

Association for Information Systems AIS Electronic Library (AISeL)

Wirtschaftsinformatik Proceedings 2009

Wirtschaftsinformatik

2009

ANALYSE VON MEINUNGEN IN SOZIALEN NETZWERKEN DES WEB 2.0

Carolin Kaiser

Universität Erlangen-Nürnberg

Follow this and additional works at: <http://aisel.aisnet.org/wi2009>

Recommended Citation

Kaiser, Carolin, "ANALYSE VON MEINUNGEN IN SOZIALEN NETZWERKEN DES WEB 2.0" (2009). *Wirtschaftsinformatik Proceedings 2009*. 111.

<http://aisel.aisnet.org/wi2009/111>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2009 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

ANALYSE VON MEINUNGEN IN SOZIALEN NETZWERKEN DES WEB 2.0

Carolin Kaiser¹

Kurzfassung

Im Web 2.0 bilden sich vermehrt soziale Netzwerke, in denen Nutzer Meinungen zu Produkten austauschen. Für Unternehmen ist es wichtig zu wissen, wie sich die Meinungen zu ihren Produkten in der Kommunikation zwischen den Nutzern der sozialen Netzwerke entwickeln um frühzeitig Chancen und Risiken zu erkennen. Eine manuelle Meinungsanalyse ist jedoch sehr aufwändig und nur in geringem Umfang möglich. In dieser Arbeit wird ein System vorgestellt, das Meinungen und Kommunikationsbeziehungen in Foren mithilfe von Text Mining automatisch extrahiert und mittels der sozialen Netzwerkanalyse Meinungsführer und Meinungstrends identifiziert.

1. Motivation

Das Internet entwickelt sich zunehmend von einem statischen zu einem interaktiven Medium. Nutzer haben nun die Möglichkeit neben der Informationsbeschaffung auch aktiv an der Gestaltung von Inhalten mitzuwirken. Dieser Wandel zum so genannten Web 2.0 wird durch verschiedene technische, wirtschaftliche und soziale Faktoren begünstigt [7]. Steigende Datenübertragungsraten zu sinkenden Preisen erlauben die Entwicklung und massenhafte Nutzung von interaktiven Webanwendungen. Das neue Selbstbewusstsein der Informationsgesellschaft führt dazu, dass sich immer mehr Nutzer zu sozialen Netzwerken zusammenschließen, um Meinungen auszutauschen. Sie interagieren miteinander und beeinflussen sich gegenseitig in ihrer Meinungsbildung. Für Unternehmen sind Meinungen bezüglich der eigenen Produkte von großer Bedeutung, um Chancen und Risiken von Produkten erkennen und Marketingmaßnahmen ergreifen zu können.

Die Diffusionstheorie beschäftigt sich mit der Verbreitung von Meinungen zu neuen Produkten in sozialen Systemen. Nach der Diffusionstheorie ist neben den Eigenschaften der betrachteten Produkten (z.B. Komplexität) die Struktur des sozialen Netzwerks ein wichtiger Einflussfaktor bei der Meinungsbildung [13]. Die Struktur wird durch die Kommunikationsbeziehungen zwischen den Akteuren des sozialen Netzwerks bestimmt. Die Tatsache, wer mit wem kommuniziert, hat damit Einfluss auf die Meinungsbildung. Um die Meinungsbildung in sozialen Netzwerken im Web 2.0 einzuschätzen, ist eine Untersuchung der Meinung einzelner Akteure sowie ihrer Kommunikationsbeziehungen erforderlich. Eine manuelle Analyse ist auf Grund der vielen Kommunikationsplattformen und Benutzer im Web 2.0 sehr aufwändig und nur begrenzt möglich. Eine automatisierte

¹ Lehrstuhl Wirtschaftsinformatik II, Universität Erlangen-Nürnberg

Analyse ist daher notwendig. In dieser Arbeit wird ein System vorgestellt, das es ermöglicht mittels Text Mining Meinungen und Kommunikationsbeziehungen in Foren automatisch zu extrahieren und mithilfe der sozialen Netzwerkanalyse Meinungsführer und Meinungstrends zu erkennen.

