Terroristensuche in Telefonnetzen?

Supercomputer durchforsten täglich Milliarden von Anrufen und Internetnachrichten. Kombiniert mit Graphentheorie und Bekanntschaftsanalysen könnten, so glaubt unser Autor, auch konspirative Gruppen aufgespürt werden.

Von Brian Hayes

»Was lässt sich aus Billionen von Telefonaten und E-Mails herausfiltern?«

Seit jenem schrecklichen Dienstagmorgen vor über fünf Jahren, als entführte Flugzeuge in das World Trade Center und das Pentagon stürzten, herrscht in Amerika eine unterschwellige Stimmung aus Sorge und Misstrauen. Es gibt natürlich Angst vor weiteren Anschlägen. Es gibt aber auch Bedenken, dass Maßnahmen, die solchen Attacken vorbeugen, individuelle Rechte und Freiheiten einschränken könnten. Im letzten Jahr gab es eine kontroverse Diskussion, nachdem bekannt geworden war, dass US-Behörden Internetverbindungen, Telefongespräche und finanzielle Transaktionen überwachen. Einige der Überwachungsprogramme sollen angeblich permanent gigantische Datenmengen durchforsten und nach Mustern suchen, die den Verdacht auf kriminelle Pläne oder Aktivitäten nahelegen.

Die Debatte über diese Programme kreist vorwiegend um rechtliche und politische Fragen. Werden verfassungsrechtliche Schutzbestimmungen hinreichend berücksichtigt? Wie sieht es aus mit Gesetzen, die Geheimdiensten verbieten, Bürger auszuspionieren? Gelingt den Überwachungsprogrammen der Balanceakt zwischen dem Recht auf Privatsphäre einerseits und dem Sicherheitsbedürfnis der Allgemeinheit andererseits? Das sind wichtige Themen, aber ich möchte sie anderen überlassen. Hier möchte ich folgender Frage nachgehen: Was kann man aus solchen umfassenden Überwachungs- und Datenauswertungsprogrammen lernen? Besitzen die Kommunikationsmuster der Terroristen so spezifische Merkmale, dass Überwachungsprogramme die konspirativen Verbindungen aus Billionen von Telefonaten oder E-Mail-Nachrichten ausfiltern können?

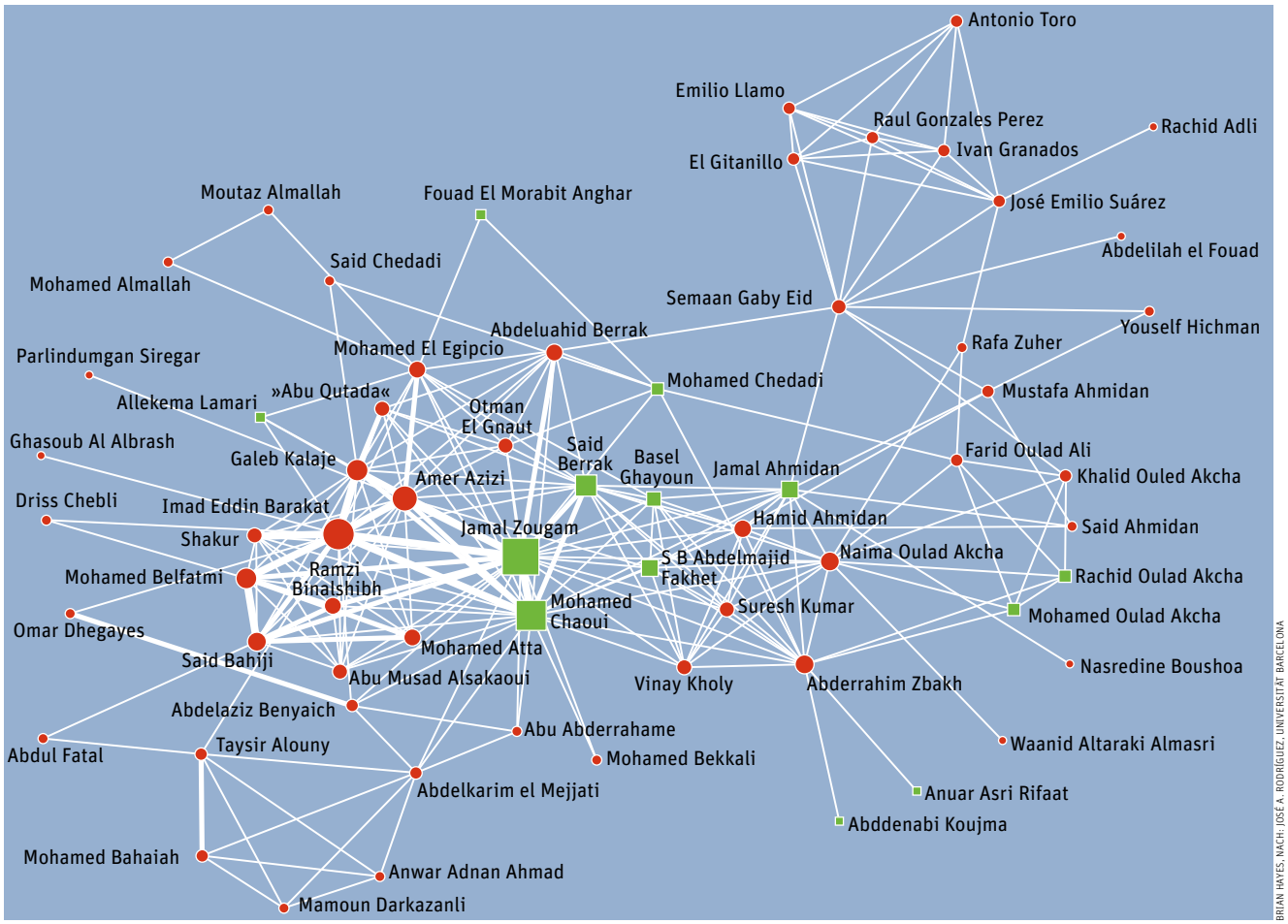
Die Beantwortung dieser Fragen stößt offenbar auf Schwierigkeiten. Bislang wurden nur sehr wenig verlässliche Informationen über Eigenschaften und Umfang solcher Software öffentlich bekannt. Mathematiker und Informatiker haben sich allerdings mit Problemen beschäftigt, die denen eines Geheimagenten ähneln, der Überwachungsdaten auswerten soll. Auch Sozialwissenschaftler interessieren sich schon seit Langem für die sozialen Netzwerke, die Personen miteinander verbinden. Vielleicht hilft ja die Kombination beider Disziplinen, um einige plausible Vermutungen anzustellen.

