Horst Hischer

Eine überraschende Eigenschaft großer realer Netzwerke

„Kleine Welten“ – ein recherchierbares Phänomen

Das Orakel verrät: Wer hat mit wem über wie viele Ecken gedreht?

Kevin Bacon ist ein amerikanischer Film- und Fernsehschauspieler, der früher nur Nebenrollen spielte und bis Ende der 1990er-Jahre kaum bekannt war. Das änderte sich schnell, als 1996 die Website „The Oracle of Bacon“ (vgl. *www.oracleofbacon.org*) des Informatikers Brett Tjaden von der Universität Virginia im Time Magazine als eine der „Top Ten“ ausgezeichnet wurde. Worum geht es dabei?

Oracle of Bacon

Graphen und Knoten zeigen Beziehungen an

Abb. 1 zeigt einen Ausschnitt aus dem „Zusammenarbeitsgraphen“ von Filmschauspielerinnen und Filmschauspielern (hier kurz „Akteure“ genannt), die die Knoten dieses Graphen bilden: Zwischen zwei Akteuren verläuft genau dann eine Kante, wenn sie gemeinsam in einem Film mitgewirkt haben, wobei zusätzlich je ein solcher „vermittelnder“ Film eingetragen ist, der die betreffende Kante markiert (es kann zwischen jeweils zwei Knoten mehrere vermittelnde Filme geben): So hat Kevin Bacon mit Donald Sutherland „zusammengearbeitet“, weil beide (zumindest) in dem Film „Animal House“ mitgewirkt haben, und gemäß **Abb. 1** hat Donald Sutherland mit Nicole Kidman in dem Film „Cold Mountain“ zusammengearbeitet und so fort. Durch diese Kette von Ereignissen „unmittelbarer Zusammenarbeit“ ergibt sich zwischen Kevin Bacon und Kate Winslet eine „mittelbare Zusammenarbeit“, die durch „4 Schritte“ in diesem Graphen gekennzeichnet ist.

Abb. 1 zeigt aber auch einen anderen „Weg“ von Kevin Bacon zu Kate Winslet, der nur 2 Schritte umfasst, nämlich über Bill Paxton. **Abb. 2** zeigt dazu einen Ausschnitt als einen auf das Wesentliche reduzierten Untergraphen, der die vermittelnden Filme nicht mehr enthält.

Ein Untergraph U eines Graphen G besteht aus einer Teilmenge der Knotenmenge von G und allen Kanten zwischen diesen Knoten, die auch in G enthalten sind. Der „Abstand“ d(A,B) zweier Knoten A und B ist die Länge eines kürzesten Weges zwischen A und B (deren es mehrere geben kann); falls zwischen A und B kein Weg existiert, ist d(A,B) := ∞.

Eine Datenbank gibt Auskunft

Die Internet Movie Data Base, *www.imdb.com*, eine ständig aktualisierte Datenbank, dient der Erfassung aller weltweit agierenden (sowohl lebenden als auch nicht mehr lebenden) Akteure und „ihrer“ Filme. Der aus all diesen Akteuren bestehende Zusammenarbeitsgraph (englisch „collaboration graph“) sei hier „Akteurs-Graph“ genannt und mit Ca bezeichnet. Für alle Akteure A ∈ Ca ist die „Bacon-Zahl von A“ der Knotenabstand d(A, Bacon). So hat Kate Winslet bisher die Bacon-Zahl 2 (es gibt derzeit noch keinen kürzeren Weg), und Donald Sutherland hat schon jetzt die Bacon-Zahl 1. Die (zeitabhängige!) Bacon-Zahl eines Akteurs beschreibt damit dessen aktuellen „filmschauspielerischen Verwandtschaftsgrad“ zu Kevin Bacon.

Auf dieser Datenbank (s. o.) basiert das bereits erwähnte von Brett Tjaden entwickelte Web-Spiel „The Oracle of Bacon“, das bei Eingabe eines beliebigen Akteurs dessen aktuelle Bacon-Zahl liefert. Dieses „Kevin-Bacon-Orakel“ wird von Patrick Reynolds ständig aktualisiert und weiterentwickelt (vgl. piki.org/patrick). So findet man z. B. bei Heinrich George (deutscher Schauspieler, 1893 – 1946) aktuell die Baconzahl 3 (**Abb. 3**, S. 42). In der Ergebnisanzeige werden vermittelnde Filme (als Kanten) und „Zwischenakteure“ (als Knoten) angegeben.