2. Themenverwandte Arbeiten

Aktuell entwickeln sich das Text Mining und die soziale Netzwerkanalyse zu beliebten Forschungsthemen. Text Mining ermöglicht die Entdeckung von interessantem Wissen aus Texten [9] und untergliedert sich in die Teilbereiche Informationssuche, Textklassifikation, Text-Clustering und Informationsextraktion [16]. Während früher der Fokus auf der Aufdeckung von Fakten lag, stehen heute zunehmend Emotionen und Meinungen im Vordergrund [9]. Es existieren bereits einige Arbeiten, die sich mit dem Einsatz von Text Mining zur Aufdeckung von Meinungen im Web 2.0 beschäftigen. Die Arbeiten von Dave et al. [6], Liu et al. [10] sowie Popescu und Etzioni [11] zum Beispiel beschreiben Systeme, die es ermöglichen Kundenmeinungen zu Produkten in Reviews zu analysieren. Hierbei werden die Meinungen in den Reviews zunächst einzeln ermittelt und anschließend aggregiert. Kommunikationsbeziehungen zwischen Kunden, die bei der Bildung von Meinungen eine große Rolle spielen, werden in keiner der Arbeiten betrachtet. Die soziale Netzwerkanalyse ermöglicht es Muster in Beziehungen zwischen interagierenden Akteuren zu analysieren [15]. Sie gliedert sich in eine individuelle Analyse, die die Handlungsmuster einzelner Akteure untersucht, und eine kollektive Analyse, die die Handlungsstruktur eines gesamten Netzwerks betrachtet. Soziale Netzwerkanalysen werden zunehmend auf das Web 2.0 angewandt. Die Arbeiten von Welser et al. [17], Chang et al. [4] und Gomez et al. [8] zum Beispiel untersuchen Rollen, Positionen und Kommunikationsmuster in Online-Konversationen. Die Inhalte der Konversationen werden jedoch hierbei nicht berücksichtigt. Eine integrierte Betrachtung von Kommunikationsinhalten und -beziehungen mittels Text Mining und sozialer Netzwerkanalyse fehlt. Ziel dieses Projektes ist die Entwicklung eines Systems, das Meinungen und Beziehungen zwischen Forenbenutzern mittels Text Mining aufdeckt und anschließend zentrale Einflussgrößen der Meinungsbildung im Netzwerk mittels sozialer Netzwerkanalyse untersucht.

3. Systemaufbau

Das System zur Analyse von Meinungen in sozialen Netzwerken (siehe *Abbildung 1*) besteht aus zwei Komponenten, die nacheinander aufgerufen werden.

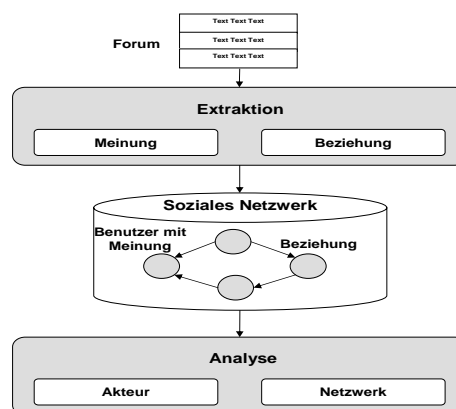


Abbildung 1: Gesamtsystem

Die Extraktionskomponente erhält als Input Foren, die von einem Crawler aus dem Internet heruntergeladen werden. Daraus extrahiert sie zunächst mithilfe von Text-Mining-Methoden für

jeden Beitrag eines Forums den Benutzernamen sowie seine Meinung bezüglich eines Produktes. Anschließend werden die Kommunikationsbeziehungen zwischen den Benutzern mittels textbasierten Relationship Mining Methoden identifiziert. Die extrahierten Benutzer, Meinungen und Beziehungen bilden ein soziales Netzwerk und werden in einer Datenbank in Form von Graphen gespeichert. Graphen bestehen aus einer Menge von Knoten und Kanten, wobei die Kanten Verbindungen zwischen Knoten darstellen. Darüberhinaus können Attribute Knoten und Kanten näher charakterisieren. Hierbei repräsentieren die Knoten die Benutzer eines Forums und die Kanten die Kommunikationsbeziehungen. Die Knotenattribute geben die Meinungen der Benutzer bezüglich des betrachteten Produkts wieder, während die Kantenattribute die Häufigkeit der Kommunikation ausdrücken. Die so entstandenen Graphen dienen als Input für die Analysekomponente. Diese ermittelt mithilfe von Kennzahlen der sozialen Netzwerkanalyse die Positionen der Forenbenutzer und Struktur des gesamten Netzwerks. Auf diese Weise können Meinungsführer und Meinungstrends erkannt werden. Die Komponenten werden am Beispiel von Foren, in denen Meinungen zum iPhone® von Apple ausgetauscht werden, illustriert und validiert.

4. Extraktion von sozialen Netzwerken

4.1. Extraktion der Meinungen von Akteuren

Zielsetzung der Extraktion der Meinungen ist es, die Einstellung eines Akteurs bezüglich eines Produktes aus seinen Äußerungen in einem Forumsbeitrag zu erkennen. Einstellungen können grundsätzlich hinsichtlich der Polarität (positiv versus negativ) und des Intensitätsgrades (sehr versus wenig) charakterisiert werden [6]. Da der Intensitätsgrad von Einstellungen schwer objektiv messbar ist, wird in diesem Projekt nur die Polarität ermittelt. Sie wird mithilfe der drei Klassen positiv, neutral und negativ modelliert.

Die automatische Erkennung von positiven, negativen und neutralen Meinungen mithilfe von Text Mining setzt sich aus zwei Phasen zusammen, der Extrahierung von Merkmalen aus dem Text und der Anwendung eines Lernalgorithmus zur Erkennung der Polarität anhand der extrahierten Merkmale.

Die Extrahierung der Merkmale kann generell mit einer linguistischen oder statistischen Analyse erfolgen [16]. Im Rahmen dieses Projekts wird eine kombinierte Analyse durchgeführt. Hierbei wird der Text eines Beitrags zunächst in einzelne Wörter zerlegt. Nach der Entfernung unwichtiger Stoppwörter (z.B. Artikel) werden die verbliebenen Wörter auf ihren Wortstamm reduziert und die Häufigkeit aller Wortstämme ermittelt. Als Merkmale für die Beiträge werden die Wortstämme zu Grunde gelegt, die besonders charakteristisch für die drei Klassen sind, d.h. die oft in einer Klasse, aber nur selten in den anderen Klassen vorkommen.

