

# BIOGRAPH

PAUL GEISLER

ZUSAMMENFASSUNG. Dieser Artikel behandelt die Verarbeitung und Analyse biographischer Daten mittels Graphen.

## 1. EINFÜHRUNG

Um biographische Daten effizienter als Curriculum-Tabellen oder Texte zu verarbeiten/analysieren, müssen sie in ein maschinerechtes Format gebracht werden. Als Werkzeug dazu wurde die Software MutaGraph, ein angepasstes Hilfsmittel zur Graphenverarbeitung erstellt. Doch zunächst davon unabhängig die mathematische Darstellung:

**Definition 1.1** (Graph, Kantenmenge, Kante, Knotenmenge, Knoten). Ein *Graph*  $G$  nennt man ein Paar  $G := (V, E)$  der *Knotenmenge*  $V$  von endlich vielen *Knoten*  $v_i \in V; i = 1 \dots n$  und der *Kantenmenge*  $E$  von *Kanten*  $e_i = (v_a, v_b) \in E \subset V^2; i = 1 \dots m$ .

Ein Graph besteht also aus einem Haufen Knoten, von denen einige paarweise durch Kanten verbunden sind, wobei eine Kante *gerichtet* ist, also einen Start- und einen Endknoten besitzt. In Mutagraph wird jedem Knoten und jeder Kante zusätzlich ein Typ zugeordnet:

$$typ_V : V \longrightarrow VT \text{ und } typ_E : E \longrightarrow ET$$

wobei  $VT$  Menge der Knotentypen bzw.  $ET$  Menge der Kantentypen.

Ein Beispiel:  $G = (V, E)$  mit

$$V = \{Paul, Pullover\}, E = \{e_1 = (Paul, Pullover)\}$$

sowie  $typ(Paul) = Person$ ,  $typ(Pullover) = Gegenstand$  und  $typ(e_1) = besitzt$ , unter Verwendung der Typmengen

$$VT = \{Person, Gegenstand\}$$

bzw.

$$ET = \{besitzt\}$$

Man sieht, das eine Kante keinen expliziten "Namen" besitzt, dieser wird hier mit ihrem Typ identifiziert.

Typgleichheit bildet offensichtlich eine Äquivalenzrelation über den Knoten bzw. Kanten, die Menge aller Äquivalenzklassen bzgl. der Knotentypen bezeichnen wir mit  $V/KT$  wobei klar:  $|V/KT| = |KT|$ .

Für die Analysen des Datenbestandes greifen wir auf weitere Begriffe aus der Graphentheorie zurück:

**Definition 1.2.** Eine Kante  $e$  nennt man zu einem Knoten  $v$  *inzident*, falls  $e = (v_a, v) \vee e = (v, v_b)$ . Die Zahl aller in einen Knoten  $v \in V$  einlaufenden Kanten  $d_<(v) := |\{e \in E | e = (v_a, v), v_a \in V\}|$  nennt man die *Eingangswalenz* des Knotens  $v$ . Analog nennt man die Zahl aller aus einen Knoten  $v \in V$  hinauslaufenden Kanten  $d_>(v) := |\{e \in E | e = (v, v_b), v_b \in V\}|$  *Ausgangswalenz* des Knotens  $v$ . Die

Gesamtzahl aller mit einem Knoten inzidierenden Kanten  $d(v) := d_{<}(v) + d_{>}(v)$  nennt man *Gesamtvalenz*<sup>1</sup>.

Eine einfache Analyse der Daten ist die Knotenauflistung nach Valenz. Man erhält eine geordnete Liste  $L_V := ((v_1, d(v_1)), \dots, (v_n, d(v_n)))$ . Bildet man die Klasseneinteilung  $L_{KT} := L_V \cap V/_{KT}$  so kann man die "wichtigsten" Elemente der Klassen suchen. Es ist zu vermuten, dass die Eltern, der engste Freund und der ärgste Feind sich auf der Liste der Klasse "Person" an den ersten Stellen befinden...

*E-mail address:* paul.geisler@web.de

---

<sup>1</sup>hier oft einfach als *Valenz* benutzt