phwt
Private Hochschule
für Wirtschaft und Technik

Soziale Netzwerkanalyse:

Interaktiv zu einem entropieoptimalen brauchbaren sozialen Netz

Elmar Reucher*, Andreas Dellnitz** und Wilhelm Rödder***

Diskussionsbeitrag der Privaten Hochschule für Wirtschaft und Technik (PHWT)

Nr. 1 / Oktober 2016

Herausgegeben vom Präsidenten der Hochschule

* Studienbereich Betriebswirtschaft der Privaten Hochschule für Wirtschaft und Technik (PHWT)
** Lehrstuhl für Wirtschaftsmathematik und Quantitative Methoden der FernUniversität in Hagen
*** Forschungsbereich OR der FernUniversität in Hagen

Inhaltsverzeichnis

Abbildungsverzeichnis

Tabellenverzeichnis

1 Einleitung

Hört man gegenwärtig von Sozialen Netzwerken (SN), so assoziiert man damit in der Regel sofort die des Internets (SNI), also Facebook usw. Soziologen verstehen unter SN jedoch jedes Beziehungsgeflecht zwischen Akteuren oder Gruppen von Akteuren; sie – die SN – werden also zunächst ganz allgemein und losgelöst vom Kommunikationsmedium definiert. Dennoch haben gerade die SNI den Soziologen, Betriebswirten und weiteren Disziplinen ein neues Feld für empirische sowie theoretische Analysen – und darunter gerade den quantitativ orientierten – geliefert.
Die quantitative Soziale Netzwerkanalyse geht dabei grundlegend auf Jakob Moreno, einen Soziologen der Wiener Schule, zurück. 1934 verwendete er erstmalig eine graphentheoretische Repräsentationsform von Sozialen Netzwerken: die Soziogramme; man vgl. dazu [6]. Akteure wurden hier durch Knoten und die zwischen ihnen herrschenden Beziehungen durch Kanten dargestellt. Zur Untersuchung von SN konnten nunmehr also auch mathematische, eben graphentheoretische, Methoden und Techniken angewendet, vor allem aber – "befeuert"durch diese neue Repräsentationsform – weitere entwickelt werden. Das Ziel ist hier insgesamt stets klar: Die Charakterisierung von Akteuren, Gruppen von Akteuren, gar ganzen SN oder den entsprechenden Beziehungen zwischen ebendiesen. Für einen tieferen Einblick in die historischen Entwicklungen und jeweiligen wissenschaftlichen Strömungen sei verwiesen auf [3] und [13]. Die mathematischen Instrumente hingegen werden umfangreich in [7] vorgestellt.
Eine neuere Form der Modellierung und Analyse von Sozialen Netzwerken lehnt sich an die graphentheoretische Sicht an, modelliert die Beziehungen darin jedoch mittels konditionaler (probabilistischer) Logik – wenn-dann. Wenn also ein Akteur oder eine Gruppe über ein immaterielles Gut (z.B. eine Information) oder eine Attitüde verfügt, so gibt bzw. vererbt er oder sie diese an die Nachfolger weiter, siehe dazu [9]. Rödder et al. entwickeln in diesem Beitrag Maße, die in diesem Modellrahmen eine Bewertung des Diffusions- bzw. Rezeptionskapitals von Akteuren und/oder ihrer Einbettung im SN vorzunehmen erlaubt. Dazu wird auf der Menge aller Zustände des Netzwerkes – Akteure haben das immaterielle Gut/die Attitüde oder eben nicht – eine Wahrscheinlichkeitsverteilung Q aufgebaut, die dem Prinzip Minimaler Relativer Entropie oder Kullback-Leibler Divergenz zur Gleichverteilung genügt; genaueres zu diesem Prinzip findet man z.B. in [2, 14, 4]. Sie – obige Autoren – betrachten in [9] aber nur sichere Weitergaben. Will heißen: Sie nehmen jedes Konditional – wenn-dann – als sicher an. Wird evident, dass ein Akteur im Besitz des immateriellen Gutes oder der Attitüde ist, so gibt er

diese mit Wahrscheinlichkeit 1 weiter. In [11] wird diese Forderung gelockert, und es wird weiterhin gezeigt, dass sich bei proportionaler Skalierung aller Weitergabewahrscheinlichkeiten (WWn) noch immer eine globale Wahrscheinlichkeitsverteilung Q auf der Menge aller Netzzustände bestimmen lässt, die zur weiteren Analyse des SN geeignet ist. Diese Eignung von Q belegen die Autoren mit dem Begriff brauchbar. Die Autoren behandeln aber auch noch den Fall der beliebigen Vergabe solcher WWn. Dies kann zu einem brauchbaren oder zu unbrauchbaren Q führen. Letzteres bedeutet, dass ein entartetes Q vorliegt, das eine Netzwerkanalyse mit den zuvor genannten Maßen nur noch begrenzt zulässt. Bei beliebigen Vergaben von Weitergabewahrscheinlichkeiten ist die Brauchbarkeit von Q also stets zu überprüfen. In welchem Bereich Weitergabewahrscheinlichkeiten von bestehenden Beziehungen noch variieren können, bevor Q unbrauchbar wird, behandeln [10]; sie unterziehen die entsprechenden Konditionale eine Ungewissheitsanalyse und erhalten dadurch ein Ungewissheitsintervall (UI). Letztlich bedeutet das, dass die beobachtete konditionale Struktur eines SN im Regelfall nicht hinreicht, um Q – und damit auch alle Weitergabewahrscheinlichkeiten – vollständig zu determinieren. Aber mehr noch: In [10] wird dargestellt, dass ein einmal bestimmtes Q auch Schätzungen von Weitergabewahrscheinlichkeiten über selbst noch nicht beobachtete Beziehungen vorzunehmen erlaubt und dazu ebenfalls das entsprechende UI bestimmbar ist.

Der vorliegende Beitrag setzt genau hier an und will beim Aufbau von brauchbaren SN mit (fast) beliebigen Weitergabewahrscheinlichkeiten unterstützen. Denn hat man ein empirisches SN erhoben und versucht dies mit konditionaler (probabilistischer) Logik zu modellieren, so kann auch hier ein unbrauchbares Q die Folge sein. Wie also sollten die erhobenen Weitergabewahrscheinlichkeiten angepasst werden, damit Q brauchbar wird? Die Behandlung dieser zentralen Frage führt zu einem interaktiven Algorithmus, in dem ebensolche Anpassungen gemeinsam mit dem Modellierer erarbeitet werden.