Zur Erkennung der Polarität der Forenbeiträge können prinzipiell überwachte und unüberwachte Lernverfahren eingesetzt werden. Beim überwachten Lernen erlernt der Algorithmus mithilfe von Beispieldatensätzen die Polarität eines Beitrags anhand seiner Merkmale zu erkennen. Beim unüberwachten Lernen werden dem Algorithmus Muster vorgegeben, anhand derer er die Polarität bestimmt. Auf Grund der besseren Lernfähigkeit kommt in diesem Projekt ein überwachtes Lernverfahren zum Einsatz. Hierbei kann die Lernaufgabe als Klassifikation angesehen werden. Die Forenbeiträge werden anhand des Auftretens charakteristischer Wortstämme in die Klassen positiv, negativ und neutral eingeteilt. *Abbildung 2* zeigt beispielhaft, wie ein Forumsbeitrag der Klasse positiv zugeordnet wird.



Abbildung 2: Klassifikation von Forenbeiträgen

Zur Klassifikation können verschiedene Methoden, wie zum Beispiel Hidden Markov Models oder Maximum Entropy, zum Einsatz kommen [16]. In diesem Projekt wird die Stützvektormethode [5] herangezogen, da sie eine große Anzahl an Merkmalen verarbeiten kann. Sie erhält als Input Beispieldatensätze, bestehend aus einer Menge an Forenbeiträgen mit zugehörigen Merkmalswerten und zugeordneter Klasse. Mit diesen Datensätzen erlernt sie die Parameter einer Regel, die die Beiträge möglichst gut klassifiziert. Die Regel ermöglicht eine Einteilung der Beiträge in zwei Klassen (binäre Klassifikation). Im Fall von drei Klassen, können drei Regeln erlernt werden: positiv versus nicht positiv, negativ versus nicht negativ und neutral versus nicht neutral. Ein Beitrag wird in die Klasse eingeteilt, die die höchste Wahrscheinlichkeit aufweist. Die Regel lässt sich im einfachen zweidimensionalen Fall als Gerade veranschaulichen. Alle Beiträge, die auf der einer Seite der Gerade liegen, gehören zur ersten Klasse, alle Beiträge, die auf der anderen Seite der Geraden liegen, gehören zur zweiten Klasse. *Abbildung 3* zeigt, wie eine Gerade vier Beispieldatensätze anhand des Vorkommens der Wortstämme „lieb“ und „super“ in die Klassen positiv und nicht positiv einteilt.

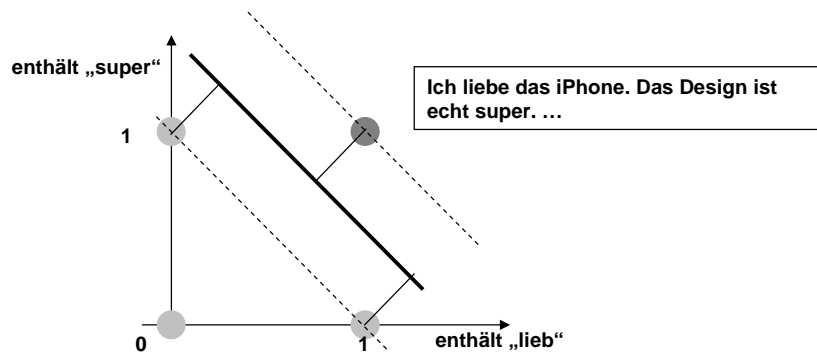


Abbildung 3: Klassifikation eines Forenbeitrags mittels der Stützvektormethode

Nach der Klassifikation werden alle Meinungen den Forumsbenutzern zugeordnet. Verfügt ein Benutzer über verschiedene Meinungen, so wird ihm seine häufigste Meinungsklasse zugewiesen.

Zur Validierung der Meinungsklassifikation wurden 929 Forenbeiträge zum Thema iPhone herangezogen und manuell in die Klassen positiv, negativ und neutral eingeteilt. Die Validierung erfolgt mittels einer fünffachen Kreuzvalidierung. Hierbei werden die 929 Forenbeiträge in fünf ungefähr gleich große Teile aufgeteilt. In fünf verschiedenen Validierungsschleifen werden jeweils vier Teile dazu verwendet die Klassifikationsregeln zu erlernen und der fünfte Teil wird zum Testen der erlernten Klassifikationsregeln genutzt. Am Ende der fünf Durchgänge wird die durchschnittliche Klassifikationsgenauigkeit für jede Klassifikationsregel berechnet. Die Klassifikationsgenauigkeit gibt den Anteil der korrekt klassifizierten Forenbeiträge an allen Forenbeiträgen an. Die Ergebnisse der Validierung sind in *Tabelle 1* dargestellt. Aus der Tabelle wird ersichtlich, dass nicht nur die Anzahl der Beispieldatensätze für den Lernerfolg entscheidend sind, sondern auch die Variation der Sprache. Während positive und negative Meinungen eher mit ähnlichen Wörtern beschrieben werden, werden neutrale Meinungen mit einer großen Vielfalt an Worten ausgedrückt.

