der Resultate bezüglich der Auswahl der Daten zu testen.

4.3 Randomisierte Graphenstrukturen

Die Randomisierung von Graphen spielt (1) bei der Beurteilung bestimmter topologischer Eigenschaften bzw. bei der (2) zufälligen oder (3) gerichteten Umstrukturierung von Graphenarchitekturen eine wichtige Rolle. Bei allen Verfahren ist darauf zu achten, dass es nicht zur Ausbildung von Selbst-Kanten (also Kanten, die einen Knoten mit sich selbst vernetzen) oder Doppelkanten (zwei Kanten vernetzen dasselbe Knotenpaar) kommt. Darüber hinaus kann es prinzipiell zur Fragmentierung von Graphen kommen. Je nachdem, ob dies erwünscht ist, muss auch dieser Effekt berücksichtigt werden.

(1) Die Erzeugung von Null-Modellen einer Graphenarchitektur ist für solche Untersuchungen von Bedeutung, bei denen die Signifikanz verschiedener struktureller Merkmale bestimmt werden soll. Hier kommt es darauf an, basale Eigenschaften eines Graphen, wie die Knoten- und Kantenzahl sowie die individuellen Knotengrade, konstant zu halten. Dazu werden in einem iterativen Prozess zwei Kanten, die keine gemeinsamen Knoten haben, zufällig ausgewählt und neu vernetzt, so dass es zum Austausch von jeweils einem Kantenende kommt. Bei genügend großer Iterationszahl (Randomisierungstiefe) erzeugt man eine Graphenstruktur, dessen Eigenschaften sich durch weitere Randomisierung nur noch unwesentlich ändern. In dieser Arbeit wurde eine maximale Randomisierungstiefe von $20m$ angenommen.

(2) Das Verfahren aus (1) lässt sich auch schrittweise einsetzen, um die graduelle Veränderung einer bestimmten Grapheneigenschaft festzuhalten. Bei der in Kap. 7.2 eingesetzten Variante dieses Verfahrens wird nur eine Kante zufällig ausgewählt und an einem Kantenende mit einem zufällig gewählten Knoten neu vernetzt. Hierdurch verändern sich die individuellen Knotengrade.

(3) Bei einem gerichteten Verfahren sollen bestimmte topologische Eigenschaften erzeugt bzw. beeinflusst werden. Dazu kann man entweder nach jedem Iterationsschritt die Veränderung einer bestimmten Eigenschaft messen und je nach Ergebnis den Iterationsschritt zulassen, oder man wählt die zu vertauschenden Kantenenden nicht zufällig aus, sondern nach bestimmten Kriterien, wie z.B. dem Knotengrad der beteiligten Knoten. Dieses Verfahren kommt in dieser Arbeit in vier Varianten zu Anwendung:

In der Variante (a), die der Reduktion der Modularität dient, werden zwei Knotenpaare zufällig aus verschiedenen Modulen ausgewählt und neu vernetzt, so dass die Menge intermodularer Kanten sukzessive ansteigt. Um die Formation einer hierarchischen Struktur zu verhindern, erfolgt die Plazierung der Kanten nur zwischen Knoten mit einem Grad $k_S < \text{median}(k_N)$ mit $N = (1, 2, 3, \dots, n)$.

Die Variante (b) reduziert den Einfluss der *Hubs* in einem Graphen, indem zunächst ein Knotenpaar mit $B > 0.4 \cdot \max(B_N)$ ausgewählt und anschließend mit einem Knotenpaar mit $B < 0.2 \cdot \max(B_N)$ neu verknüpft wird. Da ein BA-Graph nur eine geringe Zahl an Knoten mit einer *Betweenness*-Zentralität B oberhalb dieses Schwellenwertes bereithält, bedeutet das Entfernen dieser Kanten eine drastische Reduzierung der zentralen *Hub*-Dominanz mit zunehmender Randomisierungstiefe.

Die Variante (c) führt zu einer Hierarchisierung (Trusina *et al.*, 2004) eines Graphen, indem Knotenpaare zufällig ausgewählt und so neu verknüpft werden, dass nun die Knoten mit den hohen Graden verbunden sind.

In der entsprechenden Variante (d) erfolgt die Antihierarchisierung durch Neuverknüpfung von Knoten mit hohem Grad und Knoten mit niedrigem Grad.

