Modellbasierter Entwurf strukturanaloger Architekturen auf Basis der Partitionierung von Graphen

Stephan Aier

Institut für Wirtschaftsinformatik
Universität St.Gallen
stephan.aier@unisg.ch

Marten Schönherr

Institut für Wirtschaftsinformatik und quantitative Methoden
Technische Universität Berlin
mschoenherr@sysedv.tu-berlin.de

Abstract

In großen Unternehmen treffen hohe Komplexität auf die Forderung nach hoher Flexibilität, um Produkte und Prozesse den sich schneller ändernden Anforderungen anpassen zu können. Informationssysteme stellen dabei einen unterstützenden, gleichzeitig aber auch einen hemmenden Faktor dar, da sie oft weniger schnell verändert werden können, als die durch sie unterstützten Prozesse. Da die Abhängigkeiten zwischen Prozessen und IT, die gemeinsam die Unternehmensarchitektur bilden, in der Summe zunehmen, liegt die Aufgabe des Architekturmanagements darin, nachhaltige Strukturen zu schaffen. Ein Ansatz zur Erreichung dieses Ziels kann die Modularisierung der Architektur unter Berücksichtigung der Abhängigkeiten zwischen Prozessen und IT sein. Dieser Beitrag zeigt, wie die Partitionierung von Graphen durch Clusteringalgorithmen hilft, solche Module in bestehenden Architekturen zu identifizieren.

1 Problemstellung

Die IT-Systemlandschaften der meisten großen Unternehmen haben heute eine Komplexität erreicht, die nur noch schwer zu beherrschen ist. Gleichzeitig soll die IT sich immer schneller ändernde Geschäftsprozesse unterstützen. Um solche komplexen Unternehmensarchitekturen als Gesamtheit von Organisationsstrukturen, Geschäftsprozessen und IT-Systemen effizient

managen zu können, werden unter anderem verschiedene Technologien prozessorientierter IT-Systemintegration wie beispielsweise Enterprise Application Integration (EAI) oder Serviceorientierte Architekturen (SOA) diskutiert. Neben diesen Technologien sind vor allem Methoden für die Gestaltung komplexer Unternehmensarchitekturen von Bedeutung.

Ein mögliches Gestaltungsziel für Unternehmensarchitekturen, welches durch die Diskussion über SOA wieder stärker in den Mittelpunkt gerückt ist, ist die Modularisierung. Es bleibt jedoch oft unklar, nach welchen möglichst generischen Kriterien Module zu bilden sind. Dies gilt für technische wie für fachliche Aspekte. Die hier entwickelte Methode wendet Clusteringalgorithmen an, um in einer bestehenden, komplexen Unternehmensarchitektur latent vorhandene Module zu identifizieren. Die Algorithmen arbeiten auf Basis eines Enterprise Architecture Modells, welches insbesondere Aspekte der prozessorientierten Systemintegration abbildet. Durch die auf formalen Modellen basierende ganzheitliche Sichtweise auf Geschäftsprozesse und IT-Systeme sollen Architekturmodule identifiziert werden, die als Basis für eine nachhaltige Unternehmensarchitektur dienen. Dazu wird im folgenden Abschnitt kurz unser Verständnis von Unternehmensarchitekturen dargestellt, da es die Grundlage für das folgende Architekturmodell und die Architekturgestaltung bildet. Daran anschließend werden Gestaltungsansätze für nachhaltige Architekturen beschrieben, um dann unseren Architektur- und Clusteringsansatz sowie dessen technische Implementierung darzustellen.

2 Verständnis der Unternehmensarchitektur

Der Begriff der *Unternehmensarchitektur* (Enterprise Architechture, EA) ist zentraler Bestandteil dieses Beitrags. Sowohl unsere Gestaltungsvorschläge für Unternehmensarchitekturen als auch die technische Implementierung für die Modellierung und Analyse dieser Architekturen bauen darauf auf. Darum wird im Folgenden ein Überblick über das zugrunde liegende Architekturverständnis gegeben.

Verkürzt kann eine Architektur als eine abstrakte, ganzheitliche Betrachtung von Strukturen und Mustern mit Planungscharakter aufgefasst werden [Bas+03, S. 19 ff.]. Architekturen sind in der Regel das Ergebnis eines Planungsprozesses und stellen nach ihrer Definition selbst einen Masterplan für die ganzheitliche Realisierung zukünftiger Maßnahmen dar.

Diese allgemeinen Charakteristika lassen sich auf die Gestaltung von Unternehmen anwenden und werden so zur Unternehmensarchitektur. Unter einer Unternehmensarchitektur wird das

Zusammenwirken organisatorischer, technischer und psychosozialer Aspekte bei der Planung und Entwicklung betrieblicher soziotechnischer Informationssysteme verstanden [Gron03, S. 45]. Im Folgenden sollen vor allem die organisatorische und die technische Dimension der Unternehmensarchitektur erläutert werden. Dazu werden die Begriffe *Organisationsarchitektur* und *IT-Architektur* verwendet (Abbildung 1).

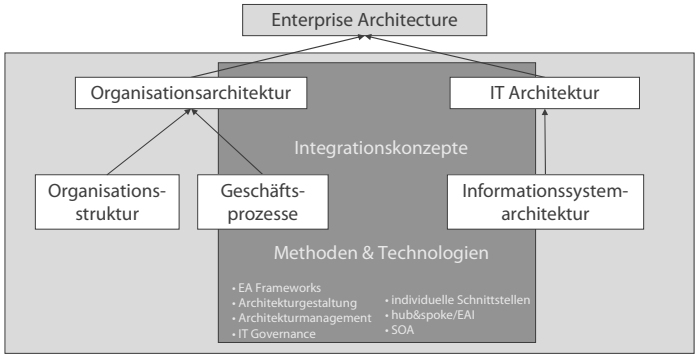


Abbildung 1: Bestandteile der Unternehmensarchitektur

Die *Organisationsarchitektur* enthält alle nichttechnischen Bestandteile der Unternehmensarchitektur und kann mit dem instrumentalen Organisationsbegriff verglichen werden, der die Gesamtheit aller generellen expliziten Regelungen zur Gestaltung der Aufbau- und Ablauforganisation umfasst [Blei91, S. 35]. Entsprechend wird die Organisationsarchitektur hier in die Organisationsstruktur und die Geschäftsprozesse unterschieden.