Tabelle 1: Ergebnisse der Meinungsklassifikation

Meinung	Anzahl	Genauigkeit
positiv	231	0,79
negativ	226	0,75
neutral	472	0,70

4.2 Extraktion der Beziehungen zwischen Akteuren

Zielsetzung der Extraktion der Beziehungen zwischen Akteuren ist es, Kommunikationsbeziehungen in einem Forum zu erkennen. Dabei besteht eine Kommunikationsbeziehung zwischen zwei Akteuren, wenn sich der eine Akteur (Sender) in mindestens einem Forenbeitrag auf einen anderen Akteur (Empfänger) bezieht. Eine Kommunikationsbeziehung wird durch den Sender, den Empfänger und die Kommunikationshäufigkeit charakterisiert.

Im Bereich des Text Mining existieren zwei Ansätze, die sich mit der Erkennung von Beziehungen zwischen verschiedenen Entitäten in Texten beschäftigen: Koreferenzauflösung und Beziehungsextraktion [16]. Die Koreferenzauflösung versucht zu erkennen, ob sich zwei Wörter in einem Text auf das gleiche Objekt beziehen. Typische Anwendung ist die Bestimmung, ob sich ein bestimmtes Pronomen auf ein zuvor erwähntes Namenswort bezieht. In den Sätzen „Das iPhone ist toll. Es hat ein großes Display.“ muss zum Beispiel entschieden werden, ob sich das Wort „iPhone“ und „es“ auf das gleiche Objekt beziehen. Die Beziehungsextraktion beschäftigt sich mit der Identifikation verschiedenartiger Beziehungen zwischen zwei Bestandteilen eines Satzes. Aufgabe kann es sein, Beziehungen der Art „ist Hersteller von“ herauszufinden. In dem Satz „Das iPhone von Apple ist super.“ ist zu erkennen, dass Apple der Hersteller vom iPhone ist. Abstrakt betrachtet besteht das Ziel darin, ein binäres Klassifikationsproblem zu lösen. Jedes Paar an möglichen Textelementen wird in die zwei Klassen „hat vordefinierte Beziehung“ und „hat keine vordefinierte Beziehung“ eingeteilt. Die Klassifikation basiert dabei auf statistisch-linguistischen Attributwerte des einen und des anderen Textelements sowie ihrer gemeinsamen Attributwerte. Zur Klassifikation können Methoden des überwachten und unüberwachten Lernens eingesetzt werden.

Im Rahmen der Klassifikation von Kommunikationsbeziehungen von Akteuren in einem Forum wird für jeden Forumsbeitrag eines Akteurs entschieden, ob eine Kommunikationsbeziehung zu einem vorhergehenden Beitrag eines anderen Akteurs besteht. *Abbildung 4* veranschaulicht die Bildung aller möglichen Paare aus drei Forenbeiträgen und deren Klassifikation.

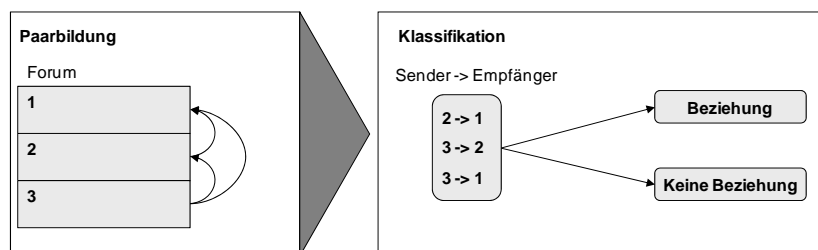


Abbildung 4: Klassifikation von Kommunikationsbeziehungen

Wichtige Hinweise auf das Vorhandensein einer Kommunikationsbeziehung sind die Erwähnung des Benutzernamens des Empfängers im Beitrag des Senders, das Zitieren von Textelementen des Empfängers durch den Sender, die Nähe der Beiträge im Forum und das Auftreten charakteristischer Worte in beiden Beiträgen. Daher werden diese Kriterien mittels String Matching Verfahren überprüft und als Attribute zur Klassifikation herangezogen. Die Klassifikation wird wie

bei der Erkennung von Meinungen mithilfe der Stützvektormethode realisiert. Nach der Klassifikation wird die Beziehungshäufigkeit aller Paare durch Addition ermittelt.

Die Validierung des Ansatzes erfolgt analog zur Validierung der Meinungsklassifikation mit dem gleichen Datensatz. Dieser besteht aus 17 Threads zum Thema iPhone mit durchschnittlich 55 Beiträgen von 22 Nutzern. Zur Validierung wurden die jeweils durchschnittlich 59 vorhandenen von 2082 möglichen Kommunikationsbeziehungen manuell markiert. Bei der Validierung konnte eine Klassifikationsgenauigkeit von rund 85% erreicht werden.