Die im letzten Jahr bekannt gewordenen Überwachungsprogramme scheinen einige sehr spezielle Aufgaben lösen zu können. Dazu zählen Lauschangriffe, bei denen jemand beispielsweise Telefongespräche abhört oder Inhalte von Internet- und E-Mail-Nachrichten aufzeichnet. Eine »Folge-dem-Geld«-Software sammelt Informationen aus Bankgeschäften. Am faszinierendsten aber finde ich Berichte über ein Projekt, bei dem eine Datenbank von Telefongesprächen analysiert werden soll, um Verbindungen zwischen Verschwörern aufzudecken.

Telefonate im Gesprächsgraphen

Diese Datenbank enthält keine Gesprächsaufzeichnungen oder sonstigen Informationen, welche die Inhalte der Gespräche verraten; sie enthält lediglich die Verbindungsdaten: die Telefonnummern der beiden jeweiligen Gesprächsteilnehmer sowie Datum, Uhrzeit und Gesprächsdauer.

Diese Gesprächsdatenbank kam mir gleich bekannt vor. Vor einigen Jahren hatte ich von Experimenten mit einem ähnlichen Datenspeicher gelesen – fast sicher einer früheren Version von der, die jetzt angeblich die Regierung testet. Die numerischen Experimente waren Algorithmentests im mathematischen Gebiet der Graphentheorie zur Analyse netzwerkartiger Strukturen. Die Telefongesprächs-



BRIAN HAYES, NACH: JOSÉ A. RODRÍGUEZ, UNIVERSITÄT BARCELONA

datenbank stellte dafür ein nützliches Testobjekt dar, da sie als riesiger mathematischer Graph betrachtet werden kann.

Erste Erwähnungen dieser Datenbank, auch »Gesprächsgraph« genannt, finden sich in den Medien. Danach hätten »Techniker (der National Security Agency NSA) bestimmte Gespräche belauscht und große Mengen von Verbindungsdaten aus Telefonaten und Internetverkehr nach Mustern durchsucht, mit denen sich Terrorismusverdächtige aufspüren lassen«.

Die NSA ist der amerikanische Spionagedienst und zuständig für Kryptografie-Belange und Signalüberwachung. Obwohl sein Budget und seine personelle Besetzung geheim sind, wird oft behauptet, er sei der größte amerikanische Überwachungsdienst und zugleich der größte Arbeitgeber für Mathematiker, womöglich weltweit.

Die Untersuchung des Graphen obliegt einem Zweig des Bereichs Signalüberwachung, die für »Verkehrsanalyse« bekannt ist. In einem Krieg kann die Situation eintreten, dass man Funksignale des Feindes auffängt, aber sich der verschlüsselte Inhalt nicht dechiffrieren lässt. Doch dabei kann es bereits erhellend sein, die Anzahl der Nachrichten zu zählen. Ein Schwall von Nachrichten signalisiert vielleicht eine kurz bevorstehende Truppenbewegung. Ein plötzlicher Nachrichtenstopp wirkt sogar noch bedrohlicher. Wenn es gelingt, die Quelle und den Empfänger jeder Nachricht zu ermitteln – wobei de facto ein Gesprächsgraph erstellt wird –, lässt sich sogar noch mehr herauskriegen. Denn auch die Kommunikationswege liefern oft Hinweise auf die Organisation der militärischen Einheiten.