Kapitel 5

Ergebnisse Teil I - Verteilungsmuster von Anregungen auf abstrakten Graphen

5.1 Einführung

Einer der zentralen Punkte dieser Arbeit beschäftigt sich mit der Frage, wie Graphenarchitekturen dynamische Prozesse beeinflussen. In diesem Kapitel, welches sich an einer Arbeit von Müller-Linow *et al.* (2006) orientiert, sollen zur Beantwortung dieser Frage globale dynamische Eigenschaften eines Graphen mittels eines einfachen dynamischen Modells (das in Kap. 3.1 vorgestellte DE-Modell) untersucht und in Beziehung zu topologischen Größen gesetzt werden. Will man den Einfluss einer topologischen Eigenschaft auf die Dynamik aus einer globalen Observablen ableiten, so ist es naheliegend, die Menge der Anregungen (normiert auf die Systemgröße, also die Anregungsdichte ρ_A) zu betrachten, die in dem System zu einem bestimmten Zeitpunkt vorliegen. Dies setzt voraus, dass die dynamischen Parameter des Modellsystems konstant gehalten werden, so dass Veränderungen in der Anregungsdichte ausschließlich auf topologische Eigenschaften des Graphen zurückgeführt werden können. Welche strukturelle Ursache letztendlich für die Kapazität eines Graphen, Anregungen unterschiedlich gut zu organisieren und simultan im Graphen vorliegen zu haben, verantwortlich ist, und wie diese die Entstehung bestimmter dynamischer Muster steuert, soll im Folgenden geklärt werden.

Dazu soll der Einfluss von drei verschiedenen topologischen Kenngrößen, (1) dem Vernetzungsgrad, (2) der Gradverteilung und (3) den Gradkorrelationen, auf die Anregungsdichte ρ_A untersucht werden (siehe hierzu auch Kap. 2).

(1) Der Vernetzungsgrad z sollte einen erheblichen Einfluss auf die Propagation von Anregungen haben. Dieser Zusammenhang wird deutlich, wenn man sich einen unvernetzten Graphen (z.B. einen ER-Graphen mit einer Vernetzungswahrscheinlichkeit von $p = 0$) betrachtet. In solchen Graphen wird die Dynamik ausschließlich von der Rate spontaner Anregungen getragen, da Anregungen nicht weitergegeben werden können. Mit steigender Vernetzungsdichte sollte auch der nicht-stochastische Beitrag (an der Anregungsdichte) durch die Propagation von Anregungen an Bedeutung gewinnen. Das Gegenstück zum unvernetzten Graph ist der vollständig vernetzte Graph, in dem jeder Knoten mit allen anderen Knoten verbunden ist. Da in einem solchen Graphen eine Anregung innerhalb eines Zeitschrittes an alle Knoten

weitergegeben wird, müsste dieser Fall für die höchsten Anregungsdichten sorgen. Dieser Zusammenhang soll in dieser Studie mit analytischen bzw. numerischen Methoden untersucht werden.

(2) Die Gradverteilung $p(k)$ ist eine der wichtigsten Eigenschaften, um verschiedene Graphenarchitekturen, die sich hinsichtlich ihrer Größe und ihrer Konnektivität ähneln, zu unterscheiden. Man könnte daher erwarten, dass sich diese topologische Eigenschaft auch im dynamischen Verhalten widerspiegelt. Betrachtet man beispielsweise eine δ -Verteilung mit $\langle k \rangle = 4$, wie sie z.B. in einem regulären Graphen aus einem Ring von Knoten vorliegt (z.B. in einem Watts-Strogatz (WS) Modell mit $p = 0$), dann sollte die Anregungsdichte bei entsprechender Größe deutlich niedriger ausfallen, als bei anderen verbundenen Systemen mit nicht-regulärer Architektur und gleichem Vernetzungsgrad, da die Dauer einer vollständigen Anregung unter anderem auch durch die Größe des Graphendurchmessers beeinflusst wird. Da reale Systeme nicht durch reguläre Strukturen charakterisiert sind, soll der Einfluss der Gradverteilung auf der Basis von drei gängigen Architekturen untersucht werden, die bis zu einem gewissen Grad Eigenschaften realer Systeme repräsentieren.