5. Analyse von Meinungen in sozialen Netzwerken

5.1 Analyse auf Akteurebene

Die extrahierten Forenbenutzer mit ihren Meinungen und Kommunikationsbeziehungen bilden ein soziales Netzwerk, das zunächst auf Akteurebene untersucht wird. Zielsetzung der Analyse auf Akteurebene ist es Meinungsführer in Foren zu identifizieren. Meinungsführer sind Akteure, die eine zentrale Rolle bei der Meinungsbildung einnehmen. Durch ihre besondere Position und ihr Kommunikationsverhalten im sozialen Netzwerk sind sie in der Lage die Meinung von vielen anderen Akteuren zu beeinflussen [2]. Sie bilden den Orientierungspunkt für Akteure, die nach Informationen suchen. Die Identifikation von Meinungsführern bietet ein hohes Potential für das Marketing. So kann die Orientierungsfunktion bei Meinungsführern mit positiver Meinung genutzt werden um über sie gezielt Informationen zu verbreiten. Meinungsführern mit negativer Meinung kann man versuchen moderativ entgegenzuwirken.

Im Bereich der sozialen Netzwerkanalyse existiert eine große Anzahl an Kennzahlen, die die Position und das Kommunikationsverhalten von Akteuren näher charakterisieren. Bei der Betrachtung von Meinungsführern sind insbesondere die Zentralitätskennzahlen von Bedeutung. Sie messen auf einer Skala von 0 bis 1 wie zentral ein Akteur für ein Netzwerk ist [14]. Dabei können drei verschiedene Kennzahlen unterschieden werden: Degree-Zentralität, Closeness-Zentralität und Betweenness-Zentralität. Die Degree-Zentralität misst wie aktiv ein Akteur mit anderen auf direktem Weg kommuniziert. Sie wird berechnet als Anteil der Summe aller ein- und ausgehenden Beziehungen eines Akteurs an der Summe aller Beziehungen [15]. Die Closeness-Zentralität erfasst die Nähe eines Akteurs zu allen anderen Akteuren. Sie ist definiert als inverse Summe aller Pfaddistanzen vom betrachteten Akteur zu allen anderen Akteuren [15]. Mithilfe der Betweenness-Zentralität wird ermittelt, wie häufig ein Akteur auf den kürzesten Verbindungspfaden zwischen anderen Akteurspaaren liegt. Die Betweenness-Zentralität eines Akteurs errechnet sich als Anteil der kürzesten Pfade, die über den Akteur laufen, zu allen kürzesten Pfaden [15].

Die Degree-Zentralität stellt einen Indikator für lokale Meinungsführer dar. Forumsbenutzer mit hoher Degree-Zentralität kommunizieren oft direkt mit anderen Benutzern und haben die Möglichkeit diese zu beeinflussen. Ihr Einfluss ist jedoch auf die direkte Umgebung begrenzt. Closeness- und Betweenness-Zentralität hingegen betrachten nicht nur direkte sondern auch indirekte Kommunikationsbeziehungen und stellen daher einen Indikator für globale Meinungsführer dar. Forumsbenutzer, die die kürzeste Distanz zu allen anderen Benutzern aufweisen, werden mit der höchsten Wahrscheinlichkeit von allen wahrgenommen. Sie sind am unabhängigsten von der Informationsweitergabe durch andere Benutzer. Forumsbenutzer, die am häufigsten auf Kommunikationspfaden zwischen anderen Benutzern liegen, verfügen über die beste Möglichkeit Informationen innerhalb dieser Kommunikation zu kontrollieren. Diese Zusammenhänge veranschaulicht auch der in *Abbildung 5* dargestellte Thread aus dem Forum computerbase.de.

Der Forumsbenutzer 28 weist mit einem Wert von 0,13 die höchste Degree-Zentralität im Thread auf. Seine positive Meinung zum iPhone spiegelt aber nur in seiner näheren Umgebung wider. Benutzer 34 hat mit einer Closeness- und Betweenness-Zentralität von jeweils 0,3 die größte Nähe zu allen anderen Benutzern und die beste Informationskontrolle. Seine negative Meinung stellt in der Gesamtmeinung des Forums die häufigste Meinungsklasse mit 49% dar (im Vergleich: 34% positive und 17% neutrale Meinungen).

