G. Rohwer Wintersemester 2004/05

Notationen für Relationen und Graphen

In diesem Text werden die in der Veranstaltung „Datengewinnung und ihre wissenschaftstheoretischen Grundlagen“ verwendeten Notationen für Relationen und Graphen erläutert.

1 Objekte und Relationen

1. Von Beziehungen wird in vielen unterschiedlichen Varianten geredet. Hier sind einige Beispiele:

- Zwei Menschen kennen sich oder sind befreundet oder sind verheiratet.
- Ein Mensch erzielt ein höheres Einkommen als ein anderer.
- Zwei Schüler sind Mitglieder derselben Schulklasse.
- Ein Mensch ist Angestellter eines bestimmten Unternehmes.
- Ein Unternehmen bezieht von einem anderen Unternehmen Vorleistungen für seine Güterproduktion.
- Zwei Computer sind durch ein Netzwerk verbunden, so dass Daten ausgetauscht werden können.

Dies sind Beispiele für *relationale Aussagen*: Aussagen, die sich gleichzeitig auf jeweils zwei (oder mehr) Objekte beziehen. Zu unterscheiden sind relationale Aussagen und Aussageformen. Zum Beispiel ist

Franz ist verheiratet mit Karin

eine *relationale Aussage*, die ihrer Intention nach einen Sachverhalt ausdrückt und infolgedessen wahr oder falsch sein kann. Dagegen ist

ω ist verheiratet mit ω'

eine *relationale Aussageform*. In diesem Fall sind ω und ω' *logische Variablen*. Relationale Aussagen, die wahr oder falsch sein können, entstehen erst dann, wenn man in die logischen Variablen (Leerstellen) bestimmte Namen einsetzt (z.B. Franz und Karin).

2. Um Schreibweisen abzukürzen, werden oft Symbole verwendet. Wir verwenden im Folgenden das Symbol $\sim$, um auf relationale Ausdrücke zu verweisen. Wenn man inhaltlich bestimmte Aussagen machen möchte, muss natürlich eine Bedeutung vereinbart werden. Zum Beispiel könnte vereinbart werden: Das Symbol $\sim$ soll bis auf weiteres als Abkürzung für den

relationalen Ausdruck ‘ist verheiratet mit’ verwendet werden. Unabhängig von der Vereinbarung einer bestimmten Bedeutung können jedoch mit dem Symbol $\sim$ relationale Aussageformen formuliert werden, die allgemein die Form $\omega \sim \omega'$ haben. In dieser Schreibweise handelt es sich also um eine Aussageform. Erst wenn man dem Symbol $\sim$ eine bestimmte Bedeutung gibt und anstelle von ω und ω' Namen für bestimmte Objekte einsetzt, entsteht eine relationale Aussage, die wahr oder falsch sein kann.

3. Allerdings muss man wissen, auf welche Arten von Objekten man sich beziehen kann, um aus relationalen Aussageformen relationale Aussagen zu machen. Die Umgangssprache orientiert sich an der Bedeutung der relationalen Ausdrücke. Ist z.B. für das Symbol $\sim$ die Bedeutung ‘ist verheiratet mit’ vereinbart worden, ist auch klar, dass man nur dann zu sinnvollen Aussagen gelangt, wenn man für ω und ω' Namen von Menschen einsetzt. Für die weiteren Überlegungen soll angenommen werden, dass man sich jeweils auf eine explizit definierte Menge von Objekten beziehen kann, deren Elemente als Objekte für relationale Aussagen verwendet werden können. Zur symbolischen Repräsentation dient die Schreibweise

$$\Omega := \{\omega_1, \ldots, \omega_n\}$$

Hierbei sind $\omega_1, \ldots, \omega_n$ Namen für die Objekte, auf die man sich gedanklich beziehen möchte, und das Symbol Ω dient zum Verweis auf die Menge dieser Namen bzw. Objekte.

4. Nach diesen Vorüberlegungen kann der Begriff einer Relation, wie er im weiteren verwendet werden soll, explizit definiert werden. Eine *Relation* besteht aus drei Bestandteilen:

a) Es muss ein relationaler Ausdruck $\sim$ eingeführt werden, mit dem relationale Aussageformen der Gestalt $\omega \sim \omega'$ gebildet werden können. (Sobald man nicht nur rein formale Betrachtungen anstellen möchte, muss natürlich auch die inhaltliche Bedeutung angegeben werden.)

b) Es muss eine Objektmenge $\Omega := \{\omega_1, \ldots, \omega_n\}$ angegeben werden, deren Elemente als Namen verwendet werden können, um relationale Aussagen bilden zu können.

c) Schließlich muss angegeben werden, welche der insgesamt möglichen relationalen Aussagen wahr bzw. falsch sind.