Die Suche nach verräterischen Mustern in Telefongesprächen könnte sich ähnliche Prinzipien zu Nutze machen, ist aber weitaus schwieriger. In Kriegslagen sind Nachrichten zwischen feindlichen Einheiten leicht als solche identifizierbar. In einer Gesprächsdatenbank indes können Gespräche zwischen ein paar Dutzend Verschwörern leicht im Meer der anderen Gespräche untergehen.

Einträge in der Gesprächsdatenbank dienen nicht der Verbesserung der nationalen Sicherheit, sondern schlicht kommerziellen Zwecken. Damit der Anschlussinhaber am Ende jedes Monats eine detaillierte Abrechnung erhalten kann, muss eine Telefongesellschaft Daten für jedes zu Stande gekommene ▷

Diese Netzwerkkarte veranschaulicht die Verbindungen zwischen Personen, die am 11. März 2004 an der Bombardierung der Pendlerzüge in Madrid beteiligt waren, und ihren Helfershelfern. Grüne Quadrate kennzeichnen die eigentlichen Bombenleger. Rote Kreise stehen für die Mitverschwörer. Weiße Linien zwischen den Knoten verbinden einerseits Personen, die miteinander verwandt sind, oder andererseits Personen, die oft einen Laden aufsuchten, der einem der Verschwörer gehörte.

▷ Telefongespräch erfassen – die Nummer des Anrufers, des Angerufenen sowie die Zeiten von Gesprächsanfang und seinem Ende. Die größten Telefongesellschaften haben pro Tag ein Gesprächsaufkommen von rund 250 Millionen gebührenpflichtigen Telefonaten. Daher fallen dort monatlich mehrere Milliarden Gesprächsdaten an. Die US-Telefongesellschaft AT&T berichtet, dass ihre Datenbank gespeicherter Gesprächsinformationen über zwei Billionen Datensätze enthält; das entspricht mehr als 300 Terabyte an Daten.

Der Gesprächsgraph dient außer zur Rechnungserstellung auch noch anderen firmeninternen Zwecken – die sich teils nicht sehr von denen der NSA unterscheiden und fast genauso geheim gehalten werden. Mit älteren Gesprächsdaten lassen sich Betrügereien aufdecken, andere Muster sind für Marketingaktionen von Interesse. Offeriert eine Firma etwa Vergünstigungen bei Gesprächen für eine bestimmte Zielgruppe, kann die Grapheninformation als Grundlage einer Kosten-Nutzen-Analyse dienen.

Im Prinzip lassen sich Verbindungsdaten wie die der Telefonate auch für andere Kommunikationsdienste erstellen. So speichern etwa Federal Express und andere Kurierdienste ihre Auftragsdaten digital. Daraus ließen sich leicht Datenbänke von Sendern und Empfängern erzeugen. Kurioserweise bietet das digitale Medium schlechthin – das Internet – keine Möglichkeit, Daten, wer mit wem kommuniziert, routinemäßig festzuhalten. Es gibt dafür keinen direkten Bedarf, weil Kunden für E-Mails eben nichts bezahlen müssen. Doch sehe ich keine prinzipiellen Hürden, um detaillierte Statistiken zu E-Mails oder anderen Arten des Internetverkehrs zu erstellen. Ein »Paket-Schnüffler«, der am Zentralrechner des Netzwerks installiert wird, müsste dazu lediglich die Kopfzeilen der Nachrichten sowie die »an«- und »von«-Adresseinträge speichern. (Es ist sogar denkbar, dass Geräte, welche die NSA in Schaltzentralen des Internets installiert hat, genau diesem Zweck dienen.)

Sich in Gesprächsgraphen zu vertiefen gerät zur Zahlenfresserei. Will man einen Berg von hunderten Terabyte von unzusammengehörigen Daten durcharbeiten, so ist dies das digitale Äquivalent zu einer Großbaustelle. Bevor sich Baggerschaufeln in Bewegung setzen, sollten wir uns klarmachen, wonach wir eigentlich suchen. Welche Kommunikationsmuster sind typisch für Terroristen und ihre Helfershelfer?

Gute Ansprechpartner für solche Fragen sind Wissenschaftler, die soziale Netzwerke studieren – also Strukturen von Gruppen, die durch ihre Verbindungen untereinander charakterisiert sind. Natürlich bildet auch die Gruppe der Sozialen-Netzwerk-Forscher selbst ein soziales Netzwerk.