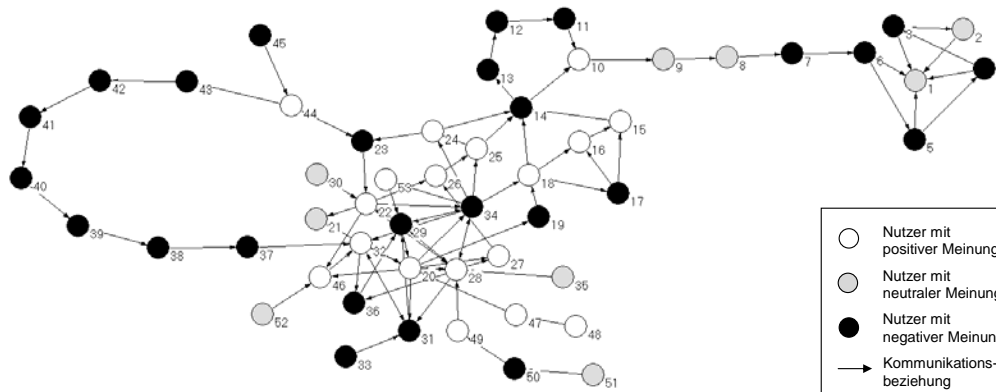


Abbildung 5: Thread aus dem Forum computerbase.de

Zur ersten Validierung wurden 17 Threads aus Foren zum Thema iPhone herangezogen und für die Netzwerke die Meinung der Benutzer mit der höchsten Closeness- und Betweenness-Zentralität sowie die Gesamtmeinung des Forum in Form der häufigsten Meinungsklasse ermittelt. Die Validierung erfolgt mithilfe des Spearmanschen Korrelationskoeffizienten, der Richtung und Stärke des Zusammenhangs zwischen ordinal skalierten Variablen auf einer Skala von -1 (negativer perfekter Zusammenhang) bis 1 (positiver perfekter Zusammenhang) misst [1]. Dabei zeigt ein Wert von 0 die Unabhängigkeit der Variablen an. In den Sozialwissenschaften werden Beträge größer als 0,3 als bedeutsame Zusammenhänge eingestuft [3]. Mithilfe des Spearmanschen Korrelationskoeffizienten konnte zwischen der Meinung des Benutzers mit der höchsten Closeness-Zentralität und der Gesamtmeinung ein positiver bedeutsamer Zusammenhang von 0,56 ermittelt werden. Für den Zusammenhang zwischen der Meinung des Benutzers mit der höchsten Betweenness-Zentralität und der Gesamtmeinung ergab sich ein positiv bedeutsamer Zusammenhang von 0,49.

5.2 Analyse auf Netzwerkebene

Ziel der Analyse auf Netzwerkebene ist die Erklärung und Prognose der Meinungsbildung eines Forums. Hierbei sollen insbesondere zwei Aspekte der Meinungsbildung näher betrachtet werden. Zum einen wird untersucht, bei welchen Charakteristika eines Netzwerks sich eine klare Tendenz der Gesamtmeinung (d.h. ein Meinungstrend) vorherrscht. Zum anderen wird analysiert, bei welchen Netzwerkeigenschaften sich die Meinung des Meinungsführers als Trend durchsetzt. Bei einer frühen Erkennung von positiven und negativen Trends können rechtzeitig verstärkende oder hemmende Marketingmaßnahmen eingesetzt und eventuelle Produktverbesserungen durchgeführt werden.

Die soziale Netzwerkanalyse stellt eine Vielzahl an Kennzahlen bereit, die die Struktur des gesamten Netzwerks näher charakterisieren. Für die Meinungsanalyse sind insbesondere die Kennzahlen Dichte, Randic Konnektivität und Closeness-Zentralisierung von besonderer Bedeutung. Die Dichte mißt die Verbundenheit eines Netzwerkes und stellt damit einen Indikator für die Kommunikations-

verteilung im Forum dar. Sie errechnet sich als Anzahl der vorhandenen Kommunikationsbeziehungen zu der Anzahl der maximal möglichen Kommunikationsbeziehungen [14]. Je höher die Dichte desto mehr Informationen können im Netzwerk verteilt werden. Die Randic Konnektivität charakterisiert den Verzweigungsgrad im Netzwerk. Sie wird als Summe der inversen Wurzeln aus den multiplizierten Knotengraden aller Akteurspaare berechnet [12]. Dabei gibt ein kleiner Wert der Kennzahl einen starken Verzweigungsgrad an, d.h. dass viele Akteure mit vielen Akteuren direkt kommunizieren. Die Closeness-Zentralisierung mißt die Variation der Closeness-Zentralität der Akteure im Netzwerk. Sie ist definiert als Verhältnis aus der Summe der Abweichungen der Closeness-Zentralität aller Akteure von der Closeness-Zentralität des zentralsten Akteurs und der theoretisch maximal möglichen Zentralisierung [15]. Ein Netzwerk ist stark zentralisiert, wenn es wenig zentrale Akteure und viele periphere Akteure gibt.

Die Kennzahlen Dichte und Randic Konnektivität werden als Indikatoren herangezogen, ob eine klare Tendenz der Gesamtmeinung in einem Forum zu erwarten ist. Je höher die Dichte und je niedriger die Randic Konnektivität desto mehr Forenbenutzer sind miteinander vernetzt und kommunizieren direkt miteinander. So können Meinungen vermehrt ausgetauscht werden. Damit steigt die Wahrscheinlichkeit, dass sich eine starke Tendenz der Gesamtmeinung bildet. Die Gesamtmeinung setzt sich aus den Anteilen der positiven, negativen und neutralen Meinungen der User bezüglich eines Produkts zusammen. Die Tendenz ist umso stärker, je mehr ein Meinungsanteil die anderen Anteile übersteigt. Die Tendenz wird als Produkt der Abweichungen des größten Meinungsanteils von den beiden kleineren Anteilen gemessen. *Abbildung 6* veranschaulicht den Zusammenhang zwischen Dichte, Randic Konnektivität und Tendenz der Gesamtmeinung am Beispiel von zwei Threads zum Thema iPhone.