In Kapitel 2 befassen wir uns daher zunächst mit Syntax und Semantik der Wissensverarbeitung nach dem Prinzip Minimaler Relativer Entropie (MinRent) sowie der Bestimmung von Ungewissheitsintervallen. Kapitel 3 ist dem Aufbau anwendungskonformer – also brauchbarer Sozialer Netzwerke – gewidmet. In Abschnitt 3.1 wird kurz mess- und wissenschaftstheoretisch die Entstehung von Unbrauchbarkeit diskutiert, um dann einen intuitiven Zugang zur Bestimmung brauchbarer SN zu schaffen. Abschnitt 3.2 präsentiert den formalen Algorithmus zur Behebung von Unbrauchbarkeiten in SN. Letztlich münden die gesamten Ausführungen in einer Anwendung des Algorithmus auf ein mittelgroßes Soziales Netzwerk in Kapitel 4. Kapitel 5 fasst zusammen und versucht einen Ausblick auf zukünftige Forschung.

2 Syntax, Semantik und Brauchbarkeit von MinREnt-Netzen

Da sich dieser Beitrag im Wesentlichen auf die Erkenntnisse von Rödder et al. [10] stützt, wird hier die darin etablierte Notation verwendet; sie wird nun kurz wiederholt.
Man betrachte eine Menge von n Akteuren $a_1, \ldots, a_n$. Jeder Akteur a_i wird im Netz mittels einer binären Variablen V_i mit Ausprägungen $V_i = v_i$ und $v_i = i/\bar{i}$ dargestellt. $v = (v_1, \ldots, v_n)$ sind somit entsprechende Konfigurationen. Für Paare von Akteuren heißen die Ausdrücke $V_j = j \mid V_i = i$ Konditionale – auch abgekürzt als $j \mid i$ –, $\mid$ ist dabei der Konditionaloperator. $V_i = i/\bar{i}$ ist hier die Proposition Akteur a_i hat die Information/Attitüde (i) oder nicht $(\bar{i})$. Die Konditionale sind hingegen bedingte Weitergaben: Wenn a_i diese hat, dann eben auch a_j.
Durch soziologische Erhebungen sind nun für einige Paare von Akteuren (a_i, a_j) Weitergabewahrscheinlichkeiten p_{ij} bekannt. Somit besteht das Soziale Netzwerk aus den Konditionalen und den entsprechenden Weitergabewahrscheinlichkeiten

$$V_j = j \mid V_i = i \text{ mit } p_{ij} \text{ für } (i,j) \in N \subseteq \{1, \ldots, n\} \times \{1, \ldots, n\}. \tag{1}$$

Gesucht wird nun nach einer Wahrscheinlichkeitsverteilung Q auf $\mathcal{V} = \{v\}$, die die Weitergabewahrscheinlichkeiten – kurz WWn – respektiert:

$$\mathsf{Q}(V_j = j \mid V_i = i) = p_{ij} \text{ für } (i,j) \in N. \tag{2}$$

Solch eine Verteilung nennen Rödder et al. in [10] *Netzbelegung*.
Nun lassen sich i.d.R. überabzählbar viele solcher Wahrscheinlichkeitsverteilungen bestimmen, siehe abermals [10]. Die Expertensystemshell SPIRIT [15] gestattet den Aufbau einer speziellen Netzbelgung, nämlich der sog. MaxEnt- bzw. MinRent-Belegung; man betrachte dazu z.B. [12]. Diese Expertensystemshell löst dazu die Aufgaben

$$\begin{aligned} \mathsf{Q}^* &= \arg\max H(\mathsf{Q}) = -\sum_v \mathsf{Q}(v) \log_2 \mathsf{Q}(v) \\ \text{u. d. N. } & \mathsf{Q}(V_j = j \mid V_i = i) = p_{ij},\ (i,j) \in N \end{aligned} \tag{3}$$

bzw.

$$\begin{aligned} \mathsf{Q}^* &= \arg\min R(\mathsf{Q}, \mathsf{P}^0) = \sum_v \mathsf{Q}(v) \log_2 \frac{\mathsf{Q}(v)}{\mathsf{P}^0(v)} \\ \text{u. d. N. } & \mathsf{Q}(V_j = j \mid V_i = i) = p_{ij}, \ (i,j) \in N \end{aligned} \tag{4}$$

Aufgabe (3) bzw. (4) respektiert alle bekannten p_{ij} und baut die Belegung Q^* maximaler Entropie H bzw. minimaler relative Entropie R zur Gleichverteilung P^0 auf $\mathcal{V}$ auf; R bezeichnet man auch als Kullback-Leibler Divergenz.

Ist diese Belegung berechnet, stellt sich die Frage nach konkreten Weitergaben einer Information/Attitüde, falls ein Akteur a_i sie/es abschickt: $V_i = i$. Im Zusammenhang unseres Kontextes heißt dieser Vorgang Evidenziierung.

Dieser Vorgang der Evidenziierung erfolgt in SPIRIT durch Lösen der Aufgabe

$$\begin{aligned} \mathsf{Q}^{**} &= \arg\min R(\mathsf{Q}, \mathsf{Q}^*) = \sum_v \mathsf{Q}(v) \log_2 \frac{\mathsf{Q}(v)}{\mathsf{Q}^*(v)} \\ \text{u. d. N. } & \mathsf{Q}(V_i = i) = 1. \end{aligned} \tag{5}$$

Evidenziieren ist also das Konditionieren einer ganzen Verteilung, und zwar unter der Bedingung, dass $V_i = i$ ist. Hat man Q^{**}, ist damit auch $\mathsf{Q}^{**}(V_j = j)$ für jedes $j \neq i$ berechenbar. $\mathsf{Q}^{**}(V_j = j)$ ist die Wahrscheinlichkeit, dass a_j die Nachricht empfängt, falls a_i sie abschickt. Ob dieser Wert aber durch die Wahrscheinlichkeitsstruktur des bekannten SN determiniert oder bloß ungewisse Vermutung ist, lässt sich erst durch eine Ungewissheitsanalyse überprüfen, siehe [12] oder [8]. Das von SPIRIT bereitgestellte Verfahren kann wie folgt zusammengefasst werden:

Bestimmung des Ungewissheitsintervalls von $V_j = j \mid V_i = i$

1. Löse (4) mit der einzigen Nebenbedingung $\mathsf{Q}(V_j = j \mid V_i = i) = \epsilon$ für ein hinreichend kleines $\epsilon > 0$. Ergebnis $\bar{\mathsf{Q}}^{u*}$
2. Löse (4) für $\bar{\mathsf{Q}}^{u*}$ statt P^0. Ergebnis $\bar{\bar{\mathsf{Q}}}^{u*}$
3. Berechne $u = \bar{\bar{\mathsf{Q}}}^{u*}(V_j = j \mid V_i = i)$
4. Löse (4) mit der einzigen Nebenbedingung $\mathsf{Q}(V_j = j \mid V_i = i) = 1 - \epsilon$ für ein hinreichend kleines $\epsilon > 0$. Ergebnis $\bar{\mathsf{Q}}^{o*}$
5. Löse (4) für $\bar{\mathsf{Q}}^{o*}$ statt P^0. Ergebnis $\bar{\bar{\mathsf{Q}}}^{o*}$
6. Berechne $o = \bar{\bar{\mathsf{Q}}}^{o*}(V_j = j \mid V_i = i)$
7. $[u, o]$ ist das Ungewissheitsintervall

Grundsätzlich heißen Belegungen Q brauchbar, wenn für alle i gilt: $0 < \mathsf{Q}(V_i = i) < 1$, sonst heißen sie unbrauchbar. Eine ausgiebige Begründung für diese Forderung findet sich in [11]. Es leuchtet jedoch sofort ein, dass für z. B. $\mathsf{Q}(V_i = i) = 0$ bzw. $\mathsf{Q}(V_i = \bar{i}) = 0$ die bedingte Wahrscheinlichkeit $\mathsf{Q}(V_j = j \mid V_i = i)$ bzw. $\mathsf{Q}(V_j = j \mid V_i = \bar{i})$ stets entartet ist. Insbesondere bei Auftreten entarteter Verteilungen ist der Algorithmus nicht anwendbar, da u, o nun nicht mehr bestimmbar sind. Eine unbedachte Vergabe von WWn kann zu solchen entarteten Verteilungen führen; in der einschlägigen Literatur nennt man so etwas auch schwache Inkonsistenz, siehe [12]. Diese Situationen sind für die von Rödder et al. in [9] vorgeschlagene Form der Sozialen Netzwerkanalyse ungeeignet.
Wie also die Bestimmung von Ungewissheitsintervallen beim Aufbau eines zur Analyse brauchbaren Sozialen Netzwerkes unterstützen kann, wird im folgenden Kapitel thematisiert.

3 Aufbau anwendungskonformer Netze

3.1 Ziele dieses Beitrages

Hat man sich nun für die Modellierung eines probabilistischen sozialen Netzes entschieden, so ist beim Aufbau eines solchen Vorsicht geboten. Denn die Vergabe aller Weitergabewahrscheinlichkeiten p_{ij} zwischen je zwei Akteuren a_i und a_j muss konsistent zueinander sein. Mit anderen Worten, es muss eine Wahrscheinlichkeitsverteilung auf dem von allen am Sozialen Netzwerk beteiligten Akteuren und den entsprechenden Weitergabewahrscheinlichkeiten aufgespannten Grundraum existieren, vgl. Kapitel 2. Ursachen für Inkonsistenzen können beispielsweise sein:

- Messfehler in der Datenauswertung
- Fehler in der Datenerfassung durch beispielsweise Nichtberücksichtigung relevanter Daten
- Fehler in der Protokollierung der erhobenen Daten
- Mangelnde Objektivität bei der Datenerhebung
- Nicht hinreichende Überprüfung auf Reliabilität, Validität der erhobenen Daten
- Fehler in der angewendeten Methodik zur Datenerhebung
- Keine Eliminierung von sogenannten Ausreißern
- Fehler in der Überführung von erhobenen Daten hin zu Weitergabewahrscheinlichkeiten; für eine tiefere Diskussion solcher Transformationsmöglichkeiten betrachte man z.B. [1].

Da man solche Fehlerquellen, vgl. hierzu auch [5] Seite 30ff., nicht immer ausschließen kann, wird im Folgenden eine Vorgehensweise beschrieben, die den Soziologen (Modellbauer) beim Aufbau eines konsistenten sozialen Netzes unterstützt. Dazu wird ihm ein Algorithmus mit an die Hand gegeben, mit dem er schrittweise die von ihm erhobenen Weitergabewahrscheinlichkeiten auf Plausibiliät hin überprüfen kann und mehr noch, ihm bei Bedarf sogar Vorschläge für notwendige Änderungen unterbreitet werden. Das Ergebnis ist schließlich eine Netzbelegung, die es ermöglicht, auch bisher noch unbekannte Nachrichtenweitergaben zwischen Akteuren festzulegen.
Der Algorithmus wird zunächst verbal und anschließend in Form eines Pseudocodes präsentiert. Anschließend findet er – der Algorithmus – an einem mittelgroßen Beispiel seine konkrete Anwendung, womit dann auch die darüber verfolgte Philosophie evident wird.
Die Ausgangssituation stellt sich wir folgt dar: Durch eine (empirische) Erhebung seien dem Soziologen für gewisse Paare (a_i, a_j) von Akteuren mit $i, j = 1, ..., n$ Wahrscheinlichkeiten p_{ij} der Nachrichtenweitergabe bekannt. Mit der bekannten Semantik aus Kapitel 2 bezeichnet eine Regel $\mathcal{R}^{i_o,j_o} = \{V_{j_o} = j_o | V_{i_o} = i_o \, [p_{i_o j_o}]\}$ für ein ausgewähltes Akteurspaar (a_{i_o}, a_{j_o}) einen solchen Sachverhalt, wobei im Folgenden aus Gründen der Übersichtlichkeit auf die Doppelindizierung verzichtet wird.
Des Weiteren wird angenommen, der Soziologe habe eine Vorstellung über die "Wichtigkeit" der von ihm beim Aufbau eines sozialen Netzes relevanten Paare von Akteuren mit den entsprechenden Weitergabewahrscheinlichkeiten und formuliert diese in einer Prioritätenliste. O.B.d.A. sei diese Liste mit $\mathcal{R}_1, \ldots, \mathcal{R}_l, \ldots, \mathcal{R}_L$, für $\mathcal{R}_1 \geq \mathcal{R}_2, \ldots, \geq \mathcal{R}_L$, und die jeweiligen Weitergabewahrscheinlichkeit mit $p_1, \ldots, p_L$ bezeichnet. $\mathcal{R}_l$ steht also für die vom Soziologen mit l-ter Priorität eingeschätzte Erhebung über das Akteurspaar (a_{i_l}, a_{j_l}) mit der entsprechenden Weitergabewahrscheinlichkeit $p_l := p_{i_l j_l}$.
Für den Fall, dass der Soziologie keine Prioritätenliste auf der Menge der erhobenen Daten angeben kann oder möchte, kann eine zufällige Priorisierung mittels eines Zufallsgenerators erzeugt werden.
Gegeben sei eine vom Soziologen formulierte Prioritätenliste über L Akteurspaare mit *gewünschten* Wahrscheinlichkeiten der Nachrichtenweitergabe $\mathcal{R}^{des} = (\mathcal{R}_1, \ldots, \mathcal{R}_L)$ mit $\mathcal{R}_1 \geq \mathcal{R}_2, \ldots, \geq \mathcal{R}_L$.
Sind sämtliche soziologisch erhobenen Weitergabewahrscheinlichkeiten zwischen je zwei Akteuren konsistent, so existiert ein probabilistisches soziales Netz in Form einer brauchbaren Belegung Q^*, die allen Regeln $\mathcal{R}^{sn} = \mathcal{R}_1 \cup \cdots \cup \mathcal{R}_L$ genügt. Im Falle von Inkonsistenz bedarf es einer Handlungsanweisung für den Modellbauer, die ihn schrittweise beim Auf-