Es wäre also eine verkürzte und potentiell irreführende Redeweise, das Symbol $\sim$ eine Relation zu nennen. Dieses Symbol bildet nur ein Hilfsmittel zur Formulierung relationaler Aussagen. Die Relation selbst besteht vielmehr in der Gesamtheit der zutreffenden relationalen Aussagen, die man mithilfe des relationalen Ausdrucks $\sim$ über alle möglichen Paare von Objekten in der Objektmenge Ω machen kann. Sobald man sich dies klargemacht hat, kann man natürlich von einer Relation $(\Omega, \sim)$ sprechen und auch abkürzend von einer Relation $\sim$, wenn der Bezug auf eine bestimmte

Objektmenge durch den Kontext gegeben ist.

5. Ein einfaches Beispiel kann die Begriffsbildungen illustrieren. Die Objektmenge besteht aus 5 Personen: $\Omega := \{\omega_1, \dots, \omega_5\}$, und es soll festgestellt werden, wer mit wem verheiratet ist. Die Bedeutung des Symbols $\sim$ wird also durch 'ist verheiratet mit' festgelegt. Mithilfe der Aussageform $\omega \sim \omega'$ können in diesem Beispiel auf insgesamt 25 unterschiedliche Weisen relationale Aussagen gebildet werden. Einige davon sind richtig, die übrigen sind falsch. Angenommen, dass ω_1 und ω_3 und auch ω_2 und ω_4 verheiratet sind, gibt es folgende Aussagen:

Zutreffende Aussagen	Unzutreffende Aussagen		
$\omega_1 \sim \omega_3$	$\omega_1 \sim \omega_1$	$\omega_2 \sim \omega_5$	$\omega_4 \sim \omega_4$
$\omega_3 \sim \omega_1$	$\omega_1 \sim \omega_2$	$\omega_3 \sim \omega_2$	$\omega_4 \sim \omega_5$
$\omega_2 \sim \omega_4$	$\omega_1 \sim \omega_4$	$\omega_3 \sim \omega_3$	$\omega_5 \sim \omega_1$
$\omega_4 \sim \omega_2$	$\omega_1 \sim \omega_5$	$\omega_3 \sim \omega_4$	$\omega_5 \sim \omega_2$
	$\omega_2 \sim \omega_1$	$\omega_3 \sim \omega_5$	$\omega_5 \sim \omega_3$
	$\omega_2 \sim \omega_2$	$\omega_4 \sim \omega_1$	$\omega_5 \sim \omega_4$
	$\omega_2 \sim \omega_3$	$\omega_4 \sim \omega_3$	$\omega_5 \sim \omega_5$

Die Relation besteht in diesem Beispiel aus der Gesamtheit der 25 Aussagen, von denen 4 zutreffend, die übrigen 21 nicht zutreffend sind.

6. Das Beispiel zeigt, dass sich eine Relation auf alle möglichen Paare von Objekten bezieht, die man aus den Elementen einer Objektmenge bilden kann. In der Mengenlehre verwendet man den Begriff eines kartesischen Produkts. Bezieht man sich allgemein auf zwei Mengen A und B, besteht ihr *kartesisches Produkt*, geschrieben $A \times B$, aus allen geordneten Paaren der Form (a, b), wobei a ein Element aus A und b ein Element aus B ist. Ist z.B. $A := \{1, 2, 3\}$ und $B := \{e, f\}$, erhält man:

$$A \times B = \{(1, e), (1, f), (2, e), (2, f), (3, e), (3, f)\}$$

Bei endlichen Mengen gilt offenbar $|A \times B| = |A| \cdot |B|$, wobei, wenn M irgendeine endliche Menge ist, mit dem Ausdruck $|M|$ auf die Anzahl ihrer Elemente verwiesen werden soll.

7. Verwendet man diese Begriffsbildung, besteht eine Relation für eine Objektmenge Ω darin, dass für jedes Element $(\omega, \omega') \in \Omega \times \Omega$ angegeben wird, ob die relationale Aussage $\omega \sim \omega'$ zutrifft oder nicht. Infolgedessen kann man eine Relation für die Objektmenge Ω durch eine Teilmenge des kartesischen Produkts $\Omega \times \Omega$ festlegen, die genau diejenigen Paare (ω, ω') enthält, für die die relationale Aussage zutrifft. In unserem Beispiel:

$$R^* := \{(\omega_1, \omega_3), (\omega_3, \omega_1), (\omega_2, \omega_4), (\omega_4, \omega_2)\}$$

Diese Methode wird *Definition einer Relation durch ein kartesisches Produkt (einer Objektmenge mit sich selbst)* genannt. Somit kann man auch sagen, dass jeder Teilmenge von $\Omega \times \Omega$ eine jeweils spezifische Relation für die Elemente von Ω entspricht.