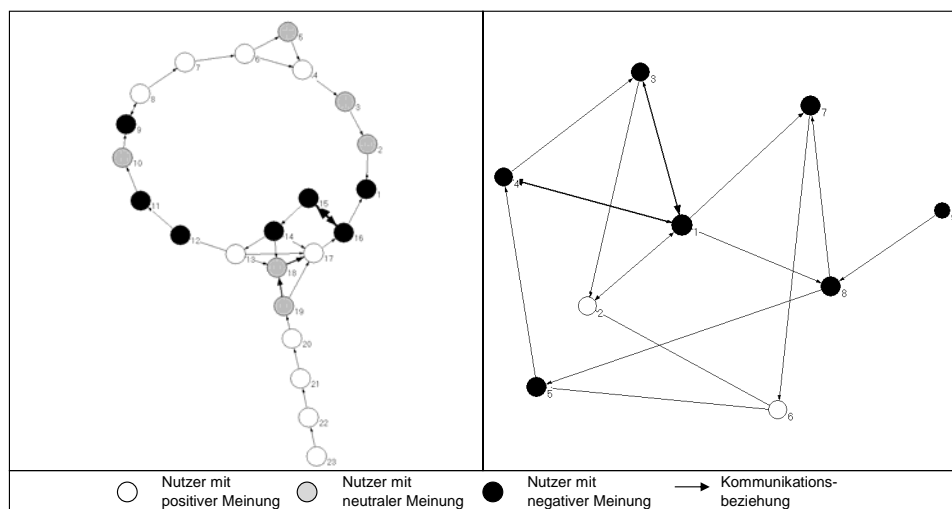


Abbildung 6: Threads aus den Foren slashfm.de und pcfreunde.de

Der Thread aus dem Forum slashfm.de (*Abbildung 6 links*) hat eine relativ niedrige Dichte von 0,08 und eine relativ hohe Randic Konnektivität von 11,33. Dies geht mit einer niedrigen Tendenz der Gesamtmeinung von 0,03 einher. 44% der Benutzer haben eine positive, 30% eine negative und 26% eine neutrale Meinung zum iPhone. Der Thread aus dem Forum pcfreunde.de (*Abbildung 6 rechts*) hingegen weist eine relativ hohe Dichte von 0,22 und eine niedrige Randic Konnektivität von 4,22 auf. Dies geht mit einer relativ hohen Tendenz der Gesamtmeinung von 0,43 einher. 78% der Forenbenutzer beurteilen das iPhone negativ, 22% bewerten es positiv.

Als Indikator, ob eine starke Tendenz von einem Meinungsführer ausgelöst wurde, kann zusätzlich die Zentralisierung verwendet werden. Ist die Closeness-Zentralisierung hoch, so gibt es nur einen

oder wenige Meinungsführer im Zentrum des Forums und viele Akteure an der Peripherie. Dadurch kann die Meinung des Meinungsführers gut verbreitet werden. Ist das Netz zusätzlich dicht vernetzt, so kann die Meinung des Meinungsführers zwischen vielen anderen Forenbenutzern intensiv kommuniziert werden. Hierdurch kann ein Trend in Richtung der Meinung des Meinungsführers entstehen. Der Thread aus dem Forum pcfreunde.de in *Abbildung 6* illustriert diesen Zusammenhang. Die Closeness-Zentralisierung ist mit einem Wert von 0,4 relativ hoch. Hier herrscht die negative Meinung des Meinungsführer (Benutzer 1), der die höchste Closeness-Zentralität von 0,7 besitzt, bei 78% der Benutzer vor.

Zur ersten Validierung der Zusammenhänge wurden 17 Threads zum Thema iPhone zu Grunde gelegt und die Kennzahlen Dichte, Randic Konnektivität, Closeness-Zentralisierung, Meinung des Meinungsführers, Tendenz und Richtung der Gesamtmeinung ermittelt. Anschließend wurden die Zusammenhänge mithilfe der Korrelationskoeffizienten von Pearson und dem Kontingenzkoeffizienten von Cramer überprüft. Der Korrelationskoeffizient von Pearson ermöglicht es Stärke und Richtung von Zusammenhängen zwischen quantitativen Variablen auf einer Skala von -1 (perfekter negativer Zusammenhang) bis 1 (perfekter positiver Zusammenhang) zu messen [1]. Ein Wert von 0 gibt an, dass kein Zusammenhang besteht. Der Kontingenzkoeffizient von Cramer misst die Stärke des Zusammenhangs zwischen zwei nominal skalierten Variablen auf einer Skala von 0 (kein Zusammenhang) bis 1 (perfekter Zusammenhang) [1]. Mithilfe des Pearsonschen Korrelationskoeffizienten konnte ein positiver bedeutsamer Zusammenhang von 0,54 zwischen Dichte und Tendenz sowie ein positiver bedeutsamer Zusammenhang von 0,50 zwischen Randic Konnektivität und Tendenz errechnet werden. Mittels des Kontingenzkoeffizienten von Cramer wurde ein bedeutsamer Zusammenhang zwischen den Netzwerkeigenschaften hohe Dichte, niedrige Randic Konnektivität und hohe Zentralisierung sowie der Bildung einer Meinungstendenz in Richtung der Meinung der Meinungsführers im Wert von 0,55 gefunden.