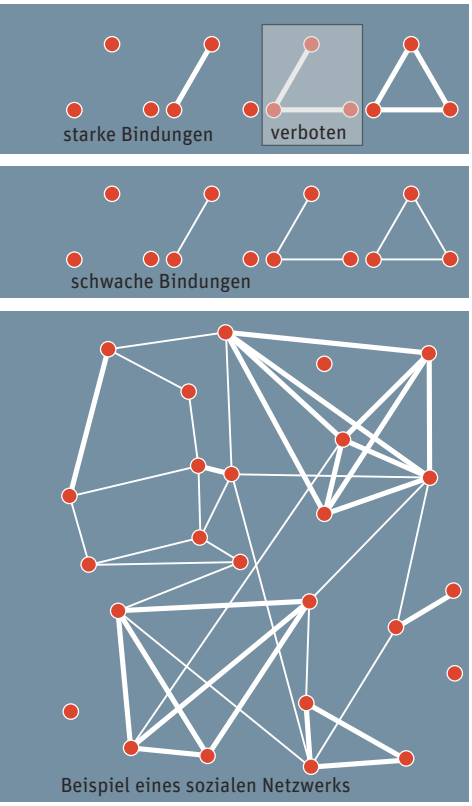
Einen grundlegenden Artikel zu diesem Thema veröffentlichte 1973 der heute an der Stanford-Universität tätige Mark S. Granovetter unter dem Titel »Die Stärke schwacher Verbindungen«. Ihm war aufgefallen, dass bei Leuten, die eng miteinander vernetzt sind – etwa nahe Freunde, Familienmitglieder oder Arbeitskollegen –, die Stärke der Verbindungen meist symmetrisch ist und dass sie einer Regel gehorchen, die man Dreiseitigkeit nennen könnte. Symmetrie bedeutet: Wenn A ein Freund von B ist, dann ist B auch ein Freund von A. Dreiseitigkeit besagt: Wenn A mit B und C befreundet ist, sind vermutlich auch B und C miteinander befreundet. Dabei handelt es sich natürlich nur um Wahrscheinlichkeiten, und jedem fallen sofort Gegenbeispiele ein: unerwiderte Liebe, gestörte Dreiecksbeziehungen. Doch für Analysezwecke ist es sinnvoll zu fragen, wie eine Gesellschaft aussehen würde, in der Symmetrie und Dreiseitigkeit streng gültig wären. Die Antwort lautet, dass die Sozialstruktur dann ausschließlich aus perfekten Cliquen bestünde – Gruppen, in denen jeder mit jedem verbunden ist.

Was die Welt zusammenhält

Starke Bindungen zwischen Individuen liefern den Klebstoff für sozialen Zusammenhalt. Doch nach Granovetters Theorie gibt es einen paradoxen Effekt. Lokal erzeugen enge Verbindungen sehr stabile Strukturen; weiträumiger betrachtet lassen sie jedoch auch voneinander isolierte Gruppen entstehen. Wegen der Alles-oder-nichts-Natur enger Beziehungen werden die Gruppen zu Inseln, die kaum miteinander kommunizieren. Was indes die Welt wirklich zusammenhält, meint Granovetter, sind die schwachen Bindungen im erweiterten Bekanntenkreis. Solche Freundschaften sind häufig symmetrisch, aber selten »dreieckig«. Man kann mit dem Kundenbetreuer bei der Bank jede Woche ein Gespräch führen, ohne alle dessen anderen Bankkunden kennen zu lernen. Solche schwachen Verbindungen, die auf den ersten Blick sozial nicht bedeutsam erscheinen, schaffen jedoch zwischen den Cliquen wichtige Querverbindungen. Die Sozialstruktur einer Gesellschaft besteht laut Granovetter aus Clustern von Personen, die intern miteinander stark verbunden sind und extern mit anderen Clustern lose.

Die Theorie der sozialen Netzwerke besitzt deutliche Parallelen zur mathematischen Graphentheorie – obwohl Menschen, die in diesen beiden Fachgebieten arbeiten, Cluster bilden, die nur schwach miteinander verbunden

Die Dreiseitigkeitsregel von Mark Granovetter besagt, dass Leute, die einen gemeinsamen Freund haben, wahrscheinlich auch selbst miteinander befreundet sind. Granovetters Modell erhebt dieses Prinzip zu einer strikten Regel – aber nur für den Fall starker sozialer Bindungen. Leute, die stark miteinander verbunden sind, bilden oft Cliquen – also Subnetzwerke, in denen jeder mit jedem verbunden ist.



sind. Graphentheorie besitzt ihre eigene Nomenklatur und beschreibt ihre Studienobjekte reichlich abstrakt. Formal betrachtet besteht ein Graph aus einer Menge aus Knoten und Kanten, wobei jede Kante zwei Knoten miteinander verbindet. Diese Definition kann auf verschiedene Weise interpretiert werden. In der Praxis jedoch lässt sich feststellen, dass sowohl Graphentheoretiker als auch Netzwerktheoretiker mit Diagrammen aus Punkten und Linien arbeiten.