Darüber hinaus können mittels der Taste „Find a different link“ weitere Verbindungswege derselben Länge angefordert werden. Und man kann die Default-Eingabe „Kevin Bacon“ durch einen anderen Akteurs-Namen ersetzen und damit also den sog. Zusammenarbeitsabstand („collaboration distance“) von zwei beliebigen, in der Datenbank erfassten Akteuren er-

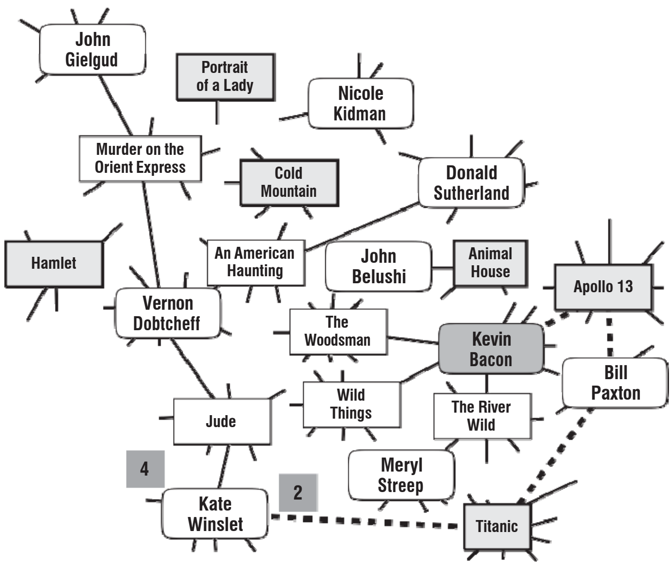

Abb. 1: Ausschnitt aus dem Zusammenarbeitsgraphen

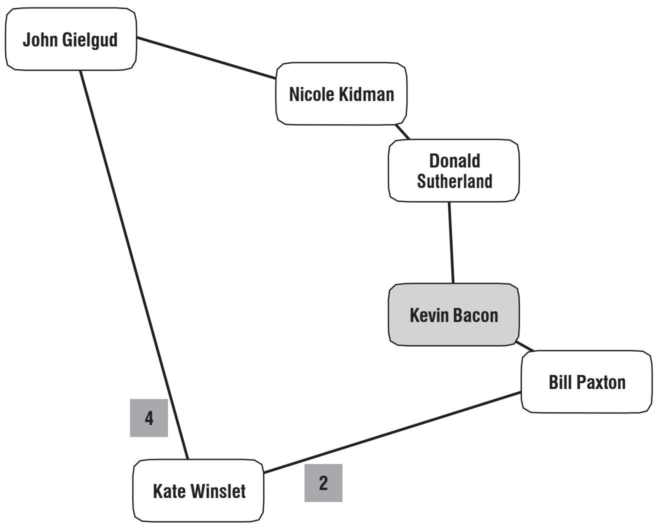

Abb. 2: Reduzierter Graph

Bacon-Zahl	1999	2009	1.9.2013	25.11.2013	14.2.2014
0	1	1	1	1	1
1	1.181	2.251	2.796	2.799	2.813
2	71.397	22.5506	311.207	313.045	324.468
3	124.975	719.767	1.059.651	1.078.865	1.120.484
4	25.665	178.784	266.847	276.680	281.275
5	1.787	12.205	21.222	22.296	21.410
6	196	1.040	2.157	2.361	2.256
7	22	165	226	251	224
8	2	17	28	24	20
mittlere Bacon-Zahl:	2,81	2,98	3,00	3,006	3,001
Datenbasis:	225.226	1.139.736	1.664.135	1.696.322	1.752.951

Tab. 1: Absolute Häufigkeiten der Bacon-Zahlen

mitteln – ein Abstand, der ggf. ∞ ist.

Es mag verwundern, dass Heinrich George von Kevin Bacon nur „3 Schritte entfernt“ ist. Allerdings: Man findet tatsächlich „meistens“ 3 als Bacon-Zahl – und das ist die eigentliche Überraschung!

Tabelle 1 zeigt die absoluten Häufigkeiten der Bacon-Zahlen zu verschiedenen Zeitpunkten. Insbesondere ist 8 die bisher größte aufgetretene endliche Bacon-Zahl, also gilt:

$d(A, \text{Bacon}) \in \{0, 1, \ldots, 8, \infty\}$ für alle $A \in Ca$. **Abb. 4** visualisiert die zeitliche Entwicklung: In den zehn Jahren von 1999 bis 2009 ist die mittlere Bacon-Zahl (gebildet aus allen endlichen Bacon-Zahlen) trotz Verfünffachung der Datenbasis kaum größer geworden, und auch in den folgenden Jahren ist sie nahezu konstant geblieben.