Komplementär zur Organisationsarchitektur steht die *IT-Architektur*. Sie umfasst alle technischen Bestandteile der Unternehmensarchitektur. Insbesondere beinhaltet sie die technischen Informationssysteme, welche ihrerseits eine eigene Architektur, die Informationssystemarchitektur, aufweisen. In der Literatur sind die Begriffe Organisationsarchitektur und IT-Architektur häufig verwendet, jedoch meist anders definiert. Abhängig von der fachlichen Herkunft des Autors umfasst die Organisationsarchitektur auch technische Komponenten [Nad+94] bzw. die IT-Architektur auch organisatorische Fragestellungen [Krcm90]. Im Folgenden sollen jedoch beide Bereiche separat und somit bezogen auf ihre Bedeutung gleichwertig betrachtet werden.

Die Aufgabe von *Integrationskonzepten* besteht zum einen darin, die technischen Komponenten der Architektur technisch zu verbinden, zum anderen dienen Integrationskonzepte zur konzeptuellen Zusammenführung von technischen Elementen mit den Elementen der Organisation – was oft unter dem Begriff Business-IT-Alignment diskutiert wird [WiLa06]. Zur Bewältigung dieser Integrationsaufgabe bedarf es *Technologien* – wie EAI oder SOA – und *Methoden*. Letztere stehen im Mittelpunkt dieses Beitrags.

3 Gestaltungsansätze nachhaltiger Unternehmensarchitekturen

Im folgenden Abschnitt sollen im ersten Teil das Konzept der intern orientierten Nachhaltigkeit dargelegt werden, um dann die Strategie der strukturellen Analogie und die Modularisierung als Gestaltungsansätze nachhaltiger Unternehmensarchitekturen abzuleiten. Dabei handelt es sich um die Fortführung der in [AiDo04; AiDo05] dargestellten Ansätze. Diese bilden in der Folge die Grundlage für die Modellierung, Analyse und Optimierung von Unternehmensarchitekturen.

3.1 Konzept der intern orientierten Nachhaltigkeit

In der Literatur existieren unzählige Definitionen und Konzepte von Nachhaltigkeit. Im Kern steht jedoch das Verständnis, das ein System langfristig nicht mehr Ressourcen verbrauchen darf als neu verfügbar werden. Aus einer Analyse dieser Definitionen lassen sich vier Strategien extrahieren, die im Ergebnis ihrer Anwendung zu mehr Nachhaltigkeit führen sollen. Dies sind die Strategie der Effizienz, die Strategie der Suffizienz, die Strategie der Konsistenz und die Strategie der Partizipation.

Der *Suffizienzstrategie* erhebt die Forderung, durch Verzicht ein ausreichendes Maß an Genügsamkeit zu realisieren und basiert auf der Einsicht, dass es keine Unbeschränktheit gibt, da jedes System Grenzen in Raum und Zeit aufweist [Hube95, S. 40]. Die *Effizienzstrategie* zielt auf die Steigerung der Produktivität, um dadurch Leistungen wirtschaftlich, d.h. mit dem kleinsten möglichen Ressourcenverbrauch zu erstellen. Zentrale Konzepte sind dabei Wiederverwendung und Langlebigkeit. Die *Konsistenzstrategie* hat entweder die vollständige Abschirmung von Systemen von deren Umwelt oder die Sicherstellung deren Verträglichkeit, Stimmigkeit mit dem sie umgebenden (Super-)System zum Ziel [Hube95, S. 41 f.]. Wenn Systeme keine Verbindungen zur umgebenden Umwelt haben, müssen diese Systeme auch nicht mit der Umwelt verträglich sein, da es keine Wirkungsbeziehungen zwischen ihnen gibt. Bestehen jedoch Beziehungen, so wird Konsistenz der Systeme mit ihrer Umwelt gefordert. Die *Partizipationsstrategie* schließlich fordert die Teilhabe, d.h. die Einbindung der von einer Systemgestaltung Betroffenen. Dies ist nötig, um das System zum einen bestmöglich für die „Betroffenen“ zu gestalten und zum anderen, um deren Akzeptanz des Systems sicherzustellen [Gron03, S. 222].

Wird Nachhaltigkeit als ein für die Unternehmung relevantes Thema betrachtet, so geht es meist um die Reduktion negativer externer Effekte auf die Umwelt oder die Gesellschaft (physische und soziale Umwelt) welche durch das Handeln des Unternehmens entstehen [LeSt03, S. 259 f.]. Dies soll im Folgenden als *extern orientierte Nachhaltigkeit* bezeichnet werden. Das hier

zugrunde gelegte Verständnis von Nachhaltigkeit soll vollkommen losgelöst von Themen der Ökologie sein. Vielmehr soll das Ziel der Betrachtung eine unternehmensinterne Sicht der Nachhaltigkeit mit dem primären Ziel der langfristigen, effizienten Unternehmensführung sein [Gron03; Klam00; Scha02]. HAHN/HUNGENBERG definieren das Oberziel einer jeden Unternehmung als Erhaltung und erfolgreiche Weiterentwicklung, als Erfüllung der Individualziele aller an der Unternehmung interessierten Gruppen [HaHu01, S. 13]. Somit kann der hier entwickelte interne Nachhaltigkeitsbegriff als ein Unterziel dieses Oberziels betrachtet werden. Dies soll im Folgenden als *intern orientierte Nachhaltigkeit* bezeichnet werden [AiDo04; AiDo05].

3.2 Strategie der strukturellen Analogie von Organisations- und IT-Architektur

Der Zusammenhang zwischen der Organisation einerseits und der IT andererseits hat in der wissenschaftlichen Literatur große Beachtung erfahren. Trotz der langjährigen Beschäftigung der Wissenschaft mit diesem Thema und den zahlreichen Veröffentlichungen lassen sich jedoch keine absoluten, verallgemeinernden Aussagen zu den Beziehungen zwischen IT und Organisation finden [MaRo88; LeHu98, S. 271 f.]. Das Bestehen monokausaler Zusammenhänge zwischen Organisation und IT wird heute weitgehend verneint. Vielmehr wird angenommen, dass Informationstechnologie eine den Gestaltungsspielraum des Organisators erweiternde Option darstellt – „IT als Enabler“ [Fres00, S. 142 f.]. Gleichzeitig wird aber auch deutlich, dass IT ein „Disabler“ für die Veränderung von Organisationsstrukturen und Geschäftsprozessen sein kann, wenn die IT nicht in ausreichendem Maße an die sich stetig verändernden Anforderungen anpassbar wird [Hage03, S. 66].