8. Eine andere Möglichkeit, um sich begrifflich auf Relationen für eine Objektmenge Ω zu beziehen, besteht in der Verwendung *relationaler Variablen*. Mit diesem Begriff sind (zunächst) Funktionen gemeint, die folgende Form haben:

$$R : \Omega \times \Omega \longrightarrow \{0, 1\}$$

R ist der Name der Funktion (der relationalen Variablen), $\Omega \times \Omega$ ist ihr Definitionsbereich und $\{0, 1\}$ ist ihr Wertebereich. Die Funktion (relationale Variable) R ordnet also jedem Element $(\omega, \omega') \in \Omega \times \Omega$ einen Wert $R(\omega, \omega') \in \{0, 1\}$ zu, wobei folgende Bedeutung vereinbart wird:

$$R(\omega, \omega') = \begin{cases} 1 & \text{wenn } \omega \sim \omega' \text{ zutrifft} \\ 0 & \text{wenn } \omega \sim \omega' \text{ nicht zutrifft} \end{cases}$$

Wie sich später zeigen wird, ist der Begriff einer relationalen Variablen sehr nützlich, weil er sich leicht verallgemeinern lässt, um in komplexerer Weise von Relationen zu sprechen. Außerdem gibt es eine gedanklich einfache Parallele zu statistischen Variablen, also zu Funktionen der Form

$$X : \Omega \longrightarrow \tilde{\mathcal{X}}$$

die jedem Element einer Objektmenge Ω einen Merkmalswert in einem Merkmalsraum $\tilde{\mathcal{X}}$ zuordnen. Der Unterschied besteht nur darin, dass eine statistische Variable jedem einzelnen Objekt, eine relationale Variable dagegen jedem Paar von Objekten einen Merkmalswert zuordnet.

9. An dieser Parallele knüpft auch eine weitere Möglichkeit zur Darstellung von Relationen an. Beziehen wir uns zunächst auf eine statistische Variable $X : \Omega \longrightarrow \tilde{\mathcal{X}}$. Ihre Werte (die Daten) können in Form einer Datenmatrix dargestellt werden, die folgende Form hat:

ω	$X(\omega)$
ω_1	$X(\omega_1)$
$\vdots$	$\vdots$
ω_n	$X(\omega_n)$

Jede Zeile bezieht sich auf jeweils ein Objekt der Objektmenge Ω. Die erste Spalte enthält den Namen des Objekts, die zweite Spalte den Merkmalswert, der dem Objekt durch die Variable zugeordnet wird. Auf ähnliche Weise kann man die Werte einer relationalen Variablen durch ein zweidimensionales Schema darstellen, das allgemein folgende Form hat:

	ω_1	$\cdots$	ω_n
ω_1	$R(\omega_1,\omega_1)$	$\cdots$	$R(\omega_1,\omega_n)$
$\vdots$	$\vdots$		$\vdots$
ω_n	$R(\omega_n,\omega_1)$	$\cdots$	$R(\omega_n,\omega_n)$

Für das oben (in Textziffer 5) angeführte Beispiel erhält man folgende Darstellung:

	ω_1	ω_2	ω_3	ω_4	ω_5
ω_1	0	0	1	0	0
ω_2	0	0	0	1	0
ω_3	1	0	0	0	0
ω_4	0	1	0	0	0
ω_5	0	0	0	0	0

Wenn ein Schema dieser Art verwendet wird, um eine Relation darzustellen, spricht man von einer *Adjazenzmatrix*.

10. Zur formalen Charakterisierung von Relationen gibt es zahlreiche Begriffsbildungen. An dieser Stelle genügen die folgenden, zu deren Erläuterung angenommen wird, dass eine Relation $(\Omega,\sim)$ gegeben ist.

a) Die Relation $(\Omega,\sim)$ wird *reflexiv* genannt, wenn gilt:

$$\text{Für alle } \omega\in\Omega:\ \omega\sim\omega$$

b) Die Relation $(\Omega,\sim)$ wird *symmetrisch* genannt, wenn gilt:

$$\text{Für alle } \omega,\omega'\in\Omega:\ \omega\sim\omega' \implies \omega'\sim\omega$$

c) Die Relation $(\Omega,\sim)$ wird *transitiv* genannt, wenn gilt:

$$\text{Für alle } \omega,\omega',\omega''\in\Omega:\ \omega\sim\omega' \text{ und } \omega'\sim\omega'' \implies \omega\sim\omega''$$

Diese Eigenschaften gelten für eine Relation immer dann, wenn sie nicht durch ein Gegenbeispiel widerlegt werden können.

Wenn alle drei Eigenschaften bestehen, spricht man auch von einer *Äquivalenzrelation*. Die oben als Beispiel verwendete Relation ist offenbar symmetrisch, jedoch weder reflexiv noch transitiv.