bau eines konsistenten probabilistischen sozialen Netzes wie folgt unterstützt.
In der Initialisierungsphase ($l = 0$) der iterativen Prozedur liegt noch kein Sachverhalt vor, die Regelmenge $\mathcal{R}^{sn}$, mit der das Soziale Netz aufgebaut werden soll, ist also noch leer, womit die Gleichverteilung P^0 gemäß (3) bzw. (4) die aktuelle Netzbelegung repräsentiert.
Im nächsten Schritt ($l = 1$) wird nun "überprüft", ob der mit höchster Priorität eingeschätzte Sachverhalt $\mathcal{R}_1$ zu $\mathcal{R}^{sn}$ konsistent ist. Trivialerweise ist das stets der Fall, so dass die aus der soziologischen Erhebung ermittelte Weitergabewahrscheinlichkeit p_1 für den mit höchster Priorität eingeschätzten Sachverhalt stets direkt übernommen werden kann. Q^{*l} für $l = 1$ bezeichnet die nunmehr aktuelle Netzbelegung gemäß (3) bzw. (4) und es gilt $\mathcal{R}^{sn} = \mathcal{R}_1$.
Nun wird für die mit zweithöchster Priorisierung eingestufte Regel $\mathcal{R}_2$ ihre Konsistenz zu $\mathcal{R}_1$ überprüft. Steht die in $\mathcal{R}_2$ formulierte Forderung nicht im Widerspruch zu der in $\mathcal{R}_1$, so kann auch in diesem Fall die Weitergabewahrscheinlichkeit p_2 übernommen werden. Es bezeichne Q^{*2} die Verteilung, in der $\mathcal{R}^{sn} = \mathcal{R}_1 \cup \mathcal{R}_2$ gilt.
Auf analoge Weise wird nun schrittweise für die weiteren Regeln $\mathcal{R}_3, \ldots, \mathcal{R}_L$ in der Reihenfolge ihrer Prioritäten verfahren.
Angenommen, im l-ten Schritt bedarf es erstmals einer Revision der vom Soziologen erhobenen Weitergabewahrscheinlichkeit p_l zwischen zwei Akteuren. Das ist genau dann der Fall, wenn man den Ausführungen bei Rödder et al. (2016b) bzw. den aus Kapitel 2 folgt und feststellt, dass die Einschätzung über die Weitergabewahrscheinlichkeit p_l nicht im dazugehörigen Ungewissheitsintervall $[u_l; o_l]$ liegt. Der Modellbauer wird in diesem Fall aufgefordert, seine aktuelle Einschätzung über die Weitergabewahrscheinlichkeit p_l des Sachverhalts $\mathcal{R}_l$ zu korrigieren. Um dabei von seiner ursprünglichen Einschätzung so wenig wie möglich abweichen zu müssen, wird er entweder den unteren oder den oberen Grenzwert dieses Ungewissheitsintervalls akzeptieren.

Kompakt zusammengefasst ist der Ablauf des Algorithmus wie folgt:
Gemäß einer vorliegenden Prioritätenliste wird also in der Reihenfolge $\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_L$ sukzessive für jeden Sachverhalt $\mathcal{R}_l$ überprüft, ob dieser konsistent mit den höher in der Priorisierung stehenden Sachverhalten $\mathcal{R}_1 \cup ... \cup \mathcal{R}_{l-1}$ ist. Ist das nicht der Fall, so wird dem Modellbauer ein Intervall $[u_l; o_l]$ für die Weitergabewahrscheinlichkeit p_l vorgeschlagen, mit dem die Beziehung zwischen (a_{i_l}, a_{j_l}) konsistent an das bereits aufgebaute Soziale Netzwerk adaptiert werden kann. Andernfalls kann er die Modellierung mit seiner vorliegenden Einschätzung p_l fortsetzen und den nächsten Sachverhalt $\mathcal{R}_{l+1}$ überprüfen.

Somit wird das Soziale Netzwerk iterativ um konsistente Beziehungen erweitert, bis das gesamte Soziale Netzwerk zur vollständigen Auswertung bereitsteht.

3.2 Interaktiver Algorithmus

Während im vorigen Abschnitt die interaktive Erzeugung eines brauchbaren Sozialen Netzwerkes verbal beschrieben wurde, wird nunmehr der formale Algorithmus dazu präsentiert.

input : A set of probabilistic rules with a given priority sequence
output: A useful social network

initialize an unconnected SN *with* n *actors*;
let $\mathcal{R}^{sn}$ *be the empty rule set for the SN to be generated*;
let $\mathcal{R}^{des} = (\mathcal{R}_1, \ldots, \mathcal{R}_l, \ldots, \mathcal{R}_\mathrm{L})$ *be the desired set of probabilistic conditionals, ordered according to the user's preferences*;
$l := 1$;
while $l \neq L + 1$ **do**
  solve $\bar{\mathsf{Q}}^{u*} = \mathit{arg\ max}\ H(\mathsf{Q})$ *with* $\mathsf{Q}(V_{j_l} = j_l | V_{i_l} = i_l) = \epsilon$ *and* $\epsilon > 0$;
  solve $\bar{\bar{\mathsf{Q}}}^{u*} = \mathit{arg\ min}\ R(\mathsf{Q}, \bar{\mathsf{Q}}^{u*})$ *with* $\mathsf{Q} \models \mathcal{R}^{sn}$;
  calculate $u_l = \bar{\bar{\mathsf{Q}}}^{u*}(V_{j_l} = j_l | V_{i_l} = i_l)$;
  solve $\bar{\mathsf{Q}}^{o*} = \mathit{arg\ max}\ H(\mathsf{Q})$ *with* $\mathsf{Q}(V_{j_l} = j_l | V_{i_l} = i_l) = 1 - \epsilon$ *and* $\epsilon > 0$;
  solve $\bar{\bar{\mathsf{Q}}}^{o*} = \mathit{arg\ min}\ R(\mathsf{Q}, \bar{\mathsf{Q}}^{o*})$ *with* $\mathsf{Q} \models \mathcal{R}^{sn}$;
  calculate $o_l = \bar{\bar{\mathsf{Q}}}^{o*}(V_{j_l} = j_l | V_{i_l} = i_l)$;
  if $p_l \in [u_l, o_l]$ **then**
    $\mathcal{R}^{sn} := \mathcal{R}^{sn} \cup \{V_{j_l} = j_l | V_{i_l} = i_l\ [p_l]\}$;
    $l := l + 1$;
  else
    if *some* $\tilde{p}_l \in [u_l, o_l]$ *is acceptable* **then**
      $\mathcal{R}^{sn} := \mathcal{R}^{sn} \cup \{V_{j_l} = j_l | V_{i_l} = i_l\ [\tilde{p}_l]\}$;
      $l := l + 1$;
    else
      stop algorithm
    end
  end
end