Was verraten uns die Prinzipien der Netzwerktheorie und der Graphentheorie über die Struktur von Terroristenzellen? Schon das Wort »Zelle« gibt einen Anhaltspunkt – es suggeriert Unterteilung. Und in der Tat ist die Welt der Spionageringe und Terroristen für ihre limitierte Kommunikation bekannt. Denn alles andere würde gefangen genommene Gruppenmitglieder nur in Gefahr bringen. Gleichzeitig jedoch müssen sie untereinander in Kontakt bleiben, um Pläne schmieden und ausführen zu können.

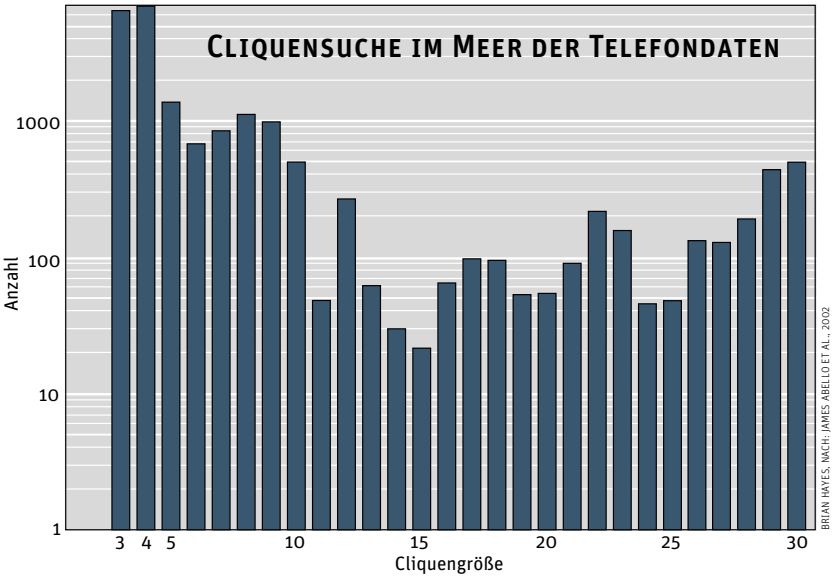
Eine erhellende Fallstudie stammt aus einem völlig anderen Bereich: Preisabsprachen bei US-Herstellern von Elektrogeräten in den 1950er Jahren. Das soziale Netzwerk heimlich sich absprechender Manager und Firmenchefs untersuchten Wayne E. Baker von der Universität von Chicago und Robert R. Faulkner von der Universität von Massachusetts. Sie fanden, dass »die Struktur der illegalen Netzwerke vor allem durch die Notwendigkeit maximaler Geheimhaltung geprägt ist, weniger durch Maximierung der Effizienz«. Freilich konnten die Preis- und Angebotsabsprachen nicht ohne Kommunikation zwischen den Beteiligten in die Wege geleitet werden – besonders bei großen Geräten. Trotz aller Risiken mussten sich die Chefs gelegentlich auch persönlich treffen, um ihre Pläne zu koordinieren.

Terroristennetzwerke unterliegen offenbar den gleichen paradoxen Zwängen. Valdis E. Krebs, der Netzwerkanalysen für die Lösung wirtschaftlicher Probleme nutzt, hat mit der gleichen Methode die Verbindungen zwischen den Flugzeugentführern vom 11. September 2001 untersucht. Eine Studie, die er wenige Wochen nach den Attacken durchführte, ergab ein überraschend »dünnes« Terroristen-Netzwerk. Zwar war jeder Entführer mit jedem anderen über einige Wege vernetzt, doch viele dieser Wege waren sehr lang und liefen über drei oder vier Mittelsmänner. Diese dünne Struktur machte die Kommunikation sehr ineffizient (siehe SdW 11/2002, S. 88).

Später, als mehr Informationen zur Verfügung standen, revidierte Krebs seine Analyse. Er hat die neue »Verbindungslandkarte« auch auf seiner Webseite http://orgnet.com/prevent.html veröffentlicht – und kommt nun zu anderen Schlussfolgerungen. Krebs wies nach, dass man, ausgehend von zwei Männern, die bereits im Januar 2000 unter Terrorverdacht standen, über bekannte Verbindung zu allen 19 Entführern gelangt – und außerdem zu weiteren Verschwörern. Jeder Knoten des Netzwerks ist mit den beiden Anfangsverdächtigen entweder direkt oder durch einen einzigen Mittelsmann verbunden.