Um eine Nachhaltigkeit der Unternehmensarchitektur zu erreichen, ist die Berücksichtigung dieser Abhängigkeiten in einer integrierten Gestaltung unabdingbar. Die Strategie der strukturellen Analogie soll die geforderte Stimmigkeit zwischen IT- und Organisationsarchitektur durch ihre strukturähnliche Gestaltung unterstützen [Gron03, S. 216; Wall96, S. 63; AiDo04]. Der Nachhaltigkeitsaspekt der strukturellen Analogie folgt mittelbar aus der Konsistenzstrategie, nach welcher IT- und Organisationsarchitektur in übereinstimmend zu gestalten sind.

IT- und Organisationsarchitekturen sind strukturanalog zueinander, wenn sie in ihren Merkmalsausprägungen weitgehend übereinstimmen. Besonders geeignet für den Strukturvergleich zwischen Organisations- und IT-Struktur erscheinen die Organisationsdimensionen Spezialisierung (Arbeitsteilung), Entscheidungsdelegation (Kompetenzverteilung) und Formalisierung. Eine Möglichkeit zur Realisierung der Strategie der strukturellen Analogie ist die Modularisierung der Unternehmensarchitektur.

3.3 Strategie der Modularisierung

Flexibilität ist eine Voraussetzung für nachhaltige Architekturen, da sich nur flexible Architekturen dauerhaft an die Anforderungen ihrer Umwelt anpassen können. Die Forderung der Komplexitätsreduktion entstammt der entscheidungstheoretischen Sicht der Nachhaltigkeit. Danach stellen sich negativ irreversible Entscheidungen vor allem als Problem komplexer Systeme dar, in welchen nicht alle Wirkungen von Entscheidungen zur irreversiblen Ressourcenverwendung erkannt werden können [Krca03, S. 23 f.]. Nachfolgend wird gezeigt, warum modulare Architekturen prinzipiell nachhaltig sein können. In Anlehnung an die im vorhergehenden Abschnitt beschriebene strukturelle Analogie zwischen Organisation und IT ist der Gedanke der Modularisierung auf beide Domänen anzuwenden (Abbildung 2).

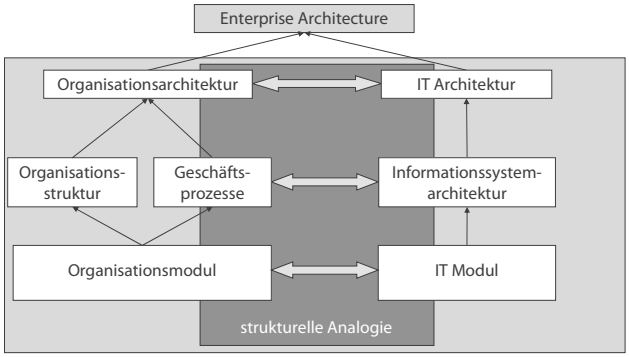


Abbildung 2: Strukturelle Analogie in einer modularisierten Architektur

Modularisierung bedeutet allgemein die Strukturierung eines Systems in relativ kleine, teilautonome und überschaubare Subsysteme (Module, Fraktale, Cluster). Die Komplexitätsreduktion ergibt sich dabei aus der Subsystembildung innerhalb des Systems Unternehmung. Die Subsystembildung wirkt komplexitätsreduzierend, da sie zum einen die subsysteminterne Komplexität im Sinne der Kapselung vor der Subsystemumwelt verbirgt, zum anderen durch die Reduktion auf wenige bekannte Schnittstellen eine Entkopplung der Subsysteme im Sinne einer Reduktion von Abhängigkeiten bewirkt. Mit anderen Worten: Komplexität ist systemtheoretisch ein Selektionszwang und Systemdifferenzierung durch Subsystembildung eine Möglichkeit der Selektion zur Verringerung der Komplexität [Stue96, S. 73]. Im Zusammenhang damit steht die Fähigkeit der einzelnen Subsysteme, selbständig zu agieren und sich selbst organisieren zu können. Die Flexibilisierung der Strukturen und damit der Abläufe ergibt sich durch die nun leichtere Rekonfigurationsmöglichkeit der entkoppelten Module über ihre definierten Schnittstellen.

Die wesentliche Frage ist nun, wie in der Praxis Module gebildet werden sollen, die zu einer strukturanalogen Architektur führen. Um die o.g. Ziele einer Modularisierung zu erreichen wird

vorgeschlagen, die Abhängigkeiten zwischen verschiedenen Modulen zu minimieren und die Abhängigkeiten zwischen Elementen innerhalb eines Moduls zu maximieren indem solche Elemente in ein Modul verschoben werden [AiSc03, S. 27 ff.]. Diese allgemein gehaltene Anweisung lässt sich praktisch nicht ohne weitere Unterstützung durch Methoden und Werkzeuge anwenden. Darum wird im folgenden Abschnitt ein Ansatz entwickelt, der mit Clusteringalgorithmen, unter Berücksichtigung bestehender Geschäftsprozesse, IT-Systeme und deren Interaktionen, Module identifiziert.

4 Modellierung als Grundlage eines Architektur-Clusterings

Modelle bilden die Basis des Clusterings der Unternehmensarchitektur. Darum wird zuerst der verfolge Modellierungsansatz vorgestellt. Darauf folgen die Einführung entsprechender Clusteringalgorithmen sowie die Darstellung der prototypischen Implementierung.

4.1 Ableitung der Modellierungsnotation

Um Clusteringalgorithmen auf Unternehmensarchitekturen anwenden zu können, müssen die hier relevanten Aspekte der Architekturen in einem Modell abgebildet werden. In unserem Ansatz liegt der Schwerpunkt darauf, prozessorientierte Systemintegration modellieren zu können. Dafür sind mindestens die folgenden Elemente notwendig:

- Geschäftsprozesse, d.h. eine Abfolgen von Aktivitäten (Aufgaben),
- IT-Systeme,
- Relationen, die die Nutzung bestimmter IT-Systeme entlang eines Geschäftsprozesses widerspiegeln,
- und die Beziehungen und Schnittstellen zwischen IT-Systemen entlang eines Geschäftsprozesses.

Wir nennen die Sicht, die durch die Abbildung der IT-System-Beziehungen entlang von Geschäftsprozessen entsteht, *dynamische Sicht*. Sie stellt eine Erweiterung statischer Sichten dar, die oft nur IT-Systeme und deren Beziehungen unabhängig von Geschäftsprozessen abbilden. In statischen Sichten ist nicht zu erkennen, welche IT-System-Schnittstellen entlang der Aus-

führungen eines bestimmten Prozesses genutzt werden. Die Abbildbarkeit dynamischer Sichten ist somit eine Voraussetzung für die Modellierung prozessorientierter Systemintegration.