Algorithm 1: Erzeugung brauchbarer Belegungen.

Wie bereits beschrieben, wird hier ein zunächst noch unverbundenes SN als Ausgangsbasis verwendet, um dieses dann sukzessive um Verbindungen zu ergänzen. Die darin geführte Fallunterscheidung ist – natürlich nur bei Auftreten von schwachen Inkonsistenzen – der interaktive Part dieses Algorithmus. Hier wird nämlich vom Modellierer eine Entscheidung erwartet: Weitergabewahrscheinlichkeit revidieren oder Algorithmus abbrechen. Wie das gesamte Prozedere in einer SN-Untersuchung angwendet werden kann, ist Gegenstand des nächsten Kapitels.

4 Eine SN-Anwendung

4.1 Ein Familienbeispiel

Das folgende Beispiel *Verwandtschaftsgrade* ist in leichter Abwandlung aus einer an der FernUniversiät in Hagen geschriebenen Bachelorarbeit entnommen.
Nachstehend die Variablenbezeichner in alphabetischer Reihenfolge mit ihrer Semantik: $AM1$ Ehemann der Cousine; $AM2$ Ehemann der Schwester; AW Ehefrau des Cousins; CW Cousine; G Großmutter; GT Großtante; M Mutter; NM Neffe 2. Grades; NW Nichte 2. Grades; $O1$ Onkel 1; $O2$ Onkel 2; S Schwester; $T1$ Tante 1; $T2$ Tante 2; V Vater.
Tabelle 3 fasst die Informationsbeziehungen zwischen den Verwandten hinsichtlich anstehender Familienfeste mit den entsprechenden Weitergabewahrscheinlichkeiten p_{ij} zusammen.

Regel	p_{ij}	Regel	p_{ij}	Regel	p_{ij}
AM1 | CM1	[0.6]	AM1 | CW	[0.6]	CW | CM1	[0.9]
G | CW	[0.9]	G | GT	[0.9]	NW | G	[0.6]
O1 | T1	[0.6]	CM3 | AW	[0.6]	G | CM3	[0.9]
S | M	[0.7]	B | M	[0.7]	M | V	[0.9]
M | T2	[0.8]	V | M	[0.9]	VF | M	[0.8]
VF | T2	[0.6]	O2 | T2	[0.9]	T2 | CM2	[0.9]
T1 | CW	[0.6]	G | CM1	[0.9]	NW | CW	[0.9]
M | CM3	[0.6]	M | G	[0.8]	NM | CW	[0.9]
G | M	[0.8]	M | T1	[0.6]	AM2 | S	[0.7]
CM1 | T1	[0.6]	B | VF	[0.7]	G | T1	[0.8]
V | T2	[0.6]	B | S	[0.7]	CM2 | T2	[0.6]
CM1 | CW	[0.9]	O2 | CM2	[0.7]		

Tabelle 1: Regelmenge Familienbeispiel.

Für dieses Beispiel – also für die Regelmenge der Tabelle 1 – lässt sich mittels (3) bzw. (4) keine brauchbare entropieoptimale Verteilung bestimmen.

4.2 Umsetzung des Algorithmus für das Familienbeispiel

Der in Abschnitt 3.1 und 3.2 beschriebene Algorithmus unterstützt den Modellierer nun schrittweise beim Aufbau eines brauchbaren sozialen Netzes. In diesem Abschnitt wird das Prinzip des Algorithmus an dem in 4.1 vorgestellten Beispiel näher beschrieben, um so die beim Aufbau eines brauchbaren probabilistischen sozialen Netzes verfolgte Idee zu verdeutlichen.

Damit der Algorithmus überhaupt angewendet werden kann, sind die Regeln zu priorisieren. Zur Veranschaulichung geben wir hier die festgelegten paarweisen Abhängigkeiten zwischen den Verwandten mit den zufällig gewählten Prioritäten (Prio) und den entsprechenden Weitergabewahrscheinlichkeiten p_l zwischen den Verwandten in Tabelle 2.

Prio l	Regel l	p_l	Prio l	Regel l	p_l	Prio l	Regel l	p_l
1	O2 | CM2	[0.7]	13	B | VF	[0.7]	25	CW | CM1	[0.9]
2	AM1 | CW	[0.6]	14	NM | CW	[0.9]	26	M | G	[0.8]
3	CM2 | T2	[0.6]	15	NW | G	[0.6]	27	G | T1	[0.8]
4	G | GT	[0.9]	16	O2 | T2	[0.9]	28	G | CM3	[0.9]
5	O1 | T1	[0.6]	17	AM1 | CM1	[0.6]	29	VF | M	[0.8]
6	CM3 | AW	[0.6]	18	T1 | CW	[0.6]	30	M | T1	[0.6]
7	S | M	[0.7]	19	G | CW	[0.9]	31	B | S	[0.7]
8	VF | T2	[0.6]	20	B | M	[0.7]	32	G | M	[0.8]
9	NW | CW	[0.9]	21	M | T2	[0.8]	33	CM1 | T1	[0.6]
10	T2 | CM2	[0.9]	22	V | M	[0.9]	34	V | T2	[0.6]
11	M | V	[0.9]	23	G | CM1	[0.9]	35	CM1 | CW	[0.9]
12	AM2 | S	[0.7]	24	M | CM3	[0.6]			

Tabelle 2: Gewünschte Regelmenge $\mathcal{R}^{des}$.