6. Zusammenfassung und Ausblick

Die klassische Form des Internet wandelt sich immer mehr von einer reinen Informationsquelle zu einer interaktiven Plattform. Kunden gruppieren sich im Web 2.0 zu sozialen Netzwerken, in denen sie über Produkte diskutieren und sich gegenseitig in ihrer Meinungsbildung beeinflussen. Für Unternehmen sind Meinungen bezüglich der eigenen Produkte von großer Bedeutung, um Chancen und Risiken einschätzen zu können. Eine Analyse lässt sich manuell nur in sehr begrenztem Umfang durchführen. Daher ist eine automatisierte Analyse nötig. Ausgehend von dem Mangel eines bestehenden Systems, wurde ein automatischer Ansatz zur Analyse von Meinungen in Foren entwickelt, der es ermöglicht Meinungen und Beziehungen aus Foren automatisch zu extrahieren und Meinungsführer und Meinungstrends zu identifizieren. Zur Realisierung kommen Text Mining Methoden und Kennzahlen der sozialen Netzwerkanalyse zum Einsatz. Mithilfe des Systems können positive Meinungsbilder genutzt werden, indem frühzeitig verstärkende Marketingmaßnahmen eingesetzt werden. Bei negativen Meinungsbildern können zeitnah Gegenmaßnahmen in Form von Produktverbesserungen oder Marketingaktionen ergriffen werden. Das System stellt ein Basiskonzept zur Meinungsanalyse in sozialen Netzwerken dar. Zukünftige Aufgaben sind die Vergrößerung der Validierungsdatenbasis. Hierbei sollen auch Foren zu anderen Produktkategorien betrachtet werden. Weiteres Ziel ist es die statische soziale Netzwerkanalyse um eine dynamische Netzwerkanalyse zu ergänzen. Hierbei sollen von verschiedenen aufeinander folgenden Zeitpunkten Momentaufnahmen von dem sozialen Netzwerk gemacht und Veränderungen analysiert werden. Auf individueller Ebene können auf diese Weise Meinungswechsler identifiziert und auf kollektiver Ebene Trendänderungen erkannt werden.

7. Literaturverzeichnis

- [1] BAMBERG, G., BAUR, F., RAPP, M., Statistik, 13. Auflage, Oldenburg Verlag, München 2007.
- [2] BÄNSCH, A.: Käuferverhalten, 9. Auflage, Oldenburg Verlag, München 2002.
- [3] BORTZ, J., Lienert, G., Boehnke, L., Verteilungsfreie Methoden in der Biostatistik, Springer Verlag, Berlin 2000.
- [4] CHANG, C. L., CHEN, D. Y., CHUANG, T. R., Browsing Newsgroups with a Social Network Analyzer, in: Proceedings of the Sixth International Conference on Information Visualization, London 2002.
- [5] CORTES, C.; VAPNIK, V. N., Support Vector Networks, in: Machine Learning, Bd. 20 (1995), S. 273-297.
- [6] DAVE, K., LAWRENCE, S., PENNOCK, D. M., Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews, in: Proceedings of the Twelfth International Conference on World Wide Web, ACM Press, Budapest 2003, S. 519–528.
- [7] GEHRKE, G., Web 2.0 – Schlagwort oder Megatrend? Fakten, Analysen, Prognosen, in: Schriftreihe Medienkompetenz des Landes Nordrhein-Westfalen, Band 6, kopaed Verlag, 2007
- [8] GOMEZ, V., KALTENBRUNNER, A., LOPEZ, V., Statistical Analysis of the Social Network and Discussion Threads in Slashdot, in: Proceedings of the International World Wide Web Conference, ACM Press, Beijing, 2008.
- [9] KAO, A.; POTEET, S.: Overview, in: KAO, A.; POTEET, S. R. (Hrsg.): Natural Language Processing and Text Mining, Springer Verlag, London 2007, S. 1-7.
- [10] LUI, B. HU, M., CHENG, J., Opinion Observer: Analyzing and Comparing Opinions on the Web, in: Proceedings of the 14th international conference on World Wide Web, ACM Press, New York 2005, S. 342-351.
- [11] POPESCU, A. M., ETZIONI, O., Extracting Product Features and Opinions from Reviews, in: A. Kao und S. R. Poteet (Hrsg.): Natural Language Processing and Text Mining, Springer Verlag, London 2007, S. 9-28.
- [12] RANDIC, M., On Characterization of Molecular Branching, in: Journal of the American Chemical Society, H. 97, 1975, S. 6609-6615.
- [13] ROGERS, E. M., Diffusion of Innovations. 5. Auflage, Free Press, New York 2003.
- [14] SCOTT, J., Social Network Analysis, a Handbook, 2. Auflage, Sage Publications, London 2007.
- [15] WASSERMANN, S.; FAUST, K.: Social Network Analysis – Methods and Applications. Cambridge University Press, Cambridge, 1999.
- [16] WEISS, S. M.; INDURKHYA, N.; ZHANG, T.; DAMERAU, F. J.: Text Mining – Predictive Methods for Analyzing Unstructured Information, Springer Verlag, New York 2005.
- [17] WELSER, H. T.; GLEAVE, E.; FISHER, D.; SMITH, M., Visualizing the Signatures of Social Roles in Online Discussion Groups, in: Journal of Social Structure, Bd. 8, 2007.