2 Ungerichtete Graphen

1. Unter einem *Graphen* versteht man allgemein eine Menge von *Knoten*, die durch *Kanten* (Linien oder Pfeile) verbunden sein können. Die

Knoten entsprechen den Objekten, auf die man sich beziehen möchte, die Kanten werden zur Darstellung von Beziehungen zwischen den Knoten (Objekten) verwendet. Zur Notation wird die Schreibweise $\mathcal{G} := (\Omega,\mathcal{K})$ verwendet, wobei $\Omega := \{\omega_1,\ldots,\omega_n\}$ die *Knotenmenge* des Graphen und $\mathcal{K} := \{\kappa_1,\ldots,\kappa_m\}$ die *Kantenmenge* des Graphen ist.

2. Diese Erläuterung zeigt bereits, dass es einen engen Zusammenhang zwischen Relationen und Graphen gibt. Zunächst besprechen wir *ungerichtete Graphen*, die den symmetrischen Relationen entsprechen. Sei also $(\Omega,\sim)$ eine symmetrische Relation. Dann kann man Ω auch als Knotenmenge eines Graphen betrachten und festlegen, dass es zwischen zwei Knoten $\omega,\omega'\in\Omega$ genau dann eine Kante gibt, wenn die relationale Aussage $\omega\sim\omega'$ zutrifft. Die Kantenmenge wird also durch

$$\mathcal{K} := \{\,\{\omega,\omega'\}\,|\,\omega\sim\omega' \text{ ist zutreffend}\,\}$$

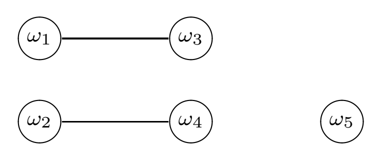
definiert. Anstelle von geordneten Paaren der Form (ω,ω') werden in diesem Fall Mengen der Form $\{\omega,\omega'\}$ verwendet, da die Relation symmetrisch ist, so dass die Reihenfolge keine Rolle spielt.

3. Zur Illustration kann zunächst das bereits im vorangegangenen Abschnitt verwendete Beispiel dienen. In diesem Fall repräsentieren die Knoten die 5 Personen und die Kanten zeigen, welche der Personen miteinander verheiratet sind. In symbolischer Notation hat dieser Graph folgende Form:

$$\mathcal{G} := (\,\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\},\ \{\{\omega_1,\omega_3\},\{\omega_2,\omega_4\}\}\,)$$

Hierbei ist $\{\omega_1,\omega_2,\omega_3,\omega_4,\omega_5\}$ die Objektmenge und $\{\{\omega_1,\omega_3\},\{\omega_2,\omega_4\}\}$ die Kantenmenge.

4. Anhand dieses Beispiels kann auch die *graphische Darstellung* von Graphen erläutert werden. Jeder Knoten des Graphen wird durch einen Punkt (oder Kreis, Rechteck, ...) und jede Kante durch eine Verbindungslinie zwischen den zugehörigen Knoten dargestellt. In diesem Beispiel kann man folgende Darstellung verwenden:

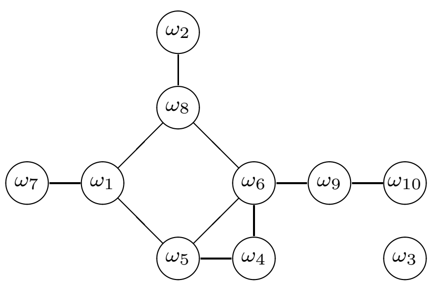


Die Anordnung der Knoten kann beliebig erfolgen, denn sie hat keine Bedeutung. Oft wählt man eine Anordnung, die möglichst keine oder nur wenige Überschneidungen der die Kanten repräsentierenden Linien erfordert. Ein Graph wird *planar* genannt, wenn man ihn vollständig ohne Überschneidungen darstellen kann.

5. Für ein weiteres Beispiel können Daten dienen, die in der ersten Stunde eines Seminars über soziale Netzwerke, an dem 10 Personen teilgenommen haben, erhoben wurden. Das Ziel war herauszufinden, welche Teilnehmer „sich bereits kennen“. Um das zu präzisieren, wurde folgende Fragestellung gewählt: Haben jeweils zwei der Teilnehmer vor Beginn des Seminars schon mindestens 5 Minuten miteinander gesprochen? Um die Daten zu gewinnen, wurde zunächst eine Liste der Teilnehmer erstellt:

$$\Omega := \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8, \omega_9, \omega_{10}\}$$

Dann wurde jeder Teilnehmer gefragt, mit welchen anderen Seminarteilnehmern er bereits vor Beginn des Seminars mindestens 5 Minuten gesprochen hat. Tabelle 1 zeigt das Ergebnis in Gestalt einer Adjazenzmatrix. Sie beschreibt einen Graphen, dessen Knoten aus den 10 Teilnehmern des Seminars bestehen. Die Einsen geben die Kanten des Graphen an und bedeuten, daß zwischen den jeweils beteiligten Knoten eine „Beziehung“ besteht, in diesem Beispiel dadurch definiert, daß bereits vor Beginn des Seminars eine Kommunikation stattgefunden hat. Da es sich um eine symmetrische Relation handelt, ist auch die Adjazenzmatrix symmetrisch und man kann zur Repräsentation einen ungerichteten Graphen verwenden, wie folgende graphische Darstellung zeigt.