José A. Rodríguez von der Universität Barcelona hat für die Terroristen, die am 11. März 2004 in Madrid Bombenattentate auf Pendlerzüge verübten, eine ähnliche Netzwerkkarte angelegt. Rodríguez deckte zwischen den Verschwörern mehrere Varianten starker Verbindungen auf. Einige waren miteinander verwandt oder schon seit ihrer Kindheit befreundet; einige stammten aus dem Umfeld eines Ladens, der einem der beiden Täter gehörte; einige waren bereits an früheren Anschlägen oder Terrorattacken beteiligt. Rodríguez untersuchte anfangs nur die 13 Männer, welche die Sprengsätze selbst gelegt und gezündet hatten, und ging dabei starken Verbindungen nach, was ein etwas seltsames Netzwerk erzeugte. Ein harter Kern aus sechs Leuten firmierte als Clique: Jeder war mit jedem anderen verbunden. Die übrigen Mitglieder waren mit der Hauptgruppe nur lose verbunden oder standen mit ihr gar nicht in Verbindung.

Das Bild änderte sich völlig, als Rodríguez rund 70 in unterschiedlicher Weise mit den Attentaten assoziierte Personen untersuchte und sowohl schwache als auch starke Verbindungen berücksichtigte. Die schwachen Verbindungen schlossen Leute ein, die an Finanztransaktionen partizipierten, den Tätern nur flüchtig begegneten oder Ähnliches. Das grö- ▷



Suche nach Gruppen in einer riesigen Datenbank aus Telefonverbindungen: Eine Clique besteht hier aus einem Datensatz von Telefonnummern. Dabei war während eines bestimmten Zeitraums jede Nummer mindestens einmal mit jeder anderen verbunden. Die Grafik zeigt Cliquen mit 3 bis 30 Mitgliedern, wie sie in 170 Millionen Telefonverbindungen eines einzigen Tages entdeckt wurden.

»Geheimdienste hätten die Attentäter des 11. September 2001 aufspüren können«

▷ ßere und vollständigere Netzwerk entspricht sehr viel stärker dem, was Granovetters Theorie erwarten lässt. Es gibt mehrere dichte Cluster, in denen die meisten Knoten stark miteinander verbunden sind; doch diese Cluster kommunizieren untereinander über relativ lose und unzuverlässige Verbindungen. So bilden etwa die Spanier, welche die Attentäter mit den Explosivstoffen versorgt hatten, einen eigenen Cluster; die meisten Verbindungen zwischen dieser Untergruppe zum Rest des Netzwerks laufen durch einen einfachen Knoten – eine ziemlich wackelige Verbindung.

Die oben beschriebenen sozialen Netzwerke wurden hinterher rekonstruiert. Ausgangspunkt waren komplette Listen der bekannten Gruppenmitglieder und ihrer biografischen Daten. Solche Strukturen bereits im Vorfeld aufzudecken – in einer Zeitspanne, wenn die Attacken sich noch in der Planung befinden und die meisten Attentäter unbekannt sind – wäre ungleich schwieriger gewesen, insbesondere wenn man mit nicht personenspezifischen Daten wie Telefon- oder E-Mail-Listen arbeitet.

Die Enthüllung eines Attentatsplans im Vorfeld – durch Datenanalyse

Solche Netzwerke auszukundschaften, das fällt klar in den Kompetenzbereich der NSA – was deren Interesse an Gesprächsgraphen erklären könnte. Ein Szenario ist leicht vorstellbar: Eine Person gerät durch bestimmte Informationen unter Verdacht und die NSA erfährt durch Auswertung ihrer Gesprächsgraphen, mit wem diese Person in den letzten Wochen und Monaten telefoniert hat. Das Ergebnis wäre ein Ring aus Kontakten, der den Verdächtigen umgibt. Dann wird jeder dieser Kontakte in gleicher Weise untersucht, was einen zweiten Ring von Kontaktpersonen zweiter Ordnung ergibt. Dies ließe sich noch weiter fortführen, aber wegen des exponentiellen Wachstums des Graphen würde der Ring dann schnell den größten Teil der Bevölkerung umfassen (insbesondere wenn die Person Anrufe von Telefonwerbern erhielt oder auch mal eine Pizza bestellte). Interessanter sind Kontaktpersonen, die ebenfalls untereinander in Verbindung stehen, da solche Dreiseitigkeit die stärksten Verbindungen kennzeichnet.

Der schwierigste Part der Netzwerkanalyse besteht nicht darin, diese Verbindungen aufzuspüren, sondern festzustellen, welche davon bedeutsam sind. Es ist denkbar, dass Geheimdienste – ausgehend von den beiden seit Januar 2000 bekannten Verdächtigen – Verbindungen zu allen Attentätern vom 11. September 2001 hätten aufspüren können, wenn sie ihr Wissen systematisch genutzt hätten. Doch tausende andere Personen wären dann ebenfalls unter Verdacht geraten. Es bringt also wenig, das Netzwerk der Verschwörer isoliert zu betrachten; der Graph ist de facto in eine viele größere Struktur eingebettet.

