Diskussion

Eine Anmerkung zu RATTINGERs Algorithmus zur Ermittlung von Cliquen*

LINTON C. FREEMAN & DAVID L. LIBHART*
University of California, Irvine

RATTINGER (1973) führte einen schnellen Algorithmus und ein entsprechendes FORTRAN-Programm zur Ermittlung von Cliquen in einer Relationenstruktur ein. Die vorliegende Anmerkung weist auf einen Fehler in RATTINGERs Verfahren hin, der es dem Algorithmus unter bestimmten Umständen unmöglich macht, alle Cliquen zu finden. Es wird eine Änderung im Algorithmus vorgeschlagen, die den aufgezeigten Mangel behebt. Diese Modifikation verlängert die Rechenzeit nur sehr wenig.

RATTINGER (1973) introduced a short algorithm and a related FORTRAN program to detect cliques in relational structures. This comment points out a mistake in RATTINGER's procedure which under certain circumstances makes it impossible to detect all cliques. An alteration of the algorithm is suggested to correct this. The modification increases computation time only slightly.

Der Begriff «Clique» bezieht sich auf die Tendenz von Menschen, sich in kleinen Gruppen zu sammeln, die durch Zusammengehörigkeitsgefühl, gegenseitigen Beistand und gemeinsame Normen und Werte gekennzeichnet sind. Es wird angenommen, daß die Existenz solcher Gruppen ein wichtiger Bestandteil der Struktur menschlicher Interaktionen ist. Durch die Betrachtung gerade dieser Gruppen hoffen wir, wichtige Aspekte individuellen und kollektiven Verhaltens besser zu verstehen.

«Clique» ist jedoch nicht nur ein intuitiver Begriff; die Clique wurde formal innerhalb der Graphentheorie von LUCE & PERRY (1949) definiert. Diese formale Definition war wichtig, denn sie erlaubte uns, die rein intuitive Spekulation über die Auswirkungen enger menschlicher Interaktionen zu überwinden. Sie ermöglichte die systematische Untersuchung von beobachtbaren Cliquenstrukturen anhand der tatsächlichen sozialen Relationen, die ihre Mitglieder aneinander binden.

Nehmen wir eine Menge von Personen und eine symmetrische soziale Relation an, die gewisse Personen verbindet, so können wir die «Struktur» der Relation durch einen Graphen darstellen (HARARY, 1969). Jede Person wird durch einen Punkt repräsentiert. Besteht zwischen zwei Personen diese soziale Relation, so wird das durch eine Linie, hier Kante genannt, dargestellt, die beide Punkte verbindet. Eine derartige Ansammlung von Punkten und Kanten bildet einen Graphen.

LUCE & PERRY definierten eine Clique als Untereinheit eines Graphen. Es gilt insbesondere, daß die Cliquen eines Graphen *maximal vollständige Teilgraphen* sind. Eine Teilmenge von Punkten aus einem Graphen bildet zusammen mit den Kanten, die sie verbinden, einen Teilgraphen. Ein Teilgraph ist vollständig, wenn jeder seiner Punkte unmittelbar mit jedem anderen Punkt verbunden ist. Ein vollständiger Teilgraph ist maximal, wenn es in dem Graphen, aus dem er gewonnen wurde, keinen Punkt gibt, der mit allen anderen Punkten des Teilgraphen direkt verbunden ist. Eine graphentheoretische Clique charak-

* Wir danken MARTIN EVERETT, daß er unsere Aufmerksamkeit auf diesen Algorithmus gelenkt hat, HANS RATTINGER dafür, daß er sein FORTRAN-Programm zur Verfügung gestellt hat, und SOPHIA WUERGER, HANS NEUMEIER und CHRISTIAN WERNER für ihre Hilfe bei der Übersetzung ins Deutsche.

terisiert also genau die Art von Gruppe, die Sozialwissenschaftler untersuchen wollen.

Schwierigkeiten mit diesem Formalismus tauchen erst bei den Berechnungen auf. Wenn wir nur wenige Beobachtungen haben, können die Cliquen eines Graphen durch optische Inspektion gefunden werden. Ist aber die Anzahl von Beobachtungen hoch, so ist es unmöglich, die Cliquen durch bloßes Durchsehen der Daten zu ermitteln. HARARY & ROSS (1957) lösten dieses Problem zumindest teilweise durch einen Algorithmus, der mittels einer Matrizendarstellung die Cliquen in einem Graphen berechnet. Sie schlugen vor, jedem Punkt in einem Graphen eine Zeile und eine Spalte in einer quadratischen Matrix zuzuordnen. Wenn eine Kante den Punkt i mit Punkt j verbindet, so wird eine 1 in die Zelle (i, j) der Matrix eingetragen, in allen anderen Fällen eine 0. Die Matrix enthält also die gleiche Information wie der Graph und eignet sich gut für Berechnungen.