(3) Die Gradkorrelationen sind ein weiteres Maß, um verschiedene Architekturen zu klassifizieren. Graphen, die sich in dieser Eigenschaft deutlich unterscheiden, sind in der Regel auch deutlich unterschiedlich strukturiert. Stark positiv korrelierte Graphen können beispielsweise ein Zentrum aus Knoten mit einem hohen Grad besitzen (in einer skalenfreien Gradverteilung wären dies die Hubs), stark antikorrelierte Graphen hingegen weisen eher eine Separierung dieser hochgradigen Knoten auf. Ein Einfluss auf die Anregungsdichte ρ_A sollte daher überprüft werden.

In einer Arbeit über das DE-Modell (dort *Forest-Fire* Modell) auf einem *Small-World* Modell konnten Graham & Matthai (2003) zeigen, dass der dynamische Parameter f einen ähnlichen Effekt wie das Einfügen von Kanten (*Shortcuts*) in eine reguläre Struktur hervorrufen kann. Um diese Vermischung topologischer und dynamischer Eigenschaften besser verstehen zu können, soll in weiteren numerischen Analysen (an unterschiedlichen Graphenarchitekturen) und analytischen Untersuchungen (an vollständig vernetzten Graphen) zudem der Einfluss der dynamischen Parameter des Modells auf die Stabilität der Anregungsmuster und damit auf die Anregungsdichte ρ_A bestimmt werden.

Für diese Studie wurden drei typische Graphenarchitekturen, der zufällige (ER) Graph (Erdős & Rényi, 1959), der Small-World (WS) Graph (Watts & Strogatz, 1998; Newman & Watts, 1999) und der skalenfreie (BA) Graph (Barabási & Albert, 1999) untersucht. Gemäß der einzelnen Erzeugungsalgorithmen (Kap. 4.1) wurden Graphen mit $n = 1000$ und variablem Vernetzungsgrad z erzeugt ($10^{-4} \leq z \leq 1$ für ER-Graphen und $10^{-3} \leq z \leq 1$ für WS- bzw. BA-Graphen), wobei der Vernetzungsgrad $z = 2m/(n(n-1))$ durch die Knoten- und Kantenzahl (n und m) festgelegt ist. Die Zustandsdichten der Knotenzustände im dynamischen Modell (erregbar E , angeregt A und refraktär R) sind durch ρ_E , ρ_A und ρ_R gekennzeichnet.

5.2 Einfluss der Vernetzungsdichte

Es existieren zwar keine analytischen Ergebnisse zum DE-Modell für die genannten Graphenarchitekturen variabler Konnektivität, aber ein erster viel versprechender Ansatz war die *Mean-Field*-Analyse der Gleichgewichtsdichten im *Forest-Fire* Modell (Christensen *et al.*, 1993; Barthélemy *et al.*, 2004), auf deren Grundlage man ungefähre Ergebnisse für die Anregungsdichte ρ_A sowie für bestimmte Verteilungseigenschaften der Anregungen als Funktion des Vernetzungsgrades z erhält. Nimmt man an, dass in jedem Zeitschritt k benachbarte Knoten zufällig ausgewählt und synchron aktualisiert werden, dann entstehen Korrelationen weder zwischen den Knoten noch zwischen den Knotengraden, da die Zahl der Nachbarn exakt k ist. Die Zahl erregbarer, angeregter und refraktärer Zustände zum Zeitpunkt $t + 1$ ergibt sich aus den dynamischen Regeln, den dynamischen Parametern f und π und der mittleren Zahl der Nachbarn k :