Übrigens: Wie kann es sein, dass von September 2013 bis Februar 2014 die absolute Häufigkeit der Bacon-Zahl 8 mehrfach gesunken ist? Eine gute Frage …

Darüber hinaus zeigt **Abb. 5** auf der Basis von **Tab. 1** die Verteilung der relativen Häufigkeiten der Bacon-Zahlen: Die Häufigkeitsverteilungen für die Jahre 2009 und 2013 sind nahezu identisch, und sie weichen selbst gegenüber der Verteilung von 1999 (punktiert dargestellt) nur geringfügig ab.

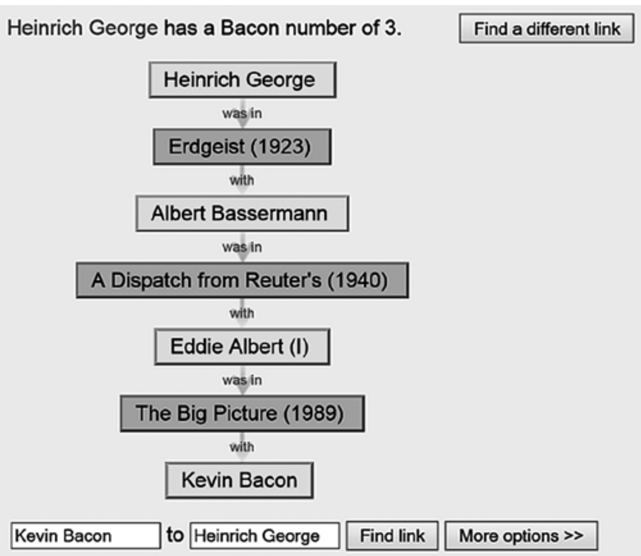

Abb. 3: Orakel-Auskunft zum deutschen Schauspieler Heinrich George

Die Erdős-Zahl

Unter Mathematikern ist schon länger die sogenannte „Erdős-Zahl“ bekannt, die ähnlich wie die Bacon-Zahl definiert ist. Sie bezieht sich auf den großen ungarischen Mathematiker Pál Erdős (1913 – 1996), und ihr liegt der (zeitabhängige) „collaboration graph“ aller (weltweit!) sowohl lebenden als auch verstorbenen, jeweils publiziert habenden Mathematiker(innen) zugrunde. Zwischen zwei Autoren verläuft genau dann eine Kante, wenn sie mindestens eine Publikation gemeinsam verfasst haben (wobei auch weitere Autoren beteiligt sein können). Die Informationen kann man der „MathSciNet“ genannten Datenbank entnehmen, vgl. www.ams.org/mathscinet/collaborationDistance.html.

Kleine Welten

Der **Tab. 1** kann man neben den mittleren Abständen zu Bacon und der jeweiligen Datenbasis (November 2013 war die mittlere Bacon-Zahl 3,01 bei einer Datenbasis von 1 696 322) auch den maximal möglichen Knotenabstand entnehmen: Wählt man nämlich in diesem Graphen – genannt „Bacon-Graph“ – zwei beliebige Schauspieler A und B, dann gibt es zwischen beiden auf jeden Fall einen Weg über Bacon, und weil gemäß **Tab. 1** sowohl $d(A, \text{Bacon}) \leq 8$ als auch $d(B, \text{Bacon}) \leq 8$ gilt, folgt $d(A, B) \leq 16$, (tatsächlich ist $d(A, B) \leq 15$). Analog zum Bacon-Graphen gibt es passend zu den Erdős-Zahlen auch einen „Erdős-Graphen“, und für den gilt $d(A,B) \leq 26$.

Der maximale Knotenabstand eines Graphen ist graphentheoretisch sein Durchmesser, der hier im Vergleich zur sehr großen Datenbasis sehr klein ist. Das hat zur Folge, dass jeweils auch der mittlere Knotenabstand „relativ klein“ ist, nämlich notwendigerweise kleiner als der jeweilige maximale Knotenabstand. Das bedeutet per saldo eine „schnelle Durchsuchbarkeit“ des Graphen. Beides zusammen – also ein kleiner Durchmesser und folglich ein kleiner mittlerer Knotenabstand – war ursprünglich kennzeichnend für „Kleine Welten“.