Der HARARY-ROSS-Algorithmus ermöglicht die Berechnung der Cliquen in einer Matrix. Aber der Algorithmus wird langsam, wenn die Anzahl der Beobachtungen, besonders die Anzahl der Kanten im Graphen steigt. Jeder Versuch, eine neue Clique zu finden, erfordert, daß alle Punktepaare, zwischen denen Kanten bestehen, untersucht werden. Das heißt: Zunächst werden Paare von Punktepaaren betrachtet, dann Tripel von Punktepaaren usw., bis eine maximale Menge von verbundenen Punkten gefunden ist. Dieses Verfahren ist sehr langwierig. Die Ermittlung von Cliquen wird sogar als ein Beispiel für sog. NP-Rechenprobleme angegeben. Das bedeutet, daß die Lösungszeit mit der Anzahl der Beobachtungen schneller wächst, als es durch Polynome dargestellt werden kann. Deshalb erforderte die Berechnung von Cliquen aus einer Datenmenge von akzeptablem Umfang bisher eine immens lange Rechenzeit.

So standen Sozialwissenschaftler vor der Veröffentlichung von RATTINGERS (1973) Werk einem schwierigen Problem gegenüber. Wir hatten theoretische Gründe, Cliquen zu suchen und ein Verfahren dazu. Das Verfahren war aber so langsam, daß man Datenmengen von realistischer Größe damit nicht angehen konnte. Die Sozialwissenschafter freuten sich also über RATTINGERS schnellen Algorithmus zur Ermittlung von Cliquen.

RATTINGERS Algorithmus reduziert die beiden

Prozesse, die die meiste Zeit benötigen, auf ein Mindestmaß, indem er allmählich gewisse Kanten eliminiert, wenn Cliquen gefunden werden. Jedesmal wenn eine Clique gefunden ist, wird der Graph reduziert. Auf diese Weise müssen weniger Punkte in Betracht gezogen werden, wenn die nächste Clique gesucht wird.

Diesem Algorithmus liegt die Beobachtung zugrunde, daß jedes Paar von Punkten, das in einer Clique enthalten und das nicht mit irgendeinem Punkt außerhalb dieser Clique verbunden ist, nicht in einer weiteren Clique auftauchen kann. Wenn also eine Clique gefunden und notiert worden ist, so können die Kanten zwischen einigen ihrer Punkte entfernt werden, ohne daß dies die Struktur der weiteren Cliquen verändert.

Wenn eine Clique gefunden ist, eliminiert RATTINGERS Verfahren Kanten zwischen Punkten in der betrachteten Clique, die die angegebene Bedingung erfüllen, bevor der Algorithmus die nächste Clique sucht. Diese Methode sorgt dafür, daß mit jeder Clique, die entdeckt worden ist, die Suche nach den restlichen Cliquen beschleunigt wird. Das ganze Verfahren - Cliquen suchen und Kanten eliminieren - geht solange weiter, bis alle Kanten eliminiert worden sind. Ab diesem Moment sind keine Cliquen mehr zu finden.

Manchmal wird eine Clique gefunden, in der es kein Paar von Punkten gibt, dessen verbindende Kante eliminiert werden kann. Wird eine derartige Clique notiert, ohne eine Kante zu eliminieren, so bleibt der Graph wie zuvor, und diesselbe Clique würde immer wieder gefunden. RATTINGER löste dieses Problem, indem er Cliquen, in denen keine Kanten eliminiert werden, einfach nicht notiert. Eine solche Clique wird «ignoriert», und es wird eine andere gesucht, in der mindestens eine Kante eliminiert werden kann. Später, nachdem die das Problem verursachenden Kanten eliminiert worden sind, können auch solche Cliquen herausgefiltert werden.

Dieses Verfahren ist elegant und ziemlich schnell. In den meisten Fällen funktioniert es gut, jedoch nicht in allen. In einigen Datensätzen wird RATTINGERS Algorithmus in eine Endlosschleife getrieben. Das geschieht immer dann, wenn in den übrig bleibenden Cliquen keine Kante eliminiert werden kann.

Man betrachte z.B. den nachfolgenden, aus sechs Punkten bestehenden Graphen. In diesem Graphen sind acht Cliquen enthalten (Abb.1).