$$A(t + 1) = fE(t) + kE(t)\theta_A(t), \quad (5.1)$$

$$E(t + 1) = (1 - f)E(t) + \pi R(t) - kE(t)\theta_A(t), \quad (5.2)$$

$$R(t + 1) = A(t) + (1 - \pi)R(t). \quad (5.3)$$

Der Term $fE(t)$ stellt den Einfluss der Spontanaktivität dar: Mit der Wahrscheinlichkeit f wechselt ein Knoten im erregbaren Zustand E in den angeregten Zustand A . Der Anteil $fE(t)$ trägt zur Erhöhung von $A(t + 1)$ bei, während der Anteil $(1 - f)E(t)$ im Zustand $E(t + 1)$ verbleibt. Der Term $kE(t)\theta_A(t)$ beschreibt den nicht-stochastischen Betrag zur Erhöhung von $A(t + 1)$ bzw. zur Verringerung von $E(t + 1)$ aufgrund der Anregungspropagation: Dieser Term hängt von der Verfügbarkeit erregbarer Knoten ($kE(t)$) und von der Wahrscheinlichkeit θ ab, die angibt, ob sich ein angeregter Knoten A in direkter Nachbarschaft zu einem erregbaren Knoten E befindet. Die Regeneration der erregbaren Zustände aus R wird durch den Term $\pi R(t)$ reguliert, die Erniedrigung von R entsprechend durch den Term $(1 - \pi)R(t)$. Zu einer Erhöhung von $R(t + 1)$ tragen alle Knoten im Zustand A bei. Die Berechnung dieses allgemeinen *Mean-Field*-Modells wurde von Carsten Marr in Müller-Linow *et al.* (2006) durchgeführt. Man erhält schliesslich einen Ausdruck für die Anregungsdichte ρ_A . Bei geringen Anregungsdichten kann man Beiträge höherer Ordnung vernachlässigen, so dass die Wahrscheinlichkeit θ_A allein aus der Anregungsdichte $\rho_A(t) = A(t)/n$ resultiert. Für $\rho(t + 1) = \rho(t)$ und $1 = \rho_A + \rho_E + \rho_R$ erhält man die Gleichgewichtslösung für die Zustandsdichte ρ_A durch Multiplikation mit $1/n$. Obwohl beide Lösungen dieser quadratischen Gleichung zu stabilen Fixpunkten führen, besitzt nur ein Fixpunkt einen positiven Wert. Bezieht man z in die Lösung mit ein und ersetzt $n - 1$ durch \tilde{n} , dann erhält man

$$\rho_A = \frac{\tilde{n}z\pi - f - \pi - f\pi + \sqrt{-4\tilde{n}z\pi^2 + (f + \pi + f\pi + \tilde{n}z\pi)^2}}{2\tilde{n}z(1 + \pi)}. \quad (5.4)$$

Dieser *Mean-Field*-Ansatz wird im Allgemeinen benutzt, um die Eigenschaften des DE-Modells auf d -dimensionalen Gittern zu verstehen. Er dient aber auch als gute Annäherung für ER-Graphen, da dort die Gradverteilung um einen mittleren Grad $\langle k \rangle$ sehr schmal ist und Gradkorrelationen fehlen. In großen regulären gitterähnlichen Graphen gibt es keine globale Synchronisation, da hier die Anregungen nur in Form von propagierenden Wellen existieren und es dadurch zu einer konstanten Anregungsstärke kommt.