Das Ablaufschema im Anhang skizziert nun den Ablauf des Prozederes.

Zu Beginn ist $\mathcal{R}^{sn} = \{\}$; für die in der Prioritätenliste ganz oben stehende Regel, hier $\mathcal{R}_1 = \{O2 \mid CM2\,[0.7]\}$, ist offensichtlich die Brauchbarkeit gewährleistet. Es bezeichne Q_1^* die entropieoptimale Verteilung, in der $\mathcal{R}^{sn} := \{\} \cup \{(O2 \mid CM2\,[0.7]\}$ gilt. Nun wird die in der Prioritätenliste nächste Regel $\mathcal{R}_2$ auf Brauchbarkeit zu $\mathcal{R}^{sn}$ überprüft. Da die Akteure $AM1$ und CW noch nicht in einem unmittelbaren Zusammenhang zu den Verwandten $O2$ und $CM2$ stehen, kann auch die Weitergabewahrscheinlichkeit $p_2 = 0.6$ beibehalten werden; Q_2^* bezeichnet die Verteilung, in der $\mathcal{R}^{sn} := \{O2 \mid CM2\,[0.7]\} \cup \{AM1 \mid CW\,[0.6]\}$ gilt. Auf analoge Weise werden die weiteren Regeln schrittweise gemäß der vorliegenden Prioritätenliste auf Brauchbarkeit hin überprüft. Dabei zeigt sich, dass die ersten 15 Regeln brauchbar sind, was man auch unmittelbar durch Berechnung der Ungewissheitsintervalle in jedem Schritt $l = 1, \ldots, 15$ leicht überprüfen kann. Allerdings mit der an Position 16 priorisierten Regel $\mathcal{R}_{16}$ geforderten Wahrscheinlichkeit von $p_{16} = 0.9$ wird das SN jedoch unbrauchbar. Eine Ungewissheitsanalyse hinsichtlich des Konditionals $O2 \mid T2$ liefert gemäß dem zuvor beschriebenen Algorithmus und bei Wahl von $\epsilon := 10^{-4}$ das Intervall $[u; o] = [0.40; 0.87]$. Mit dieser Information wird der Soziologe nun aufgefordert, seine bisherige Einschätzung

über die Weitergabewahrscheinlichkeit $p_{16} = 0.9$ zu revidieren und sich für einen Wert $0.4 \leqslant p_{16} \leqslant 0.87$ zu entscheiden, um den Aufbau eines brauchbaren sozialen Netzes sicherzustellen; akzeptiert der Modellbauer die vorgeschlagene Korrektur jedoch nicht, so würde das Verfahren hier abgebrochen.
Bei Akzeptanz der Weitergabewahrscheinlichkeit hingegen von $\tilde{p}_{16} = 0.87$ liegt der Soziologe mit seiner ursprüngliche Einschätzung am nächsten. Somit ergibt sich $\mathcal{R}^{sn} := \{\mathcal{R}_1 \cup \cdots \cup \mathcal{R}_{15} \cup \{O2 \mid T2\,[0.87]\}\}$.Für sämtliche restliche Regeln 17 bis 35 ist keine Revision der bekannten Weitergabewahrscheinlichkeiten $p_{17}, \ldots, p_{35}$ erforderlich, womit sich das brauchbare Soziales Netz $\mathcal{R}^{sn} = \{\mathcal{R}_1 \cup \cdots \cup \mathcal{R}_{15} \cup \{O2 \mid T2\,[0.87]\} \cup \mathcal{R}_{17} \cup \cdots \cup \mathcal{R}_{35}\}$ ergibt, das in Abbildung 6 im Anhang dargestellt ist.

5 Zusammenfassung und Ausblick

Empirische Erhebungen sind bekanntlich mit vielen Schwierigkeiten behaftet, z.B. zu kleinen Stichproben, Messfehlern usw. Bei der Analyse Sozialer Netzwerke gilt dies natürlich ebenso. Und gerade hier können solche Erhebungen zu inkonsistenten, in besonderen Fällen sogar zu widersprüchlichen probabilistischen Strukturen führen. Solche Inkonsistenzen aus vorliegendem Datenmaterial zu beseitigen ist Gegenstand dieses Beitrags. Dazu wird ein interaktiver Algorithmus entwickelt, mit dem Weiergabewahrscheinlichkeiten (WWn) zwischen Akteuren im Konsens mit dem Anwender modifiziert werden, wenn sie – die WWn – mit der übrigen Struktur nicht harmonieren. Allerdings wird hier vom Anwender erwartet, dass er die zu überprüfenden Weiergabewahrscheinlichkeiten priorisiert. Gibt es also ein Vorgehen, das unabhängig vom jeweiligen Anwender Inkonsistenzen zu beheben erlaubt? Es gibt also weiteren Forschungsbedarf.

Anhang

Initialisiere das unverbundene SN mit $\mathcal{R}^{sn} = \{\}$*; vgl. Abbildung 1*;
$l := 1$;
$\epsilon := 10^{-4}$;
while $l \neq 35 + 1$ **do**
Für $l = 1$ *und* $(V_{j_1} = j_1 | V_{i_1} = i_1) = (O2 \mid CM2)$ *erhält man*;
$u_1 = \bar{\bar{\mathsf{Q}}}^{u*}(O2 \mid CM2) = \epsilon$;
$o_1 = \bar{\bar{\mathsf{Q}}}^{o*}(O2 \mid CM2) = 1 - \epsilon$;
$p_1 = 0.7 \in [\epsilon, 1 - \epsilon]$;
$\mathcal{R}^{sn} := \{\} \cup \{O2 \mid CM2\,[0.7]\}$;
Der aktuelle Graph ist in Abbildung 2 dargestellt!;
$l := 2$;
Für $l = 2$ *und* $AM1 \mid CW$ *erhält man*;
$u_2 = \bar{\bar{\mathsf{Q}}}^{u*}(AM1 \mid CW) = \epsilon$;
$o_2 = \bar{\bar{\mathsf{Q}}}^{o*}(AM1 \mid CW) = 1 - \epsilon$;
p_2 = 0.6 $\in [\epsilon, 1 - \epsilon]$;
$\mathcal{R}^{sn} := \{O2 \mid CM2\,[0.7]\} \cup \{AM1 \mid CW\,[0.6]\}$;
Der aktuelle Graph ist in Abbildung 3 dargestellt!;
$l := 3$;
$\vdots$
$\mathcal{R}^{sn} := \mathcal{R}^{sn} \cup \{NW \mid G\,[0.6]\}$;
Der aktuelle Graph ist in Abbildung 4 dargestellt!;
$l := 16$;
Für $l = 16$ *und* $O2 \mid T2$ *erhält man*;
$u_{16} = \bar{\bar{\mathsf{Q}}}^{u*}(O2 \mid T2) = 0.4$;
$o_{16} = \bar{\bar{\mathsf{Q}}}^{o*}(O2 \mid T2) = 0.87$;
$p_{16} = 0.9 \notin [0.4, 0.87]$;
$\tilde{p}_{16} = 0.87$ *wird akzeptiert*;
$\mathcal{R}^{sn} := \mathcal{R}^{sn} \cup \{O2 \mid T2\,[0.87]\}$;
Der aktuelle Graph ist in Abbildung 5 dargestellt!;
$l := 17$;
$\vdots$
Keine weiteren Inkonsistenzen!;
Das brauchbare SN ist in Abbildung 6 dargestellt!
end

Algorithm: Ablaufschema zum Familienbeispiel.