Bei dem Entwurf unseres Modellierungsansatzes war es nicht das Ziel, eine vollständig neue Notation zu entwickeln. Vielmehr bestand das Ziel darin, Unternehmensarchitekturen als Basis für die Anwendung von Clusteringalgorithmen modellieren zu können. Der Schwerpunkt lag dabei auf der Darstellung der Verknüpfung zwischen Geschäftsprozessen, IT-Systemen und der Integration dieser IT-Systeme entlang der Geschäftsprozesse.

Als Basisnotation für die Modellierung kamen nach der Analyse verschiedener Notationen die Unified Modeling Language (UML) sowie ereignisgesteuerte Prozessketten (EPK) in die engere Auswahl. Beide Notationen unterscheiden sich im Bereich der reinen Prozessabbildung faktisch nur unwesentlich. Die Vorteile der EKP liegen in der Möglichkeit IT-Systeme verschiedenen Prozessschritten zuordnen zu können, sowie in der weiteren Verbreitung bei Fachanwendern. Darüber hinaus lässt sich eine einfache EPK auch automatisiert in ein UML Activity Diagram transformieren. Die entgegengesetzte Transformation ist zwar auch möglich, jedoch komplexer. Darum wurde als Ausgangspunkt für die Visualisierung die EPK gewählt. Darüber hinaus wurde die Integration verschiedener IT-Systeme in der dynamischen Prozesssicht, sowie in einer statischen IT-Sicht visualisiert. Um in beliebigen Detaillierungsebenen konsistent modellieren zu können, wurden Eingangs- und Ausgangselemente hinzugefügt (Abbildung 3). Ein solches Modell kann auch als Graph bzw. Netzwerk betrachtet werden. Im folgenden Abschnitt wird eine kurze Einführung in die Graphentheorie sowie in die Algorithmen zur Partitionierung von Graphen gegeben, um danach unsere Implementierung dieser Algorithmen für das Clustering von Unternehmensarchitekturen vorzustellen.

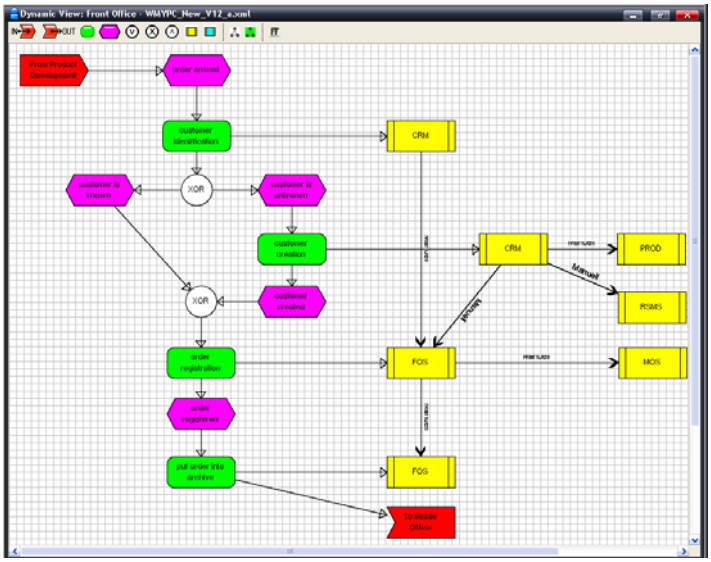


Abbildung 3: Modellierungsnotation auf EPK-Basis

4.2 Graphentheorie und Clusteringansätze

Ein *Graph*/*Netzwerk* besteht aus einer Menge von *Knoten* (vertices) *V* und einer Menge von *Kanten* (edges) *E*. Im hier vorliegenden Kontext kann das Architekturmodell als Graph aufgefasst werden. Alle strukturell wesentlichen Elemente (Module, Funktionen, IT-Systeme) lassen sich als Knoten und die Verbindungen zwischen diesen Elementen als Kanten abstrahieren. Ereignisse und Operatoren werden nicht in den Graphen übernommen, da sie für eine strukturelle Analyse nur von untergeordneter Bedeutung sind.

Graphen deren Kanten durch Zahlenwerte attributiert werden, werden *gewichtete Graphen* genannt. Kanten, die einen definierten Anfang und ein definiertes Ende haben, heißen *gerichtete Kanten*. Kanten bei denen das nicht der Fall ist heißen *ungerichtete Kanten*. Die *Entfernung* zwischen zwei Knoten ist definiert als der kürzeste Pfad zwischen diesen. Bei gewichteten Graphen wird oft auch von der *gewichten Entfernung* gesprochen [OMa+05, S. 4 f.]. Ein solches Netzwerk mit *n* Knoten lässt sich nun als Nachbarschaftsmatrix *A* der Form $n \times n$ beschreiben.

$$A_{ij} = \begin{cases} 1 & \text{wenn die Knoten } i \text{ und } j \text{ verbunden sind} \\ 0 & \text{sonst} \end{cases} \tag{1}$$

In einem gewichteten Netz stellt A_{ij} das Gewicht der Kante, welche die Knoten *i* und *j* verbindet, dar. Weiterhin lässt sich der *Grad* (degree) *k* eines Knotens *i* definieren als die Anzahl der Kanten, die mit diesem Knoten verbunden sind:

$$k_i = \sum_j A_{ij} \tag{2}$$

Die vorliegenden Modelle können in einen solchen Graphen transformiert werden. Dazu ist es zunächst unerheblich, welchem Typ die Elemente entsprechen. Sie werden gleichrangig als Knoten abgebildet. Die Kanten werden als ungerichtete Kanten in den Graphen übernommen, da die Richtungsinformation hier nicht wesentlich für die strukturellen Zusammenhänge ist.

Ein Gestaltungsziel für Architekturen bestand in der Nachhaltigkeit. Zur Umsetzung dieses Ziels wurde eine Modularisierung der Architektur vorgeschlagen. Als die wichtigsten Kriterien für die Modulbildung werden die Minimierung der Abhängigkeiten zwischen verschiedenen Modulen und die Maximierung der Abhängigkeiten zwischen Elementen innerhalb eines Moduls, im Sinne einer Verschiebung solcher Elemente in ein Modul, vorgeschlagen.

Bereits bei kleineren Ausschnitten einer Architektur, wird die Anwendung einer solchen Vorschrift ohne Werkzeugunterstützung schnell schwer überschaubar. Darum wurde versucht, Algorithmen zu finden und anzuwenden, die aus der Struktur eines Graphen die Zusammengehö-

rigkeit verschiedener Elemente bewerten. Dabei wird von der inhaltlichen Dimension des Modells abstrahiert und ausschließlich die Struktur ausgewertet. Entsprechend des Optimierungsansatzes sollten alle Elemente mit einer hohen Zusammengehörigkeit ein Modul bilden.