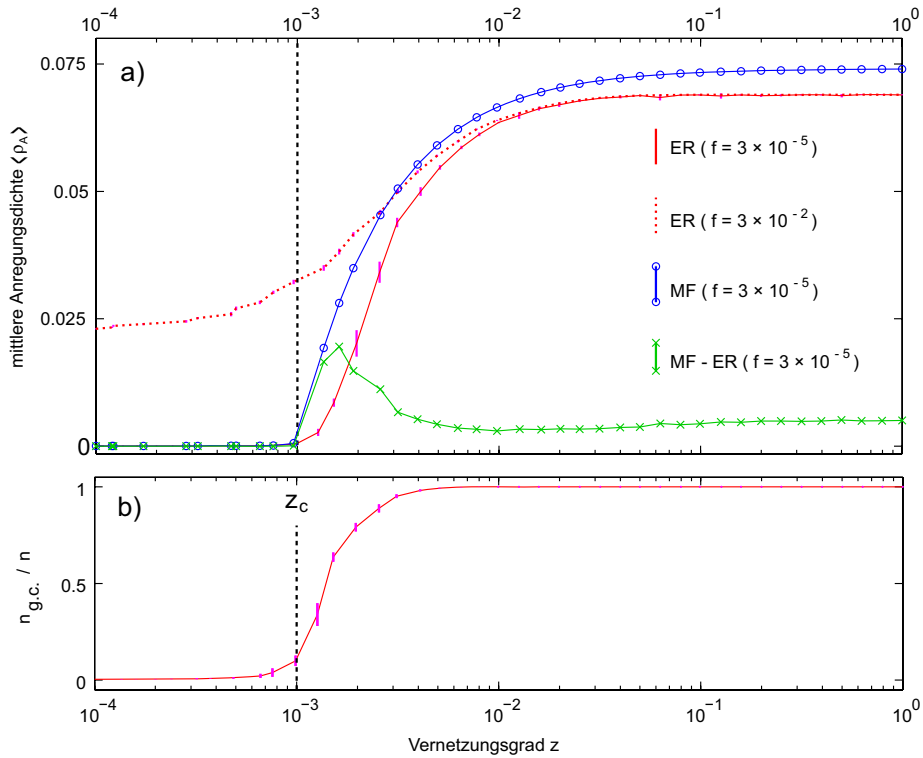


Abbildung 5.1: (a) Beziehung zwischen der mittleren Anregungsdichte $\langle \rho_A \rangle$ und dem Vernetzungsgrad z in einem zufälligen (ER) Graph (simulierte Resultate (ER) und *Mean-Field*-Annäherung (MF) mit $f = 3 \times 10^{-5}$ bzw. $f = 3 \times 10^{-2}$ und $\pi = 0.08$). Die Differenzkurve ($MF - ER$) zeigt die Unterschiede zwischen analytischen und simulierten Resultaten. Die maximale Abweichung im Bereich von $z = 1.5 \times 10^{-3}$ deutet den Vernetzungsgrad an, bei dem der mittlere Grad $\langle k \rangle$ der *Giant Component* (g.c.) den Wert 2 übersteigt. (b) Die ER-Graph Perkolationschwelle wird repräsentiert durch die relative Größe der größten verbundenen Komponente $n_{g.c.}$ und die Perkolationschwelle z_c .

Für $\pi \gg f$ kann man annehmen, dass der Parameter π nur Auswirkungen auf die Zeitskala der Dynamik hat. Dies wurde an ER-Graphen und verschiedenen Parameterkonstellationen mit dem Ergebnis überprüft, dass die Anregungsmuster über einen weiten Bereich von π unverändert bleiben und die Anregungsdichte ρ_A mit π skaliert. Im Folgenden wurden zwei Parametereinstellungen ($\pi = 0.08$, $f_1 = 3 \times 10^{-5}$ bzw. $f_2 = 3 \times 10^{-2}$) zur Bestimmung der Topologieabhängigkeit der Anregungsmuster im ER-Graphen getestet. 10 Simulationen mit einer Dauer von $t_{max} = 25000$ wurden für jede Graphen-Parameterkonstellation durchgeführt.

Abb. 5.1a zeigt die mittlere Anregungsdichte $\langle \rho_A \rangle$ als Funktion des Vernetzungsgrades z für die simulierten und analytischen Ergebnisse. Zur Interpretation dieser Resultate müssen Perkolations-effekte berücksichtigt werden. Perkolations bezeichnet im ER-Modell den Übergang von einem unverbundenen zu einem verbundenen Graph bei zunehmender Vernetzungsdichte. In diesem Übergangsbereich kommt es zur Ausbildung eines großen Knotenclusters (*Giant Component*), welcher von vielen kleinen nicht verbundenen Fragmenten umgeben ist. Ab einem bestimmten Vernetzungsgrad ist dieser Cluster mit dem gesamten Graph identisch. Die Perkolations des ER-Graphen wurde durch Berechnung dieser Clustergröße

bestimmt (siehe Abb. 5.1b). Die theoretische Perkolationschwelle liegt bei $z_c = 1/n$ (Albert & Barabási, 2002). Aus dem Unterschied zwischen numerischer Simulation und den analytischen Resultaten in Abb. 5.1 wird deutlich, dass die Analytik oberhalb von z_c vergleichsweise höhere Anregungsdichten ermittelt. Der Phasenübergang an dieser Schwelle (das ist der Punkt, an dem sich ein großer Knotencluster bildet) ist in dem Gleichgewichtsmodell bereits berücksichtigt ($\rho_A \rightarrow 0$ für $z \rightarrow z_c$ in Gl. (5.4)). Im Gegensatz zur Differenz zwischen beiden Kurven kann der starke Abfall von ρ_A für kleine Werte von z nicht durch Perkolationsphänomene erklärt werden. Bei genauerer Betrachtung der einzelnen Zeitreihen für kleine Werte von z fallen zwei Regionen auf, die sich deutlich in ihren Anregungsmustern unterscheiden (siehe Abb. 5.2a). Das erste Regime zeichnet sich durch starke kurze Anregungen (deutlich voneinander getrennte *Spikes*) aus, die als ein globales dynamisches Ereignis interpretiert werden können (Abb. 5.2b), während das zweite Regime durch langandauernde Fluktuationen (*Bursts*) charakterisiert ist, die in Folge lokaler dynamischer Ereignisse entstehen und deutlich reduzierte Intensitäten aufweisen (Abb. 5.2c).