6. An dieser Stelle kann auch bereits ein weiterer Begriff erläutert werden: der *Grad eines Knotens*. Bei einem ungerichteten Graphen versteht man darunter die Anzahl der Kanten, die von dem Knoten ausgehen bzw. in ihn münden. Um den Grad eines Knotens ω zu bezeichnen, verwenden wir die Notation $g(\omega)$. Die Berechnung kann am einfachsten mithilfe der Adjazenzmatrix des Graphen erfolgen. Bezeichnet $\mathbf{A} = (a_{ij})$ die Adjazenzmatrix, gilt nämlich:

$$g(\omega_i) = \sum_{j=1}^{n} a_{ij} = \sum_{j=1}^{n} a_{ji}$$

D.h. bei ungerichteten Graphen ist der Grad eines Knotens gleich der

Tabelle 1 Adjazenzmatrix der Seminardaten.

	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
ω_1	0	0	0	0	1	0	1	1	0	0
ω_2	0	0	0	0	0	0	0	1	0	0
ω_3	0	0	0	0	0	0	0	0	0	0
ω_4	0	0	0	0	1	1	0	0	0	0
ω_5	1	0	0	1	0	1	0	0	0	0
ω_6	0	0	0	1	1	0	0	1	1	0
ω_7	1	0	0	0	0	0	0	0	0	0
ω_8	1	1	0	0	0	1	0	0	0	0
ω_9	0	0	0	0	0	1	0	0	0	1
ω_{10}	0	0	0	0	0	0	0	0	1	0

Zeilen- oder Spaltensumme der zu diesem Knoten gehörigen Zeile oder Spalte der Adjazenzmatrix. In unserem Beispiel findet man:

ω	ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8	ω_9	ω_{10}
$g(\omega)$	3	1	0	2	3	4	1	3	2	1

Offenbar liefert der Grad eines Knotens eine Information darüber, in welchem Ausmaß der Knoten in das Netzwerk eingebunden ist. Gibt es insgesamt n Knoten, kann der Grad eines Knotens bei einer nicht-reflexiven Beziehung maximal den Wert $n - 1$ annehmen. Der minimale Wert ist natürlich Null. Knoten, die den Grad Null haben, werden auch *isolierte Knoten* genannt.

3 Gerichtete Graphen

1. Oft sind Relationen nicht symmetrisch. Dann werden *gerichtete Graphen* verwendet. Zur symbolischen Notation kann wie bei ungerichteten Graphen die Formulierung $\mathcal{G} := (\Omega, \mathcal{K})$ verwendet werden. Es muss nur berücksichtigt werden, dass bei gerichteten Graphen die Kantenmenge $\mathcal{K}$ aus *geordneten* Paaren von Knoten besteht, so dass bei zwei Knoten ω und ω' zwischen Kanten, die von ω zu ω' führen, und Kanten, die von ω' zu ω führen, unterschieden werden kann. Zur Unterscheidung wird von *gerichteten Kanten* gesprochen. In der graphischen Darstellung werden deshalb nicht Linien, sondern Pfeile verwendet.

2. Als Beispiel betrachten wir eine Objektmenge Ω, die aus 5 Unternehmen besteht. Mit den relationalen Aussagen der Form $\omega \sim \omega'$ soll erfasst werden, ob das Unternehmen ω' Produkte des Unternehmens ω als Vorleistungen verwendet. Es werden vielleicht folgende Beziehungen festgestellt:

$$\omega_1 \sim \omega_2,\;\; \omega_3 \sim \omega_2,\;\; \omega_4 \sim \omega_3,\;\; \omega_4 \sim \omega_5$$

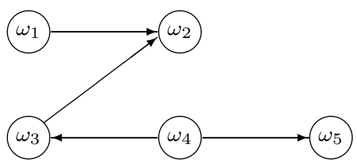
so dass die Adjazenzmatrix folgendermaßen aussieht:

$$\mathbf{A} := \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Offenbar ist diese Adjazenzmatrix und dementsprechend auch die Relation nicht symmetrisch. Zur Darstellung wird deshalb ein gerichteter Graph verwendet, dessen Kantenmenge durch

$$\mathcal{K} := \{(\omega_1, \omega_2), (\omega_3, \omega_2), (\omega_4, \omega_3), (\omega_4, \omega_5)\}$$

definiert ist. Es handelt sich um gerichtete Kanten, und die graphische Darstellung sieht folgendermaßen aus:



3. Bei gerichteten Graphen muß unterschieden werden zwischen der Anzahl der Kanten, die in einen Knoten einmünden, und der Anzahl der Kanten, die von ihm ausgehen. Im ersten Fall spricht man vom *Eingangsgrad*, im zweiten Fall vom *Ausgangsgrad* eines Knotens. Bezieht man sich auf eine Adjazenzmatrix $\mathbf{A} = (a_{ij})$, erhält man folgende Definitionen:[1]

$$g_a(\omega_i) := \sum_{j=1}^{n} a_{ij} \quad \text{und} \quad g_e(\omega_i) := \sum_{j=1}^{n} a_{ji}$$

D.h. anhand der Adjazenzmatrix kann für jeden Knoten sein Eingangsgrad durch die Spaltensumme und sein Ausgangsgrad durch die Zeilensumme abgelesen werden. In unserem Beispiel findet man folgende Werte:

ω	$g_a(\omega)$	$g_e(\omega)$
ω_1	1	0
ω_2	0	2
ω_3	1	1
ω_4	2	0
ω_5	0	1

[1] Bei der Darstellung eines gerichteten Graphen durch eine Adjazenzmatrix $\mathbf{A} = (a_{ij})$ wird stets von der Konvention ausgegangen, dass die Richtung von i (Zeilen) zu j (Spalten) gegeben ist.

In diesem Beispiel gibt der Eingangsgrad eines Unternehmens an, von wie vielen anderen Unternehmen es Vorleistungen bezieht; der Ausgangsgrad gibt an, an wie viele andere Unternehmen Güter als Vorleistungen abgegeben werden.

4 Weitere Definitionen

1. Kanten, die in einem (ungerichteten oder gerichteten) Graphen Knoten mit sich selbst verbinden, werden *Schlingen* genannt. Graphen ohne Schlingen werden *einfache Graphen* genannt. Bei vielen Begriffsbildungen bezieht man sich nur auf einfache Graphen.

2. So wird z.B. der Begriff der *Dichte* eines Graphen normalerweise nur für einfache Graphen verwendet und folgendermaßen definiert:

$$\text{Dichte} := \frac{\text{Anzahl der vorhandenen Kanten}}{\text{Anzahl der möglichen Kanten}}$$

wobei es bei einem gerichteten Graphen mit n Knoten $n(n-1)$ mögliche Kanten und bei einem ungerichteten Graphen mit n Knoten $n(n-1)/2$ mögliche Kanten gibt. Ein Graph, bei dem alle möglichen Kanten vorhanden sind, wird ein *vollständiger Graph* genannt.

3. Ein Graph $(\Omega', \mathcal{K}')$ heißt ein *Teilgraph* eines Graphen $(\Omega, \mathcal{K})$, wenn Ω' eine Teilmenge von Ω und $\mathcal{K}'$ eine Teilmenge von $\mathcal{K}$ ist. Ein maximaler Teilgraph, der aus mindestens drei Knoten besteht und bei dem jeder Knoten mit jedem anderen Knoten verbunden ist, wird eine *Clique* genannt.

4. Zur Definition von Komponenten beziehen wir uns zunächst auf einen ungerichteten Graphen mit der Knotenmenge {1,...,n}. Dann ist mit dem Begriff *Weg* eine Folge von Knoten $i_0, \ldots, i_m$ gemeint, so dass es zwischen je zwei aufeinander folgenden Knoten eine Kante gibt.[2] Man sagt auch, dass ein solcher Weg vom Knoten i_0 zum Knoten i_m führt; die Anzahl der Kanten, also m, wird *Länge des Weges* genannt. Offenbar kann man den Weg auch als eine Folge von Kanten beschreiben, die der Reihe nach durchlaufen werden.

Im allgemeinen kann es zwischen jeweils zwei Knoten eines ungerichteten Graphen einen, mehrere oder auch keinen Weg geben. Darauf bezieht sich der Begriff einer Komponente: Eine *Komponente eines ungerichteten Graphen* ist ein maximaler Teilgraph, bei dem für jeweils zwei Knoten gilt, dass sie durch mindestens einen Weg miteinander verbunden sind. Ein ungerichteter Graph, der nur aus einer einzigen Komponente besteht, wird *zusammenhängend* genannt. Als Beispiel kann man an den in Abschnitt 2

[2] Bei dieser allgemeinen Definition ist also zugelassen, dass dieselbe Kante innerhalb eines Wegs mehrfach auftreten kann. Wenn dies ausgeschlossen werden soll, sprechen wir von Wegen ohne Kantenwiederholungen.