Solche Fragestellungen werden unter den Schlagworten *Partitionierung* bzw. *Clustering* von Graphen diskutiert. Ein Cluster ist dabei eine Menge von Elementen, die in der einen oder anderen Form ähnlich zueinander sind [OMa+05, S. 18.]. Im hier vorliegenden Fall ergibt sich Ähnlichkeit aus der Tatsache, dass verschiedene Elemente eine gemeinsame Untermenge von Nachbarn haben. Insbesondere in soziologischen Forschungsrichtungen finden Clusteringansätze und netzwerkbezogene Analysen Anwendung wenn es darum geht, Gruppen und Organisationen also soziale Systeme zu analysieren [WaFa99; Scot05].

Girvan/Newman entwickelten ein Clusteringalgorithmus, um Communities in sozialen Netzwerken zu bestimmen. Ein Netzwerk besteht dabei aus einer Menge von Personen (Knoten) die Beziehungen zueinander haben (Kanten). Die Beziehung kann beispielsweise durch das gegenseitige Kennen bestehen. Die Frage ist nun, welche Personen kennen sich in einem Netzwerk direkt und lassen sich diese Personen und die Communities (Cluster) zu denen sie gehören identifizieren [GiNe02, S. 7821]? Abbildung 4 stellt ein solches Netzwerk dar in welchem die Cluster (Communities) hervorgehoben worden sind. Girvan/Newman analysierten bestehende Clusteringalgorithmen nach ihrer Leistungsfähigkeit Cluster in Netzen zu erkennen, deren Strukturen bekannt sind. Dabei handelt es sich meist um hierarchisch arbeitende Algorithmen in welchen Cluster durch Baumstrukturen abgebildet werden (Abbildung 5).

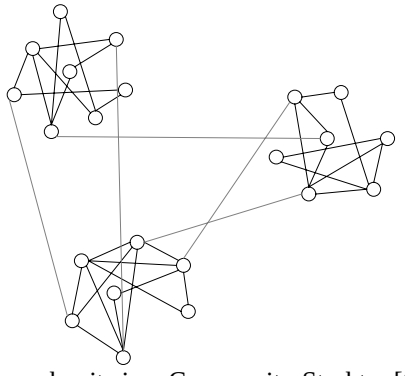


Abbildung 4: Netzwerk mit einer Community-Struktur [GiNe02, S. 7822]

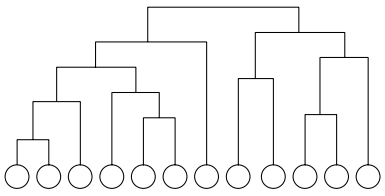


Abbildung 5: Hierarchisches Clustering in Baumstrukturen [GiNe02, S. 7822]

In bestimmten Konstellationen, insbesondere bei Knoten die am Rand eines Netzwerks liegen und nur durch eine einzige Kante mit dem Netz verbunden sind, liefern diese Algorithmen schlechte Ergebnisse. Elemente werden aus einem Cluster ausgeschlossen obwohl sie richtiger Weise dazugehören sollten. Die Grundidee des *„Betweenness"-Algorithmus* von Girvan/Newman ist es nicht, die zentralen Kanten eines Netzes zu finden, sondern die Kanten die am wenigsten zentral sind und somit am meisten *zwischen* (*between*) den Communities liegen.[1]
In der Vergangenheit wurde die Knoten-Betweenness als ein Maß des Einflusses eines Knotens auf ein Netzwerk untersucht. Zuerst von Freeman vorgeschlagen, wird die *Betweenness-Centrality* eines Knotens *i* definiert als die Anzahl der kürzesten Pfade zwischen Paaren anderer Knoten die durch *i* verlaufen [Free77]. Girvan/Newman verallgemeinern nun Freemans Betweenness-Centrality und definieren die *Kanten-Betweenness* einer Kante als die Anzahl der kürzesten Pfade zwischen Paaren von Knoten die über diese Kante laufen. Wenn in einem Netzwerk nun Communities bestehen, dann sind diese nur über wenige Kanten miteinander verbunden. Das heißt, dass die kürzesten Pfade zwischen diesen Communities über diese wenigen Kanten laufen. Darum werden diese Kanten eine hohe Kanten-Betweenness haben. Durch das Entfernen dieser Kanten, können die Communities separiert werden und es kann die zugrunde liegende Community-Struktur aufgedeckt werden. Der Algorithmus wird darum auch *Edge-Remover-Clusterer* genannt.
Der Edge-Remover-Algorithmus für gewichtete Graphen ist wie folgt anzuwenden [Newm04]:

1. Berechnung der Betweenness für alle Kanten in einem Netzwerk.
2. Division der Betweenness durch das Gewicht der jeweiligen Kante
3. Entfernung der Kante mit der höchsten resultierenden Betweenness.
4. Neuberechnung der resultierenden Betweenness für alle übrigen Kanten.
5. Wiederholung der Schritte ab Schritt 3 bis keine Kanten übrig bleiben.

Das dargestellte Problem und der Lösungsalgorithmus wurde von uns auf Modularisierung von Architekturen übertragen. Bei Bedarf können die Kanten – insbesondere die der Nutzung und Verbindung von und zwischen IT-Systemen – mit Gewichten versehen werden.

[1] Im Folgenden wird angelehnt an den Originalbeitrag von [GiNe02] weiter von *Betweenness* gesprochen als einem Grad, wie sehr ein Element zwischen den anderen liegt. Es gibt, unseres Erachtens nach, keine entsprechende deutsche Übersetzung.

Durch die wiederholte Anwendung des Algorithmus im Schritt fünf, lässt sich auch eine mehrstufige Modularisierung erreichen. Eine weitere Frage ist, wann eine gute Modularisierung durch das Entfernen von Kanten erreicht ist bzw. wann andernfalls noch weitere Kanten entfernt werden sollten. Girvan/Newman definieren dazu die *Modularität Q* [NeGi04, S. 7 f]. Sie berechnen den Anteil der Kanten eines Graphen welche sich innerhalb der Communities befinden mit [Newm04, S. 7 f]:

$$\frac{\sum_{ij} A_{ij}\delta(c_i,c_j)}{\sum_{ij} A_{ij}} = \frac{1}{2m}\sum_{ij} A_{ij}\delta(c_i,c_j) \qquad (3)$$