Betrachtet man die durchschnittliche Anregungsdichte in beiden Regimen, so ist diese im *Spike*-Regime deutlich geringer als im *Burst*-Regime. Im *Spike*-Regime hängt sie von der mittleren Distanz l_{Spike} zwischen zwei *Spikes* (Abb. 5.2b) und der mittleren Zahl angeregter Zustände während eines *Spike*-Ereignisses ab. Die Gl. (5.4) zeigt, dass in schwach vernetzten Graphen (mit $z < z_c$ und $f \ll \pi$) die durchschnittliche Anregungsdichte $\langle \rho_A \rangle$ im Wesentlichen durch die spontane Anregungsrate f bestimmt wird. Dies erklärt auch den Anstieg von $\langle \rho_A \rangle$ für $f_2 = 3 \times 10^{-2}$. Simuliert man mit kleinem f auf einem Graph ohne Kanten, so ist $\langle \rho_A \rangle \approx f$. Im Grenzwertbereich unendlicher Systemgrößen mit $z = 1$ lässt sich Gl. (5.4) zu $\rho_A = \frac{\pi}{1+\pi}$ vereinfachen. Beide Grenzwerte stimmen mit den Asymptoten der jeweiligen Kurve numerisch überein (siehe Abb. 5.1a).

Der analytische Ansatz zum DE-Modell wurde von C. Marr für vollständig vernetzte Graphen ($z = 1$) in Müller-Linow *et al.* (2006) erweitert. Auf dieser Grundlage ist es möglich für jede beliebige Parameterkonstellation von f und π die durchschnittliche Anregungsdichte $\langle \rho_A \rangle$ mit den anteiligen Beiträgen der entsprechenden dynamischen Regime (z.B. dem temporalen Anteil des *Burst*-Regimes a_{burst}) zu berechnen (siehe Gl. (5.3) und (5.4)). Die Analytik der vollständig vernetzten Graphen ($n = 500$, $\pi = 0.01$ und $10^{-6} \leq f \leq 1$) wurde mit den entsprechenden Simulationsergebnissen (25 Simulationen für jede Parameterkonstellation mit $t_{max} = 25000$) verglichen. Beide Ergebnisse (hier nicht extra aufgeführt) zeigen konsistentes Verhalten. Von geringen Variationen abgesehen ist die Analytik prinzipiell in der Lage, eine gute Vorhersage der simulierten Ergebnisse zu leisten, was sie für weitere Studien des DE-Modells auf Graphenarchitekturen mit variabler Konnektivität und variablen Gradkorrelationen interessant macht.

5.3 Einfluss der Gradverteilung

Um reine Perkolationseffekte als Ursache für den Wechsel in den Anregungsmustern endgültig auszuschließen und die tragende Rolle der Vernetzungsdichte weiter zu unterstreichen, wurden die Resultate entsprechender Simulationen an WS- und BA-Graphen mit denen der ER-Graphen verglichen. Abb. 5.3a zeigt die Ergebnisse der drei Graphentypen für $\pi = 0.08$ und $f = 3 \times 10^{-5}$. Überraschenderweise weisen alle Kurven einen ähnlichen Übergang für ρ_F oberhalb der Perkolationschwelle z_c aus. Abb. 5.3b zeigt für verschiedene Konnektivitäten (exemplarisch am BA-Graphen) Ausschnitte aus drei Zeitreihen. Die