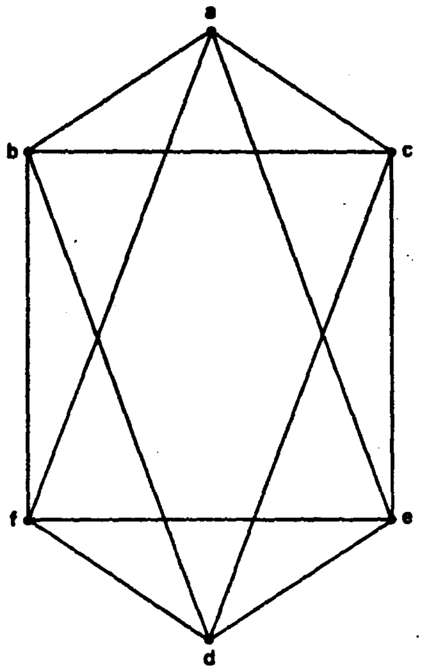

Abb. 1: Graph einer Cliquenstruktur.

Keine dieser Cliquen kann aber mittels RATTINGERS Algorithmus ermittelt werden. Wir wollen das Tripel *abc* einmal näher betrachten. Diese drei Punkte sind alle miteinander verbunden und bilden daher einen vollständigen Teilgraphen. Da kein anderer Punkt mit allen dreien verbunden ist, ist der Teilgraph maximal und deshalb eine Clique. RATTINGERS Algorithmus würde diese Clique finden und versuchen, eine Kante zu eliminieren.

Die Clique *abc* enthält drei Kanten: *ab, ac* und *bc*. Jede dieser Kanten könnte eliminiert werden, wenn es keinen externen Punkt gäbe, der mit beiden Endpunkten der Kante verbunden wäre. Das ist aber für keine Kante der Fall. Die Kante *ab* z. B. verbindet *a* mit *b*; *a* und *b* sind aber beide mit dem externen Punkt *f* verbunden, darum muß die Kante *ab* bleiben. Auf ähnliche Weise sind *a* und *c* an *e* und *b* und *c* an *d* gebunden. Es kann also keine Kante eliminiert werden. Die Clique bleibt in dem Graphen enthalten, und es muß eine andere gesucht werden.

Hier entsteht das Problem mit RATTINGERS Algorithmus. Alle sieben der anderen Cliquen in dem Graphen sind genauso in den Graphen eingebettet wie *abc*. Keine von ihnen erlaubt die Eliminierung irgendeiner Kante, und der Graph kann deshalb nie reduziert werden. In diesem Fall gerät der Algorithmus in eine Endlosschleife und ermittelt überhaupt keine Cliquen. Dieses Problem ergibt sich immer dann, wenn alle übrig gebliebenen Cliquen so in den Graphen eingebettet sind, daß keine Kanten eliminiert werden können.

Dies würde bedeuten, daß wir wieder den Algorithmus von HARARY-ROSS verwenden müssen. RATTINGERS Algorithmus ist jedoch um soviel schneller, daß es wünschenswert ist, möglichst viel von dieser Stärke zu nutzen. Tatsächlich findet RATTINGERS Algorithmus in den meisten Fällen in kurzer Zeit die korrekte Lösung.

Die Lösung des Problems scheint in der Kombination von RATTINGERS Algorithmus mit dem HARARY-ROSS Algorithmus zu liegen. Ein verbessertes Programm würde mit RATTINGERS Algorithmus beginnen und so lang wie möglich damit fortfahren. Bleiben dann noch Cliquen übrig (d.h. sind noch Einsen in der Matrix), die RATTINGERS Algorithmus nicht eliminieren kann, fährt das Programm mit dem HARARY-ROSS Algorithmus fort. Diese Methode nutzt die schnellere Geschwindigkeit von RATTINGERS Algorithmus solange dies möglich ist und macht erst dann vom HARARY-ROSS-Algorithmus Gebrauch, wenn es unbedingt erforderlich ist.

Literatur

HARARY, F. 1969. Graph Theory. Reading, MA: Addison-Wesley.

HARARY, F. & ROSS, I.C. 1957. A procedure for clique detection using the group matrix. Sociometry, 20, 205–215.

LUCE, R.D. & PERRY, A.D. 1949. A method of matrix analysis of group structure. Psychometrika, 14, 95–116.

RATTINGER, H. 1973. Eine einfache Methode und ein FORTRAN-Programm zur Ermittlung von Cliquen. Zeitschrift für Sozialpsychologie, 4, 5–14.