Bekannt über mehrere Ecken

Dieses „Kleine-Welt-Phänomen“ passt zu der Erfahrung, wenn man etwa als völlig Fremder zu einer Party oder einer anderen Versammlung stößt und nach kurzer Unterhaltung feststellt, dass man mit einem anderen Teilnehmer einen gemeinsamen Bekannten hat – was man dann etwa kommentiert mit: „Ach, wie ist die Welt doch klein!“

Die englische Bezeichnung „small-world phenomenon“ wird dem Psychologen Stanley Milgram zugeschrieben, der 1967 ein Experiment durchgeführt hat: Er schickte Hunderte von Briefen an Menschen in Nebraska und bat sie, diese Nachricht an einen persönlichen Bekannten weiterzugeben, der sie wieder an einen Bekannten weiterreichen sollte, und so weiter, bis sie schließlich bei dem namentlich genannten Empfänger, einem Börsenmakler in Boston, eintreffen würde. Die „Mitwirkenden“ an diesem Versuch kannten also lediglich

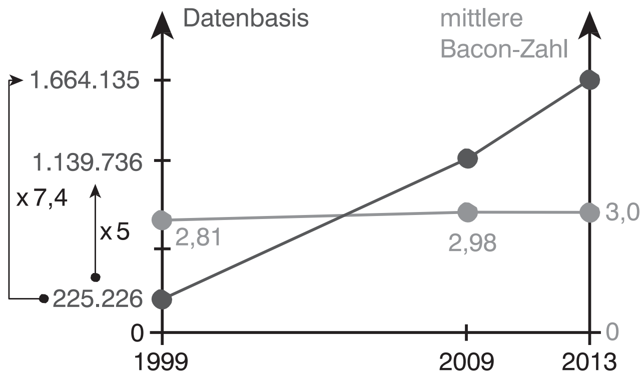

Abb. 4: Zeitliche Veränderung der mittleren Bacon-Zahl aufgrund der mit den Jahren zunehmenden Datenbasis

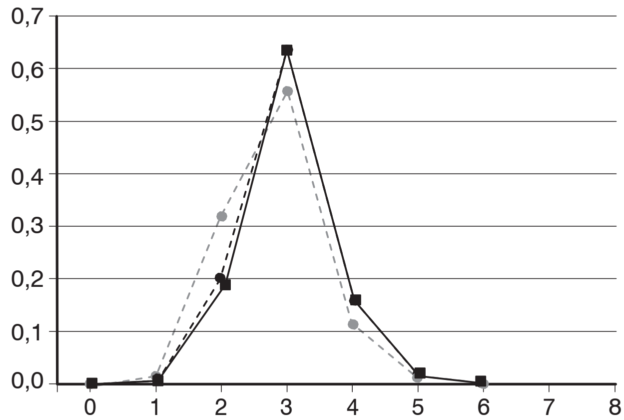

Abb. 5: Verteilung der relativen Häufigkeiten der Bacon-Zahlen 1999 (------), 2009 (------) – fast identisch mit 2013 (——)

den Namen des vorgesehenen Empfängers, nicht aber dessen Adresse. Diejenigen Briefe, die ihren Bestimmungsort erreichten, hatten im Durchschnitt sechs Zwischenstationen durchlaufen, was dann „Six Degrees of Separation“ genannt wurde und 1990 zu einem gleichnamigen Schauspiel von John Guare führte, das 1993 oscarreif verfilmt wurde. (Trailer unter www.imdb.com/video/screenplay/vi3416524057/). Die Protagonistin Ouisa sagt hier:
Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and evereybody else on this planet. The president of the United States. A gondolier in Venice ...

Weitere „Kleine Welten“

Das Kleine-Welt-Phänomen ist typisch für viele reale, große „Netzwerke“. So konnte z. B. gezeigt werden, dass zwei beliebige chemische Stoffe in einer Zelle fast immer über nur drei Reaktionen miteinander verbunden sind, und im World Wide Web mit seinen (im Stand von 2004) mehr als drei Milliarden Dokumenten liegen zwei Web-Seiten typischerweise nur 19 Klicks auseinander. Dabei muss man übrigens zwischen dem Internet als einem ungerichteten Graphen (Knoten: Server, Router, Computer; Kanten: Datenleitungen unterschiedlichen Typs) und dem World Wide Web (WWW) als einem gerichteten Graphen (Knoten: Web-Seiten; Kanten: Hyperlinks) unterscheiden, wobei auch das Internet eine „Kleine Welt“ ist. (Streng genommen kann man daher nicht im Internet surfen, sondern nur im WWW!)