(Textziffer 4) angegebenen Graphen denken, der aus 5 Knoten und 2 Kanten besteht. Dieser Graph ist nicht zusammenhängend, sondern zerfällt in 3 Komponenten. Es gilt allgemein, dass isolierte Knoten als separate Komponenten angesehen werden.

5. Die eben angegebene Definition gilt nur für ungerichtete Graphen. Bei gerichteten Graphen kann man zunächst in zwei unterschiedlichen Weisen von Wegen sprechen:

a) Eine Folge von Knoten $i_0, \ldots, i_m$ wird ein *(gerichteter) Weg* von i_0 nach i_m genannt, wenn jeweils zwei aufeinander folgende Knoten i_k und i_{k+1} durch eine gerichtete Kante von i_k nach i_{k+1} verbunden sind.

b) Eine Folge von Knoten $i_0, \ldots, i_m$ wird ein *Semi-Weg* von i_0 nach i_m genannt, wenn jeweils zwei aufeinander folgende Knoten i_k und i_{k+1} durch eine Kante verbunden sind, die von i_k nach i_{k+1} *oder* von i_{k+1} nach i_k führt.

Dementsprechend unterscheidet man bei gerichteten Graphen zwischen zwei Arten von Komponenten: Eine *Komponente* ist ein maximaler Teilgraph, bei dem jeweils zwei Knoten durch mindestens einen Weg verbunden sind; dagegen spricht man von einer *Semi-Komponente*, wenn nur gefordert wird, dass jeweils zwei Knoten durch mindestens einen Semi-Weg verbunden sind. Ein gerichteter Graph, der nur aus einer einzigen Komponente besteht, wird *zusammenhängend* oder auch *unzerlegbar* genannt.

5 Bewertete Graphen

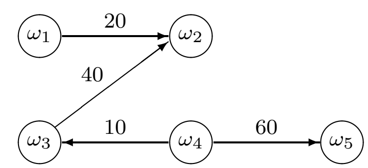
1. Bei einer Relation $(\Omega, \sim)$ wird nur festgestellt, ob für jeweils zwei Objekte $\omega, \omega' \in \Omega$ die relationale Aussage $\omega \sim \omega'$ zutrifft oder nicht. Zum Beispiel: Zwei Personen sind verheiratet oder nicht verheiratet. Oft ist es jedoch von Interesse, qualitative oder quantitative Unterschiede in der Art der Beziehung zu erfassen. Zum Beispiel könnte man bei persönlichen Beziehungen zwischen Bekanntschaften und Freundschaften unterscheiden; oder bei dem in Abschnitt 3 verwendeten Beispiel könnte man unterscheiden, in welchem Ausmaß Vorleistungen bezogen werden. Um solche Unterscheidungen berücksichtigen zu können, werden *bewertete Graphen* verwendet: Jeder (gerichteten oder ungerichteten) Kante des Graphen wird dann eine Zahl zugeordnet, die die durch die Kante repräsentierte Beziehung charakterisiert.

2. Als Beispiel verwenden wir wieder eine Objektmenge, die aus 5 Unternehmen besteht. In diesem Fall soll es sich jedoch um Aktiengesellschaften handeln, so dass man feststellen kann, wie viel Prozent des Aktienkapitals eines Unternehmens von einem anderen Unternehmen gehalten wird. Solche Daten können wiederum in Form einer Adjazenzmatrix dargestellt

werden, wobei jetzt aber in den einzelnen Feldern der Matrix die Prozentanteile des Kapitalbesitzes eingetragen werden. In unserem Beispiel sieht die Adjazenzmatrix vielleicht folgendermaßen aus:

$$\mathbf{A} := \begin{pmatrix} 0 & 20 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 40 & 0 & 0 & 0 \\ 0 & 0 & 10 & 0 & 60 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Das Unternehmen ω_1 hält am Unternehmen ω_2 20 % der Kapitalanteile usw. Man erhält dann folgende graphische Darstellung:



3. Zur symbolischen Notation bewerteter Graphen wird in der Literatur oft die Formulierung

$$\mathcal{G} := (\Omega, \mathcal{K}, v)$$

verwendet. Ω ist die Knotenmenge, $\mathcal{K}$ die Kantenmenge. Hinzu kommt eine Funktion[3]

$$v : \mathcal{K} \longrightarrow \mathbf{R}$$

die jeder Kante $\kappa \in \mathcal{K}$ eine Zahl $v(\kappa) \in \mathbf{R}$ zuordnet und als *Bewertung der Kante* bezeichnet wird (wobei natürlich eine jeweils sinnvolle Bedeutung vereinbart werden muss). In unserem Beispiel sieht diese Funktion folgendermaßen aus:

κ	$v(\kappa)$
(ω_1, ω_2)	20
(ω_3, ω_2)	40
(ω_4, ω_3)	10
(ω_4, ω_5)	60

4. Als einheitlicher begrifflicher Rahmen für Graphen aller Art eignen sich

[3]Dabei bedeutet $\mathbf{R}$ die Menge der reellen Zahlen.

am besten relationale Variablen, die in allgemeiner Weise als Funktionen der folgenden Form definiert sind:

$$R : \Omega \times \Omega \longrightarrow \tilde{\mathcal{R}}$$

Hierbei ist Ω eine Objektmenge und $\tilde{\mathcal{R}}$ ein im Prinzip beliebig konzipierbarer Merkmalsraum. Die relationale Variable R ordnet jedem Element $(\omega, \omega') \in \Omega \times \Omega$ einen Merkmalswert $R(\omega, \omega') \in \tilde{\mathcal{R}}$ zu. Wie bereits besprochen wurde, genügt für unbewertete Graphen ein Merkmalsraum $\tilde{\mathcal{R}} := \{0, 1\}$, da nur unterschieden werden muss, ob zwischen zwei Objekten eine Beziehung besteht oder nicht. Wenn man differenziertere Merkmalsräume verwendet, können jedoch auch beliebige bewertete Graphen repräsentiert werden. Für das zuvor besprochene Beispiel kann man als Merkmalsraum z.B. die Zahlen von 0 bis 100 verwenden, und $R(\omega, \omega')$ bedeutet dann den Prozentanteil des Kapitals des Unternehmens ω', den das Unternehmen ω besitzt. Relationale Variablen bieten also sehr flexible Formulierungsmöglichkeiten. Außerdem lassen sich viele Überlegungen und Unterscheidungen, die für statistische Variablen bereits eingeführt worden sind, unmittelbar übertragen.

6 Bi-modale Graphen

1. Zum Abschluß soll noch kurz erwähnt werden, dass man sich auch für Beziehungen zwischen Objekten interessieren kann, die unterschiedlichen Arten von Objektmengen angehören. Als Beispiel kann man sich auf die Frage beziehen, an welchen Lehrveranstaltungen die Studenten eines bestimmten Studiengangs teilnehmen. Dann gibt es zwei Objektmengen. Erstens eine Objektmenge $\Omega := \{\omega_1, \ldots, \omega_n\}$, die die Studenten repräsentiert, zweitens eine Objektmenge $\Omega^* := \{\omega_1^*, \ldots, \omega_m^*\}$, die die Lehrveranstaltungen repräsentiert. Ist nun $\omega \in \Omega$ und $\omega^* \in \Omega^*$, soll die Aussage $\omega \sim \omega^*$ bedeuten, dass der Student ω an der Lehrveranstaltung ω^* teilnimmt. Da es in diesem Fall zwei Objektmengen gibt, spricht man von einer *bi-modalen Relation* bzw. von einem *bi-modalen Graphen*. Daten können durch eine *bi-modale Adjazenzmatrix* erfasst werden, die folgende allgemeine Form hat:

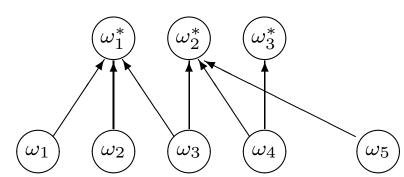
	ω_1^*	$\cdots$	ω_m^*
ω_1	a_{11}	$\cdots$	a_{1m}
$\vdots$	$\vdots$		$\vdots$
ω_n	a_{n1}	$\cdots$	a_{nm}

Wenn $a_{ij} = 1$ ist, nimmt der Student ω_i an der Lehrveranstaltung ω_j^* teil, andernfalls nicht.

2. Als Beispiel kann man sich vorstellen, dass es 5 Studenten und 3 Lehrveranstaltungen gibt und dass die bi-modale Adjazenzmatrix folgendermaßen aussieht:

$$\mathbf{A} := \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Den entsprechenden bi-modalen Graphen kann man sich dann durch folgende Darstellung veranschaulichen:



3. Wiederum kann man auch relationale Variablen verwenden, die jetzt folgende allgemeine Form haben:

$$R : \Omega \times \Omega^* \longrightarrow \tilde{\mathcal{R}}$$

Jedem Element $(\omega, \omega^*) \in \Omega \times \Omega^*$ wird ein bestimmter Wert $R(\omega, \omega^*) \in \tilde{\mathcal{R}}$ zugeordnet, der entweder nur feststellt, ob eine Beziehung besteht, oder (bei einem bewerteten bi-modalen Graphen) diese Beziehung näher charakterisiert. Man spricht dann von einer *bi-modalen relationalen Variablen*.