Dabei ist c_i die Community der der Knoten *i* angehört. Die Funktion $\delta(u,v)$ ist dann 1 wenn $u = v$, sonst ist sie 0. Weiterhin ist $m = \frac{1}{2}\sum_{ij} A_{ij}$ die Anzahl der Kanten des Graphen. Wenn nun die Grad k_i für alle Knoten *i* beibehalten werden, jedoch die Kanten zufällig im Netzwerk verteilt werden, dann ist die Wahrscheinlichkeit, dass zwischen den Knoten *i* und *j* eine Kante existiert $k_i k_j / 2m$. Daraus ergibt sich die Modularität *Q* zu

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i,c_j)\,. \qquad (4)$$

Daraus ergeben sich für *Q* Werte zwischen 0 und 1. Ein Wert von 0 bedeutet, dass es nach einem Clustering keine Kanten mehr in einer Community gibt, die man dort gemäß dem Zufall erwarten würde. Reale Werte für *Q*, welche ein gutes Clustering vermuten lassen, liegen zwischen 0,3 und 0,7. Prinzipiell können bei der sukzessiven Entfernung von Kanten nach dem Girvan/Newman Algorithmus mehrere lokale Maxima auftreten. Diese Maxima beschreiben gute (hierarchische) Modularisierungen.

4.3 EA Builder Software System

Der Modellierungs- und Clusteringansatz wurde prototypisch im Kontext von Unternehmensarchitekturen in einem Softwaresystem namens EA Builder implementiert (Abbildung 3) [http://www.ea-builder.com].

Im Gegensatz zu vielen der am Markt erhältlichen EA-Tools [Mat+05] unterstützt das Metamodell des EA Builders die Abbildung prozessorientierter Systemintegration über die so genannte dynamische Sicht. Dies ist unabhängig davon, welcher Integrationsansatz gewählt wurde.

Die Leistung des eingesetzten Clusteringalgorithmus wurde an einer Reihe spezieller Testszenarien untersucht [Aier06]. Im Folgenden wird die Funktionsweise des EA Builders anhand des

fiktiven Unternehmens WMYPC (We Make Your PC) demonstriert. Das Unternehmen baut und vertreibt individuell konfigurierte Computersysteme – angefangen beim kleinen Multimediasystem bis hin zu größeren Serversystemen. Die Geschäftsprozesse werden durch sechs individuell implementierte Informationssysteme sowie ein Standardsoftwaresystem unterstützt.

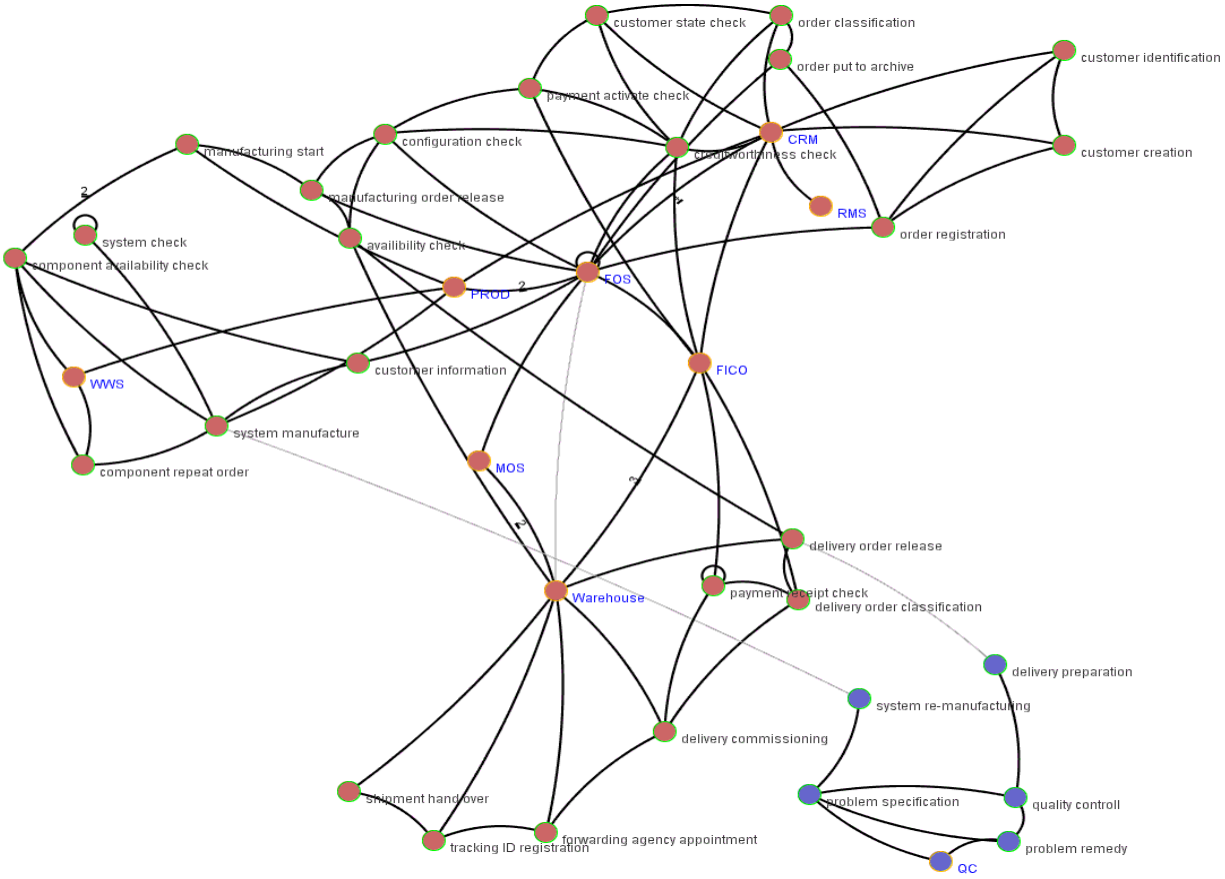


Abbildung 6: Modell transformiert in einen Graphen, drei Kanten entfernt

Abbildung 6 zeigt das zugehörige EA-Modell, welches in einen Graphen transformiert wurde. Bei 20 entfernten Kanten erreicht die Modularitätsfunktion ihr Maximum. Nach der Anwendung des Clusteringalgorithmus und dem Entfernen von 20 Kanten (Abbildung 7) ist der Graph in fünf Cluster aufgeteilt.

Das klassische Szenario für die Anwendung unseres Ansatzes ist die Transformation komplexer IT-Landschaften beispielsweise in eine serviceorientierte Architektur. Der Ansatz kann als Bottom-up-Ansatz bezeichnet werden, da die Cluster aus einer bestehenden Ist-Architektur abgeleitet werden. Verglichen mit einem Top-down-Ansatz führt dies zu einer potenziell evolutionären Architekturtransformation.