Naben: Mittelpunkte in einem Graphen

Typisch für Kleine Welten ist das Vorhandensein sogenannter „Naben“. Das sind Knoten mit einem – im Vergleich zur Mehrheit der anderen Knoten – besonders hohen Knotengrad, also mit besonders vielen Kanten zu anderen Knoten. Solche Naben sind in gewissem Sinn „Mittelpunkte“ in ihrem zugehörigen Graphen, weil sie einen sehr kleinen mittleren Abstand zu den anderen Knoten aufweisen. Sie treten auch in den vorgestellten Zusammenarbeitsgraphen auf. So ist beispielsweise Kevin Bacon im Akteursgraphen eine Nabe, also ein „Mittelpunkt“, und dazu gehört die Bacon-Zahl von aktuell etwa 3,00 als mittlerer Abstand zu den anderen Akteuren. Aber Patrick Reynolds teilt auf der von ihm gepflegten Website oracleofbacon.org mit, dass er (Stand: April 2013) insgesamt 369 Akteure gefunden habe, die ein „besserer Mittelpunkt“ als Bacon seien, beispielsweise Sean Connery mit dem mittleren Abstand 2,937 zu den anderen Akteuren (so gibt es auch eine „Connery-Zahl“). All diese Akteure sind ebenfalls Naben in diesem Zusammenarbeitsgraphen.

Naben ermöglichen kurze Verbindungen in ihren Graphen, und sie sind die eigentliche Ursache für das Kleine-Welt-Phänomen. Sie sind aber für „natürliche“ große Graphen, die man auch „Netzwerke“ nennt, ein großes Problem: Während solche Netzwerke den Ausfall einzelner, stochastisch gewählter Knoten problemlos verkraften, sind sie extrem anfällig gegenüber gezielten zerstörenden Angriffen auf Naben. Andererseits kann das z. B. pharmakologisch vorteilhaft genutzt werden bei der Bekämpfung von Erregern (die als Naben auftreten).

Cliquenhaftigkeit und Kleine Welten

Kleine Welten sind nun nicht nur durch das Vorliegen eines kleinen Durchmessers (bzw. eines kleinen mittleren Knotenabstands) und durch Naben gekennzeichnet, sondern auch zusätzlich durch eine hohe Cliquenhaftigkeit, die sich in einem „möglichst hohen“ globalen Clusterkoeffizienten zeigt – in folgendem Sinne:

In vollständigen Graphen sind je zwei Knoten durch eine Kante verbunden. Ein vollständiger Untergraph eines Graphen ist eine Clique; so bilden etwa die Eckpunkte eines Tetraeders mit den Tetraederkanten eine Clique. (Anschaulich: In einer Clique ist jeder mit jedem bekannt oder befreundet.)

Der von den „Nachbarknoten“ eines Knotens P gebildete Untergraph U(P) eines Graphen G sei Cluster (Haufen) von G genannt. Ein Cluster kann ggf. eine Clique sein. C(P), der lokale Clusterkoeffizient von P, ist ein numerisches Maß für die Cliquenhaftigkeit von P mit $0 \leq C(P) \leq 1$. Genau dann ist $C(P) = 1$, wenn das Cluster U(P) vollständig ist, und genau dann ist $C(P) = 0$, wenn U(P) keine Kanten enthält. Der globale Clusterkoeffizient von G ist das arithmetische Mittel aller seiner lokalen Clusterkoeffizienten.

Fazit

Der Einsatz des Computers steht hier nicht im Vordergrund, obwohl mit Hilfe eines Browsers wichtige Primärerfahrungen über das „Oracle of Bacon“ zu gewinnen sind, die dann zu einem Staunen führen sollten und zu analysieren sind. Insbesondere ist es ratsam, über einen längeren Zeitraum Daten gemäß **Tab. 1** zu erfassen und diese graphisch auf geeignetem Wege zu strukturieren. Ziel sollte es sein, das Kleine-Welt-Phänomen auf diese Weise exemplarisch kennenzulernen und dabei die Bedeutung von Naben für große Netzwerke wie z. B. das Internet zu erfahren. Denn nicht der „Computereinsatz“ ist wichtig, sondern „Bildung ist das Paradies“!

Literatur

Hischer, H. (2014): Kleine Welten und Netzwerke und ihr mögliches Potential für Didaktik, Unterricht und Pädagogik. Fachrichtung Mathematik, Universität des Saarlandes, Preprint 342, www.math.uni-sb.de/service/preprints/.

Hischer, H. (2012): Mathematikunterricht und Neue Medien – oder: Bildung ist das Paradies! Fachrichtung Mathematik, Universität des Saarlandes, Preprint 67, www.math.uni-sb.de/service/preprints/.

Links geprüft am 7. 4. 2015