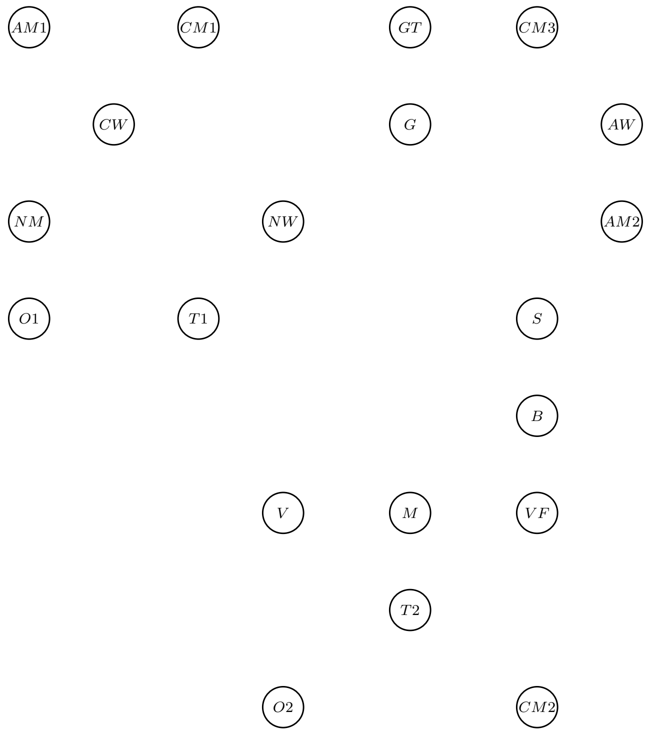

Abbildung 1: Unverbundenes Soziales Netzwerk $\mathcal{R}^{sn} = \{\}$.

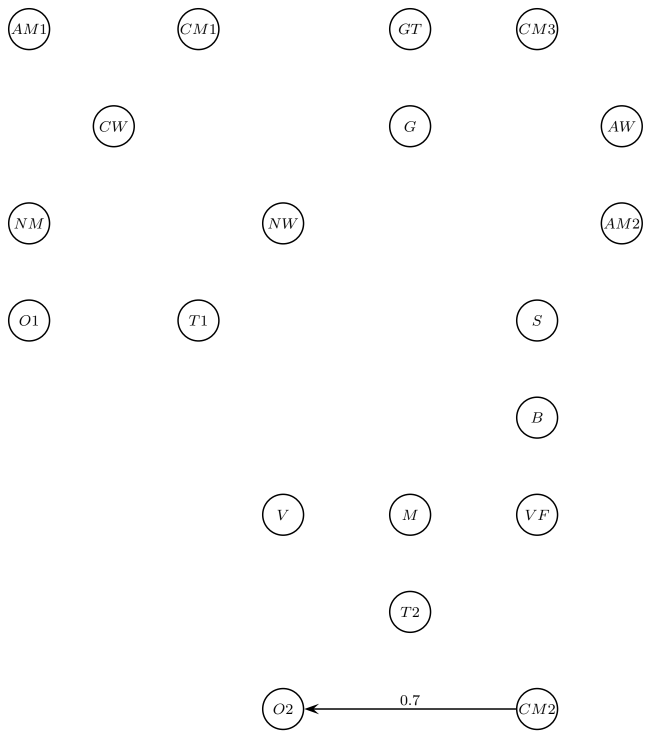

Abbildung 2: Iteration $l = 1$.

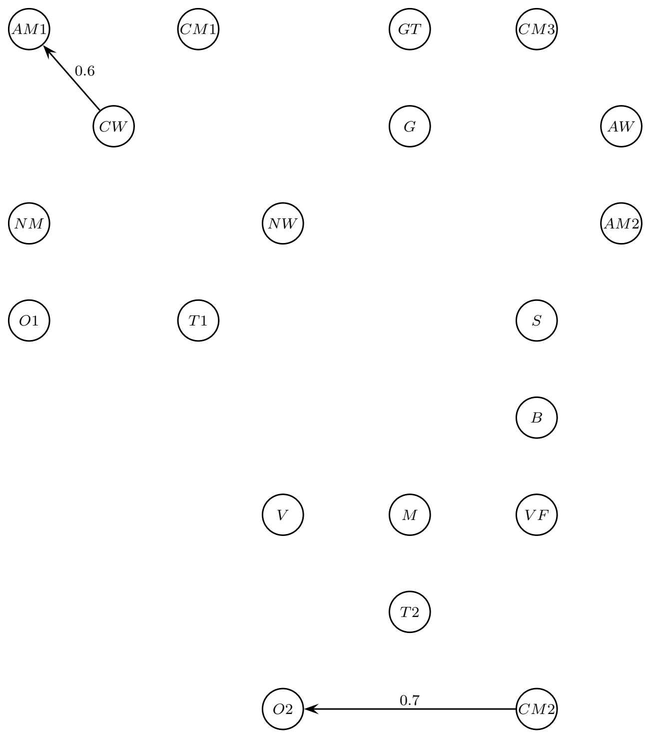

Abbildung 3: Iteration $l = 2$.