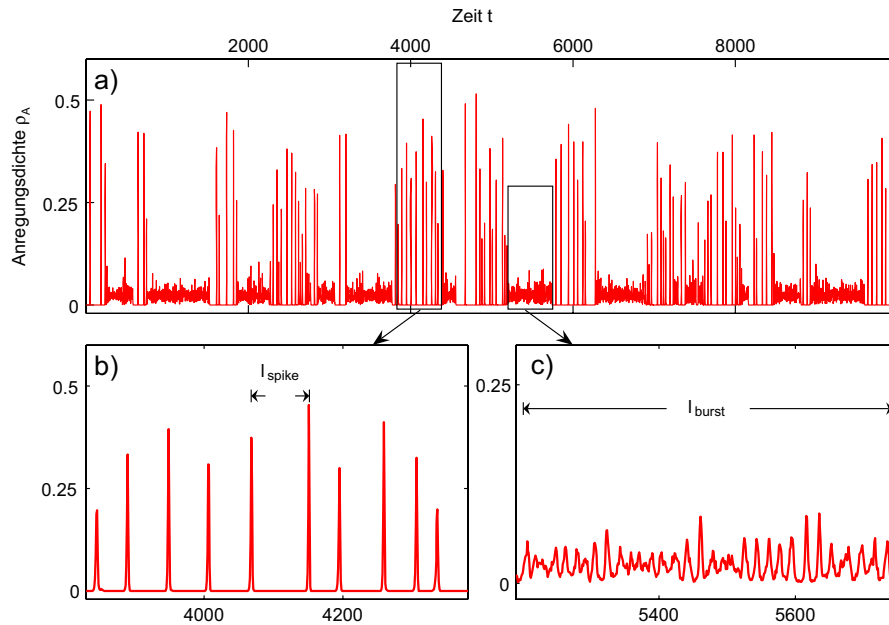


Abbildung 5.2: (a) Zeitreihe der Anregungsdichte ρ_A in einem zufälligen (ER) Graph ($z = 0.006$; $n = 1000$) mit $f = 3 \times 10^{-5}$ und $\pi = 0.08$. Die Ausschnitte zeigen zwei typische dynamische Verhaltensformen, und zwar (b) das *Spike*-Regime (hier gibt l_{spike} die Dauer zwischen zwei benachbarten *Spike*-Ereignissen an) und (c) das *Burst*-Regime (l_{burst} steht für die Dauer eines *Bursts*).

Zeitreihe für $z_1 = 0.002$ ist durch ein großes *Spike*-Regime, die Zeitreihe für $z_3 = 0.063$ durch ein großes *Burst*-Regime und der Übergangsbereich bei $z_2 = 0.0025$ durch die Koexistenz beider Regime charakterisiert. Folgerichtig muss eine Erhöhung der Konnektivität z einen Wechsel im Verteilungsmuster der Anregungen und damit auch in der mittleren Anregungsdichte $\langle \rho_A \rangle$ induzieren. Die Ähnlichkeit aller Kurven zeigt zudem, dass die Gradverteilung nur einen sehr geringen Einfluss auf die Anregungsdichte ρ_A ausübt, obwohl sie das beste Maß zur Charakterisierung der drei Graphenklassen darstellt und, sofern von Bedeutung, gerade bei niedrigen Konnektivitäten (also im Übergangsbereich) alle Graphen deutlich voneinander trennen sollte. Offensichtlich scheint sie aber bei der Regulierung der Anregungsmenge in einem Graph nur eine untergeordnete Rolle zu spielen.

Um zu überprüfen, ob die Erhöhung von ρ_A wirklich auf einen qualitativen Wechsel in den Verteilungsmustern der Anregungen zurückzuführen ist, wurde die Anregungsdichte innerhalb der *Burst*-Regime (als Funktion der Konnektivität) für die BA- und die ER-Graphen quantifiziert ($\langle \rho_{burst} \rangle = \rho_{burst} \cdot t_{burst} / t_{max}$) und mit der mittleren Anregungsdichte $\langle \rho_A \rangle$ verglichen. Ein *Burst*-Regime unterscheidet sich von einem *Spike*-Ereignis durch die Dauer der Anregung. Während die Länge eines *Bursts* von der Verteilung der erregbaren Knotenzustände E in einem Graphen abhängt, wird die Dauer eines *Spike*s durch die Längenverteilung der kürzesten Pfade in einem Graphen reguliert. Ein *Burst*-Regime entspricht in dieser Definition einem Ausschnitt aus einer Zeitreihe mit einer Anregungsdichte, die für jeden Zeitpunkt größer als Null ist, und mit einer Länge $l_{burst} > d$, wobei der Durchmesser d eines Graphen die Maximaldauer