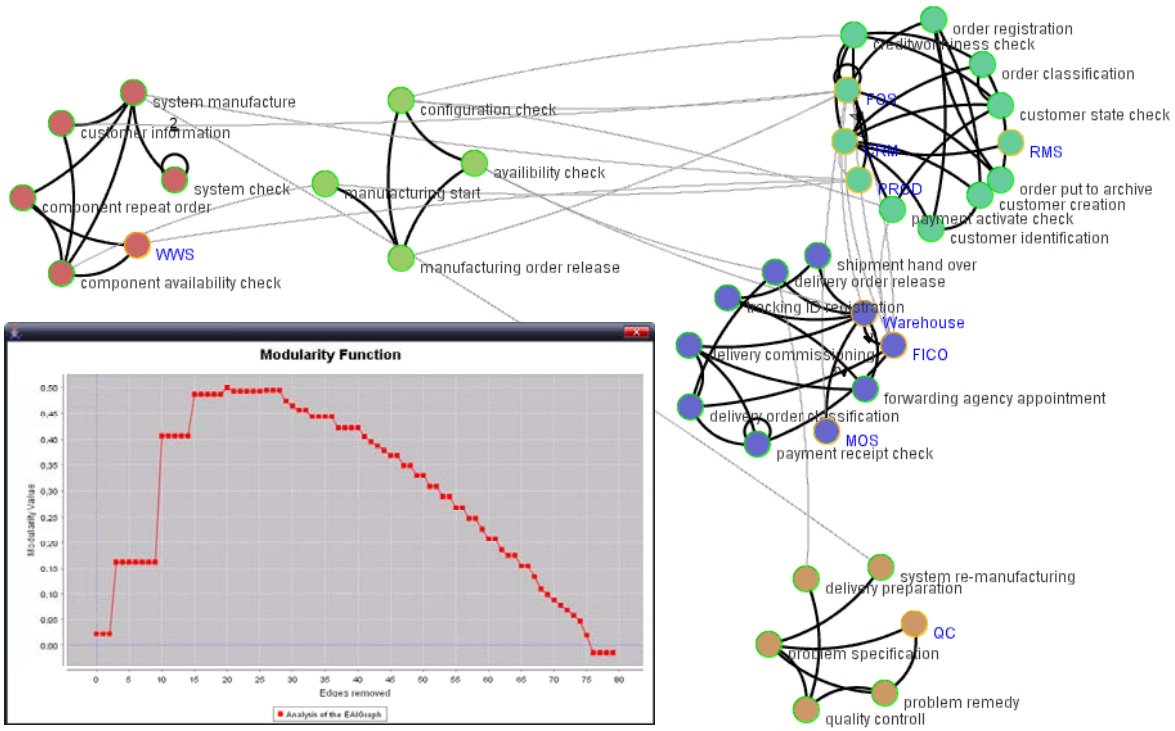


Abbildung 7: Screenshot des EA Builder Modells transformiert in einen Graphen, nach entfernen von 20 Kanten entstehen fünf Cluster, gruppierte Ansicht, Verlauf der Modularity-Function

Die zu Beginn genannte Forderung eines Business-IT-Alignment – hier strukturelle Analogie – wird durch diesen Ansatz unterstützt. Nach der Transformation des EA-Modells in einen Graphen, wird für das Clustering nicht mehr zwischen Prozessen und IT-Systemen differenziert. Das bedeutet, dass auch keine isolierte Optimierung von Prozess- bzw. IT-System-Strukturen erfolgt – obwohl auch diese möglich wäre. Die gebildeten Cluster reflektieren das tatsächliche Zusammenspiel von Prozessen und IT-Systemen und führen damit potenziell zu struktureller Analogie zwischen beiden.

5 Fazit

Ausgehend von der Nachhaltigkeit als Gestaltungsziel für Unternehmensarchitekturen, haben wir die Modularisierung der Architekturen als mögliche Umsetzungsstrategie vorgeschlagen. Betont wurde dabei die Notwendigkeit, Geschäftsprozesse und sie unterstützende IT-Systeme gemeinsam zu betrachten. Dafür wurde das Konzept der strukturellen Analogie genutzt. Basierend auf einem Architekturmodell, bestehend aus den hier als relevant erachteten Elementen, wurde der Einsatz von Clusteringalgorithmen zur Identifikation von Modulen in der Architektur vorgeschlagen. Dieser Ansatz wurde prototypisch im EA Builder-Werkzeug implementiert.

Die so entstandenen Module (Cluster) beschreiben jedoch „nur“ die latent vorhandene Ist-Struktur. Diese Ist-Struktur ist jedoch nicht notwendiger Weise mit dem Organigramm oder einer Karte der Anwendungslandschaft identisch, da es die tatsächliche, nicht die geplante oder vermutete Struktur beschreibt. Abhängig von einer Architektur- oder Integrationsstrategie muss nun entschieden werden, wie mit den Informationen aus dem Clustering weiter verfahren wird. Wird beispielsweise eine serviceorientierte Architektur (SOA) als Paradigma zugrunde gelegt, sind die identifizierten Module Kandidaten für so genannte Service Domains [BaHe04] für welche dann individuelle Integrations- und Managementkonzepte entwickelt werden können.

Der vorgeschlagene Ansatz liefert kein abgeschlossenes Vorgehensmodell für die Gestaltung nachhaltiger Architekturen. Der Nutzen liegt vielmehr darin, latent vorhandene Informationen über komplexe Architekturen zu erhalten, die durch eine einfache Visualisierung nicht sichtbar werden. Diese Informationen bilden dann wiederum die Grundlage für die gezielte Beeinflussung der Architekturen.

Weiterer Forschungsbedarf wird insbesondere in der Analyse des Einflusses weiterer Eigenschaften der modellierten Elemente auf das Clustering-Ergebnis gesehen. So können Elementen und Verbindungen zwischen diesen Attribute gegeben werden, die dann über Gewichte im Clustering berücksichtigt werden. In diesem Zusammenhang erscheint eine Erweiterung des Metamodells um weitere Elementtypen ebenfalls sinnvoll.

Literatur

[Aier06] Aier, Stephan: How Clustering Enterprise Architectures helps to Design Service Oriented Architectures. In: Proceedings of the IEEE International Conference on Services Computing (SCC'06), (Chicago, USA, 2006), IEEE Computer Society, Los Alamitos, CA, USA, S. 269–272.

[AiDo04] Aier, Stephan; Dogan, Turgut: Nachhaltigkeit als Gestaltungsziel von Unternehmensarchitekturen. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Serviceorientierung und nachhaltige Architekturen. Gito, Berlin 2004, S. 75–122.