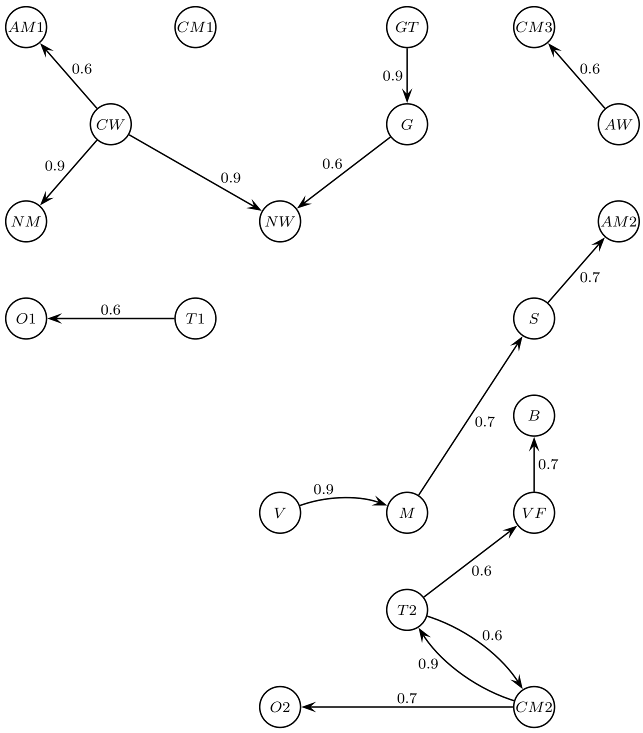

Abbildung 4: Iteration $l = 15$.

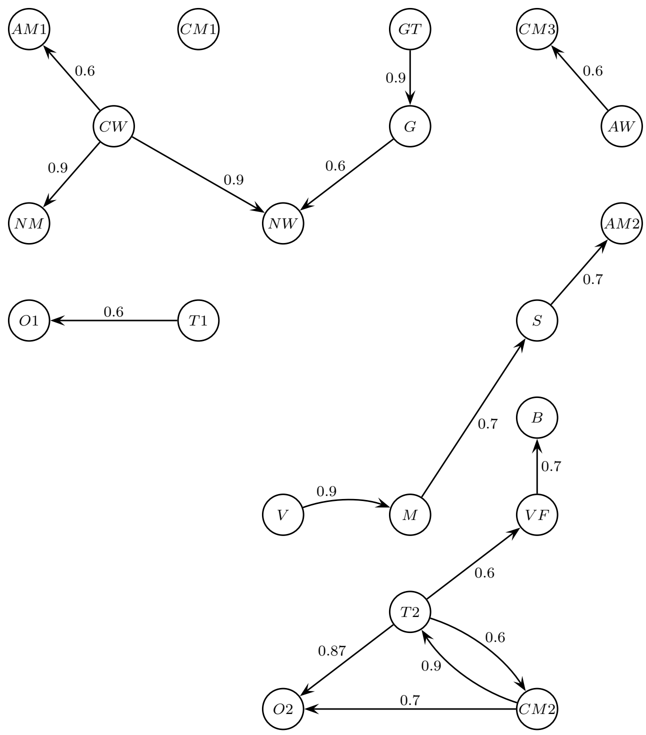

Abbildung 5: Iteration $l = 16$.

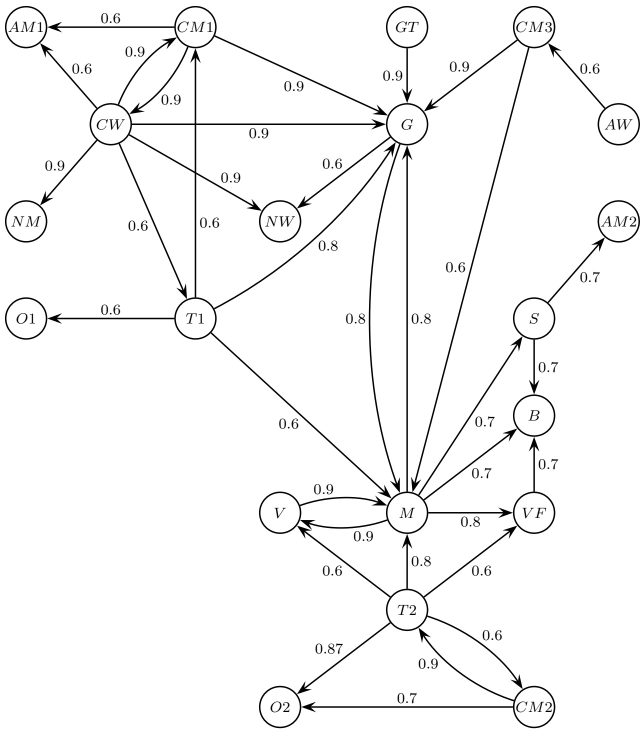

Abbildung 6: Brauchbares Soziales Netzwerk.

Literatur

[1] S. P. Borgatti and Daniel S. Halgin. *Analyzing affiliation networks.* Thousand Oaks, CAThe Sage Handbook of Social Network Analysis, Available at: http://www.bebr.ufl.edu/sites/default/files/Analyzing2011.

[2] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):148–158, 1975.

[3] D. Jansen. *Einführung in die Netzwerkanalyse.* Leske + Budrich, Opladen, 1999.

[4] G. Kern-Isberner. Characterizing the principle of minimum cross-entropy within a conditional-logical framework. *Artificial Intelligence*, 98(1-2):169–208, 1998.

[5] A. Kuß. *Marktforschung: Grundlagen der Datenerhebung und Datenanalyse.* Springer, Wiesbaden, 2012.

[6] J.L. Moreno. *Who Shall Survive: A New Approach to the Problem of Human Interrelations.* Nervous and Mental Disease Publishing Co., Washington, DC, 1934.

[7] M.E.J. Newman. *Networks: An introduction.* Oxford University Press, 2012.

[8] E. Reucher. *Modellbildung bei Unsicherheit und Ungewißheit in konditionalen Strukturen.* Logos Verlag, Berlin, 2002.

[9] W. Rödder, D. Brenner, and F. Kulmann. Entropy based evaluation of net structures deployed in social network analysis. *Expert Systems with Applications*, 41(17):7968–7979, 2014.

[10] W. Rödder, A. Dellnitz, and I. Gartner. Missing links in attenuated social networks. *Working Paper*, 2016. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2780969.

[11] W. Rödder, F. Kulmann, and A. Dellnitz. A new rationality in network analysis – status of actors in a conditional-logical framework –. In C. Beierle, G. Brewka, and M. Thimm, editors, *Computational Models of Rationality*, volume 20, pages 348–364. College Publications, 2016.

[12] W. Rödder, E. Reucher, and F. Kulmann. Features of the expert-system-shell spirit. *Logic Journal of IGPL*, 14(3):483–500, 2006.

[13] J. Scott. *Social Network Analysis*. Sage Publications, London, 2000.

[14] J. Shore and R. Johnson. Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. *IEEE Transactions on Information Theory*, 26(1):26–37, 1980.

[15] SPIRIT, 2011. Last accessed on 2015-03-18.