Will man die Verbindungen zwischen bekannten Verdächtigen nachverfolgen, so wäre der direkte Zugriff auf deren Gesprächsgraphen hilfreich, aber er ist nicht notwendig. Bei namentlich bekannten Personen könnte die Herausgabe der Gesprächsdaten auf gesetzlichem Wege erwirkt werden, wie es bei Strafverfolgungsbehörden gängige Praxis ist. Würde der Gesprächsgraph mit seinen 10^{12} Datensätzen nur daraufhin durchforstet, die paar hundert oder paar tausend relevanten Gespräche herauszusuchen, erschiene der Aufwand in der Tat unangemessen hoch.

Aktuelle Berichte in den Medien lassen jedoch vermuten, dass Gesprächsgraphen weit interessantere Informationen liefern können: Damit lassen sich nicht nur die Kontaktpersonen bekannter Terroristen aufspüren, sondern auch ein konspirativer Attentatsplan im Vorfeld enthüllen – einfach indem die Datenbestände nach Mustern durchsucht werden, die auf Verdächtige hindeuten. Das hört sich vielleicht wie Magie an: Man wirft einen Blick auf das riesige und komplizierte Gebilde des Gesprächsgraphen und sieht – ohne auch nur die Namen der Anschlussinhaber zu kennen – einige Verbindungsmuster, die als Gefahrensignale gewertet werden können. Lässt sich dieser Trick von der Welt der Magie in die Welt der Algorithmen übertragen?

Wenn solche typischen Muster existieren, legen die Erkenntnisse der Netzwerktheorie nahe, dass sie bestimmte Kombinationen aus starken und schwachen Verbindungen enthalten. Eine Terroristenzelle kann auf Grund spezieller Gesprächsmuster als eine Gruppe von Leuten gekennzeichnet werden, die viel miteinander telefonieren, aber dem Rest der Welt wenig mitzuteilen haben. Daher ist das Alarm auslösende Muster ein dichter Subgraph, der relativ isoliert von seiner Umgebung auftritt.

Nur die NSA selbst weiß, ob es möglich ist, solche Muster tatsächlich auszumachen. Doch man kann stellvertretend ein etwas simpleres Problem betrachten, um die Schwierigkeit der Aufgabe abzuschätzen. Dieses simplere Problem besteht darin, eine große Clique innerhalb eines Gesprächsgraphen aufzuspüren. An der Lösung versuchte sich Ende der 1990er Jahre ein Team um James Abello, der damals an den AT&T Bell Laboratories tätig war.

Die größte Clique in einem Graph zu finden ist ein typisches »schweres Problem«. Nach der Brute-force-Methode (gemeint ist

hier rohe Rechenkraft) untersucht man dazu einfach jede mögliche Untereinheit von Knoten und prüft, ob sie jeweils alle miteinander verbunden sind. Die Zahl von Untereinheiten wächst dabei so schnell, dass der Algorithmus schon bei einem Graphen aus 50 Knoten stecken bleibt. Bei 50 Millionen Knoten beispielsweise wäre sein Einsatz schlicht unmöglich. Die einzig praktikable Alternative sind Annäherungen und probabilistische Methoden, die meist eine gute Näherung finden, aber nicht unbedingt garantieren können, dass sie auch die optimale Lösung darstellen.

14 000 Cliquen mit 30 Mitgliedern

Der in Abellos Experiment benutzte Graph umfasst die Einträge eines einzigen Tages. Er hatte 53 767 087 Knoten (entspricht der Zahl der Telefonnummern) und mehr als 170 Millionen Kanten (entspricht der Zahl der Anrufe). Der Algorithmus begann mit einer kleinen Clique und versuchte, daraus eine größere zu konstruieren. In der Anfangsphase suchte das Programm wiederholt nach neuen Knoten, die mit allen bereits gefundenen verbunden waren. Wenn keine neuen Knoten, die diese Bedingung erfüllen, mehr entdeckt wurden, wechselte das Programm die Suchstrategie. Nun versuchte es, einen Knoten zu entfernen und dafür zwei andere neu dazuzubekommen. Um den kompletten Datensatz eines einzigen Tages zu durchsuchen, benötigte das Programm, das auf einem Rechner mit vier Prozessoren und vier Gigabyte Arbeitsspeicher lief, rund fünf Stunden.