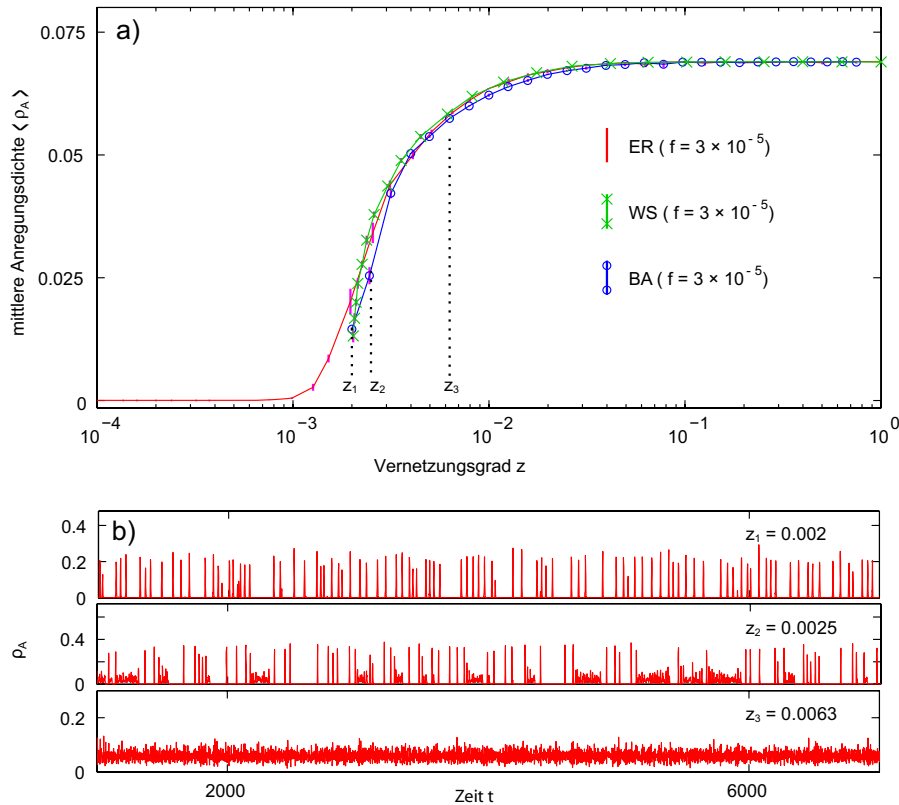


Abbildung 5.3: (a) Beziehung zwischen der mittleren Anregungsdichte $\langle \rho_A \rangle$ und dem Vernetzungsgrad z für drei verschiedene Graphenarchitekturen: der zufällige Graph (ER), der skalenfreie Graph (BA) und der *Small-World* Graph (WS) wurden mit $f = 3 \times 10^{-5}$ und $\pi = 0.08$ simuliert. (b) Die Ausschnitte zeigen die Zeitreihen der mittleren Anregungsdichte im BA-Graphen für verschiedene Konnektivitäten. Im Bereich der kleinsten Vernetzungsdichten, z.B. $z_1 = 0.002$, werden die Anregungsmuster ausschließlich durch *Spike*-Dynamik induziert, für Werte oberhalb $z_3 = 0.0063$ dominiert die *Burst*-Dynamik. Die Übergangsregion zeigt beide Anregungsmuster zu unterschiedlichen Anteilen (siehe z.B. $z_2 = 0.0025$).

eines *Spikes* festlegt. Die Länge t_{burst} entspricht der Summe aller *Burst*-Regime. Die Ähnlichkeit der Ergebnisse (hier nicht gezeigt) legt den Schluss nahe, dass die Anregungsstärke durch das Verhältnis von globalen und lokalen Koinzidenzen bzw. durch den Wechsel von *Spikes* zu *Bursts*, getragen wird. Bei niedrigen Konnektivitäten zeigt der ER-Graph vergleichsweise höhere Werte für $\langle \rho_{burst} \rangle$. Dieses Verhalten kann ähnlich wie zuvor interpretiert werden: Die Ergebnisse für den ER-Graphen zeigten bereits, dass auch unterhalb der Schwelle von $z = 0.002$, an der erst ein verbundener BA-Graph entsteht, die *Giant Component* des ER-Graphen bereits *Burst*-Verhalten erzeugen kann und damit mehr zur Anregungsstärke beiträgt, als man von der globalen Vernetzungsdichte erwarten würde.