[AiDo05] Aier, Stephan; Dogan, Turgut: Indikatoren zur Bewertung der Nachhaltigkeit von Unternehmensarchitekturen. In: Ferstl, O.K.; Sinz, E.J. (Hrsg.): Wirtschaftsin-

formatik 2005: eEconomy, eGovernment, eSociety. Physica, Heidelberg 2005, S. 607–626.

[AiSc03] Aier, Stephan; Schönherr, Marten: Flexibilisierung von Organisations- und IT-Architekturen durch EAI. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Management komplexer Architekturen. Gito, Berlin 2003, S. 3–56.

[Bas+03] Bass, Len; Clements, Paul; Kazman, Rick: Software Architecture in Practice. 2. Auflage, Pearson Education Inc., Boston 2003.

[BaHe04] Bath, Uwe; Herr, Michael: Implementation of a service oriented architecture at Deutsche Post MAIL. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Serviceorientierung und nachhaltige Architekturen. Gito, Berlin 2004, S. 279–297.

[Blei91] Bleicher, Knut: Organisation: Strategien, Strukturen, Kulturen. 2. Auflage, Gabler, Wiesbaden 1991.

[Free77] Freeman, L. C.: A Set of Measures of Centrality based upon Betweenness. In: Sociometry 40 (1977) S. 35–41.

[Fres00] Frese, Erich: Grundlagen der Organisation: Konzept – Prinzipien – Strukturen. 8. Auflage, Gabler, Wiesbaden 2000.

[GiNe02] Girvan, Michelle; Newman, Mark E. J.: Community Structure in Social and Biological Networks. In: Proceedings of the National Academy of Science 99 (2002) 12, S. 7821–7826.

[Gron03] Gronau, Norbert: Wandlungsfähige Informationssystemarchitekturen – Nachhaltigkeit bei organisatorischem Wandel. Gito, Berlin 2003.

[Hage03] Hagen, Claus: Integrationsarchitektur der Credit Suisse. In: Aier, Stephan; Schönherr, Marten (Hrsg.): Enterprise Application Integration – Management komplexer Architekturen. Gito, Berlin 2003, S. 61–83.

[HaHu01] Hahn, Dietger; Hungenberg, Harald: PuK – Wertorientierte Controllingkonzepte. 6. Auflage, Gabler, Wiesbaden 2001.

[Hube95] Huber, Joseph: Nachhaltige Entwicklung durch Suffizienz, Effizienz und Konsistenz. In: Fritz, P.; Huber, Joseph; Levi, Hans Wolfgang (Hrsg.): Nachhaltigkeit in naturwissenschaftlicher und sozialwissenschaftlicher Perspektive. Hirzel, Stuttgart 1995, S. 31–46.

[Klam00] Klamma, Ralf: Vernetztes Verbesserungsmanagement mit einem Unternehmensgedächtnis-Repository Fakultät für Mathematik, Informatik und Naturwissenschaften, Rheinisch-Westfälischen Technischen Hochschule Aachen, Aachen, 2000.

[Krca03] Krcal, Hans-Christian: Systemtheoretischer Metaansatz für den Umgang mit Komplexität und Nachhaltigkeit. In: Leisten, Rainer; Krcal, Hans-Cristian (Hrsg.): Nachhaltige Unternehmensführung – Systemperspektiven. Gabler, Wiesbanden 2003, S. 3–30.

[Krcm90] Krcmar, Helmut: Bedeutung und Ziele von Informationssystem-Architekturen. In: Wirtschaftsinformatik 32 (1990) 5, S. 395–402.

[LeSt03] Leitschuh-Fecht, Heike; Steger, Ulrich: Wie wird Nachhaltigkeit für Unternehmen attraktiv? – Business Case für nachhaltige Unternehmensentwicklung. In: Linne, Gudrun; Schwarz, Michael (Hrsg.): Handbuch Nachhaltige Entwicklung. Leske + Budrich, Opladen 2003, S. 257–266.

[LeHu98] Lewin, Arie Y.; Hunter, Starling D.: Information Technology & Organizational Design: A Longitudinal Study of Information Technology Implementations in the U.S. Retailing Industrie, 1980–1996. In: Glaser, Horst; Schröder, Ernst F.; Werder, Axel v. (Hrsg.): Organisation im Wandel der Märkte. Gabler, Wiesbaden 1998, S. 251–286.

[MaRo88] Markus, M. Lynne; Robey, Daniel: Information Technology and Organizational Change: Causal Structure in Theory and Research. In: Management Science 34 (1988) 5, S. 583–589.

[Mat+05] Matthes, Florian; Wittenburg, André; Ernst, Alexander M.; Lankes, Josef; Schweda, Christian M.: Enterprise Architecture Management Tool Survey, Software Engineering for Business Information Systems (sebis), Ernst Denert-Stiftungslehrstuhl, Chair for Informatics 19, TU München, München, 2005.

[Nad+94] Nadler, David A.; Gerstein, Marc S.; Shaw, Robert B.: Organisationsarchitektur. Campus, Frankfurt/Main, New York 1994.

[Newm04] Newman, Mark E. J.: Analysis of Weighted Networks. In: Phys. Rev. E 70 (2004) 056131,

[NeGi04] Newman, Mark E. J.; Girvan, Michelle: Finding and Evaluating Community Structure in Networks. In: Phys. Rev. E 69 (2004) 026113,

[OMa+05] O'Madadhain, Joshua; Fisher, Danyel; Smyth, Padhraic ; White, Scott ; Boey, Yan-Biao: Analysis and Visualization of Network Data using JUNG. In: Journal of Statistical Software (2005) noch nicht erschienen.

[Scha02] Schäffner, Lothar: Der Beitrag der Veränderungsforschung zur Nachhaltigkeit von Organisationsentwicklung. Hampp, München et al. 2002.

[Scot05] Scott, John: Social Network Analysis: A Handbook. 2. Auflage, SAGE, London u.a. 2005.

[Stue96] Stünzer, Lilia: Systemtheorie und betriebswirtschaftliche Organisationsforschung. Duncker & Humblot, Berlin 1996.

[Wall96] Wall, Friederike: Organisation und betriebliche Informationssysteme – Elemente einer Konstruktionslehre. Gabler, Wiesbaden 1996.

[WaFa99] Wasserman, Stanley; Faust, Katherine: Social Network Analysis: Methods and Applications. Cambridge Univ. Press, Cambridge u.a. 1999.

[WiLa06] Winter, Robert; Landert, Karl: IT/Business Alignment als Managementherausforderung - Editorial zum Schwerpunktthema. In: Wirtschaftsinformatik 48 (2006) 5, S. 309.