Die größte gefundene Clique besaß 30 Knoten. Überlegen Sie mal, was das bedeutet: Eine Gruppe von 30 Telefonanschlüssen war dadurch gekennzeichnet, dass im Lauf eines einzigen Tages von jedem Anschluss aus alle anderen 29 Nummern angerufen wurden oder dieser von den anderen, was sich zu mindestens 435 Anrufen summiert – ein äußerst eifriger Telefonzirkel! Aber es gab nicht nur eine solche Clique. Abellos Team spürte über 14 000 Cliquen aus 30 Knoten aus.

Das erstaunliche Ergebnis dieses Experiments erlaubt mehrere vorläufige Schlüsse. Erstens steht die Rechenkraft zum Analysieren eines Gesprächsgraphen jederzeit zur Verfügung – wenn auch nicht gerade auf jedem Aldi-Tischrechner. Abello konnte bei seinen Analysen den Datensatz eines ganzen Tages durchforsten; heutige Computer könnten sicherlich noch größere Datenmengen bewältigen.

Zweitens bereitet es offenbar kein Problem, Beispiele eines vorgegebenen Musters im Gesprächsgraphen aufzuspüren. Das Problem besteht darin, ein Muster zu definieren, das selektiv genug ist, um eine Zielgruppe zu entdecken, ohne gleichzeitig 14 000 andere ins Visier zu nehmen. Der Algorithmus muss auf irgendeine Weise einige Dutzend potenzielle Terroristen von anderen Gruppen ähnlicher Größe und Struktur unterscheiden, die lediglich ein Familientreffen organisieren, die Nachbarschaft nach einer entlaufenen Katze absuchen, eine Ratsversammlung einberufen oder um die Wette telefonieren, um Gratis-Konzerttickets von einem Radiosender zu erhalten.

Egal mit welchem Verfahren dieses Problem angegangen wird – die Geheimdienste stehen vor einer gewaltigen Aufgabe: Sie müssen eine riesige Population (bis zur gesamten, über sechs Milliarden Personen umfassenden Menschheit) nach einer winzigen Untergruppe (die Gewalttaten plant) durchforsten.

Der Mathematiker James Abello merkt aber an, dass aus den Graphen auch noch andere Informationen extrahiert werden können. So existieren einige Cliquen für mehrere Tage, andere verschwinden rasch wieder. Informationen dieser Art können helfen, die Gruppen gegeneinander abzugrenzen. Abello zitiert auch neue Studien, die sich zum Ziel gesetzt haben, selbst organisierte Gemeinschaften in anderen Umfeldern zu identifizieren: von den Mitgliedern von Chat-Gruppen bis hin zu E-Bay-Kunden.

Es liegt in der Natur von Geheimdienstprogrammen, dass sie geheim sind. Dennoch kann das, was die Regierung im Geheimen möglicherweise tut, nicht allein ihr überlassen werden. Meine eigene Meinung habe ich mir bislang noch nicht abschließend gebildet. Terroristen mittels eines Gesprächsgraphen aufzuspüren erscheint mir als sehr schwierige Aufgabe. Aber das bedeutet ja nicht, dass sie nicht gelöst werden kann!

Gesprächsgraphen könnten nicht nur versteckte Terrorzellen aufspüren, sondern auch bei anderen Fahndungsaufgaben hilfreich sein. Hier ein Beispiel: Die Bush-Regierung hat ihren Unwillen darüber bekundet, dass die Öffentlichkeit von allen neuen Überwachungsprogrammen Kenntnis erhalten hat, und wüsste gern, über welches Schlupfloch dieses Wissen nach außen gelangte. Der Gesprächsgraph könnte das ideale Werkzeug zur Beantwortung auch solcher Fragen sein. Man muss zum einen nur alle Personen auflisten, die Zugang zu diesen Informationen hatten, und zum anderen die Journalisten, die darüber berichteten. Dann muss man diesen Graphen nach direkten und indirekten Verbindungen zwischen den beiden Knotensätzen durchsuchen. Das Kuriose ist, dass derjenige, der die Information preisgab, sicherlich sehr genau wusste, wie weit er sich damit aus dem Fenster lehnte. ◁

Brian Hayes ist Mathematiker und Redakteur von American Scientist.

AUTOR UND LITERATURHINWEISE

Graph mining: laws, generators, and algorithms. Von D. Chakrabarti und C. Faloutsos in: ACM Computing Surveys, Bd. 38, S. 1, 2006

The March 11th terrorist network. Von J. A. Rodríguez in: EPP-LEA, Nr. 3 (Vorabdruck), 2005

Massive quasi-clique detection. Von J. Abello et al. in: LATIN 2002. Lecture Notes in Computer Science, Bd. 2286, S. 598, Berlin, Springer-Verlag 2002

Mapping networks of terrorist cells. Von V. E. Krebs in: Connections, Bd. 24(3), S. 43, 2001

Weblinks finden Sie unter www.spektrum.de/artikel/862186.